THE UNIVERSITY OF TEXAS AT AUSTIN
CENTER FOR TRANSPORTATION RESEARCH

TECHNICAL REPORT 0-6838-3
TxDOT PROJECT NUMBER 0-6838

Bringing Smart Transport to Texans: Ensuring the Benefits of a Connected and Autonomous Transport System in Texas (Phase 2)—Final Report

Dr. Kara Kockelman (Research Supervisor)
with Dr. Stephen Boyles, Purser Sturgeon, Dr. Christian Claudel, Lisa Loftus-Otway, Wendy Wagner, Dr. Duncan Stewart, Dr. Guni Sharon, Dr. Michael Albert, Dr. Peter Stone, Josiah Hanna, Yantao Huang, Krishna Murthy Gurumurthy, Dongxu He, Abduallah Mohamed, Rahul Patel, Tian Lei, Michele Simoni and Sadegh Yarmohammadisatri

April 2018; Published July 2018

http://library.ctr.utexas.edu/ctr-publications/0-6838-3.pdf

Technical Report Documentation Page

1. Report No. FHWA/TX-18/0-6838-3	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Phase 2 - Bringing Smart Transport to Texans: Ensuring the Benefits of a Connected and Autonomous Transport System in Texas – Final Report		5. Report Date April 2018; Published July 2018 6. Performing Organization Code
7. Author(s) Dr. Kara Kockelman, Dr. Stephen Boyles, Dr. Christian Claudel, Dr. Peter Stone, Lisa Loftus-Otway, Purser Sturgeon, Dr. Guni Sharon, Krishna Murthy Gurumurthy, Yantao Huang, Michele Simoni, Tian Lei, Rahul Patel, Dongxu He, Abduallah Mohamed, Dr. Jun Liu, Sadegh Yarmohammadi, Eric Thorn, Wendy Wagner		8. Performing Organization Report No. 0-6838-3
9. Performing Organization Name and Address Center for Transportation Research The University of Texas at Austin 3925 W. Braker Ln, Stop D9300 Austin, TX 78759		10. Work Unit No. (TRAIS) 11. Contract or Grant No. 0-6838
12. Sponsoring Agency Name and Address Texas Department of Transportation Research and Technology Implementation Division P.O. Box 5080 Austin, TX 78763-5080		13. Type of Report and Period Covered Technical Report August 2016 – April 2018 14. Sponsoring Agency Code
15. Supplementary Notes Project performed in cooperation with the Texas Department of Transportation.		
16. Abstract This project develops and demonstrates a variety of smart-transport technologies, policies, and practices for highways and freeways using connected autonomous vehicles (CAVs), smartphones, roadside equipment, and related technologies. The intent is to maximize the benefit of these technologies in terms of improved driver safety, reduced congestion, and agency cost savings. For example, in a well-implemented system, advanced CAV technologies may reduce current crash costs by at least $390 billion per year. A poorly implemented system could significantly detract from or reverse these benefits.		
17. Key Words Self-driving vehicles, connected vehicles, connected autonomous vehicles, automated vehicles, smart intersections, transport planning, transport law	18. Distribution Statement No restrictions. This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161; www.ntis.gov.	

19. Security Classif. (of report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of pages 402	22. Price

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

THE UNIVERSITY OF TEXAS AT AUSTIN
CENTER FOR TRANSPORTATION RESEARCH

Phase 2 - Bringing Smart Transport to Texans: Ensuring the Benefits of a Connected and Autonomous Transport System in Texas—Final Report

Dr. Kara Kockelman (Research Supervisor)
with Dr. Stephen Boyles, Purser Sturgeon, Dr. Christian Claudel, Lisa Loftus-Otway, Wendy Wagner, Dr. Duncan Stewart, Dr. Guni Sharon, Dr. Michael Albert, Dr. Peter Stone, Josiah Hanna, Yantao Huang, Krishna Murthy Gurumurthy, Dongxu He, Abduallah Mohamed, Rahul Patel, Tian Lei, Michele Simoni and Sadegh Yarmohammadisatri

CTR Technical Report:	0-6838-3
Report Date:	April 2018; Published July 2018
Project:	0-6838
Project Title:	Bringing Smart Transport to Texans: Ensuring the Benefit of a Connected and Autonomous Transport System in Texas
Sponsoring Agency:	Texas Department of Transportation
Performing Agency:	Center for Transportation Research at The University of Texas at Austin

Project performed in cooperation with the Texas Department of Transportation and the Federal Highway Administration.

Center for Transportation Research
The University of Texas at Austin
1616 Guadalupe, Suite 4.202
Austin, TX 78701

http://ctr.utexas.edu/

Disclaimers

Author's Disclaimer: The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official view or policies of the Federal Highway Administration or the Texas Department of Transportation (TxDOT). This report does not constitute a standard, specification, or regulation.

Patent Disclaimer: There was no invention or discovery conceived or first actually reduced to practice in the course of or under this contract, including any art, method, process, machine manufacture, design or composition of matter, or any new useful improvement thereof, or any variety of plant, which is or may be patentable under the patent laws of the United States of America or any foreign country.

Engineering Disclaimer

NOT INTENDED FOR CONSTRUCTION, BIDDING, OR PERMIT PURPOSES.

Project Engineer: Kara Kockelman
Professional Engineer License State and Number: Texas No. 93443
P. E. Designation: Research Supervisor

Acknowledgments

The authors express appreciation to Project Manager Darrin Jensen, and TxDOT employees Jianming Ma, Travis Scruggs, Melisa Montemayor, Danny Magee, Janie Temple, Joseph Carrizales, Jack Foster, Alex Power, Dale Picha, and Becky Blewett, who served as members of the Project Monitoring Committee. Scott Schauer-West and Amy Banker provided most of the administrative support needed to carry out this project, and Maureen Kelly provided editing support. This research project was ended 5 months early (at month 13 in Phase 2 of this multi-phase project, rather than month 18), so this report is not as comprehensive as originally envisioned, but it remains an excellent reflection of the intense effort and highly meaningful contributions the team has made throughout this project.

Related TxDOT Projects

This report builds on findings produced under TxDOT Projects 0-6838 (Phase 1), 0-6847, and 0-6849, which go deeply into various legal, economic, traffic, and safety impacts of connected and automated vehicles. For details and associated project publications for those and other TxDOT research initiatives, please see the CTR-hosted TxDOT library catalog at http://ctr.utexas.edu/library/.

Table of Contents

List of Tables

List of Figures

List of Acronyms

ABM	activity-based modeling
ADC	analog-to-digital converter
ADS	automated driving system
AIM	Autonomous Intersection Management
ANN	artificial neural network
ATMS	Advanced Traffic Management System
Atruck	autonomous truck
AV	autonomous vehicle (fully automated)
BPR	Bureau of Public Roads
BSM	basic safety message
CAV	connected autonomous vehicle (a communicating and self-driving vehicle)
C/AV	connected and/or automated vehicle (not necessarily fully automated)
CDF	cumulative distribution function
CP	congestion pricing
CTM	cell transmission model
CV	connected vehicle
DMCP	*Difference between Marginal Cost Paths* (heuristic function)
DOT	department of transportation
DP	dynamic programming
DRS	dynamic ride-sharing
DSRC	dedicated short-range communication
DTA	dynamic traffic assignment
DUE	dynamic user equilibrium
FAF	Freight Analysis Framework
FCC	Federal Communications Commission
FCFS	first-come, first-served
FFT	Fast Fourier Transform
FHWA	Federal Highway Administration
FLH	Fast Lax-Hopf
FTC	Federal Trade Commission
GA	genetic algorithm

GUE	generalized-cost UE
H-AIM	Hybrid Autonomous Intersection Management
HAV	highly automated vehicle
Htruck	human-driven truck
HV	human-driven vehicle
IIA	independence from irrelevant alternatives
IMU	inertial measurement unit
LD	long-distance
LH	Lax-Hopf
LSTM	long short-term memory
LTM	link transmission model
LWR	Lighthill–Whitham–Richards model
MCT	marginal cost toll
MCU	microcontroller unit
MDP	Markov Decision Processes
NHTS	National Household Travel Survey
NHTSA	National Highway Traffic Safety Administration
NUMA	National Use Microdata Area
OBU	onboard unit
OD	origin-destination
PDE	partial differential equation
POD	portable onboard device
PSR	present serviceability rating
PTT	*Path Travel Time* (heuristic function)
RUBMRIO	random-utility-based multiregional input-output
RSU	roadside unit
SAM	Statewide Analysis Model
SAV	shared AV
SO	system optimum
SOGA	system-optimal genetic algorithm
SwRI	Southwest Research Institute
TAZ	traffic analysis zone
TBR	tile-based reservation

TE	*Time Evaluation* (heuristic function)
TMC	traffic management center
TNC	transportation network company
TSTT	total system travel time
TxDOT	Texas Department of Transportation
UE	user equilibrium
VMT	vehicle-miles traveled
VOT	value of time
VOTT	value of travel time
WTP	willingness to pay
WWD	wrong-way driver

Chapter 1. Introduction and Report Summary

1.1. Purpose

Smart-driving technologies are changing the landscape of transportation. Substantial mobility, crash reduction, and environmental benefits may ultimately emerge from these technologies, which enable safer and less burdensome road travel. However, in order to optimally capitalize on such benefits for Texas' transportation systems, thoughtful operational strategies, predictions, policymaking, and infrastructure innovations are needed.

This project work developed and demonstrated a variety of smart-transport technologies, predictions, policies, and practices for Texas highways and freeways using highly automated or fully autonomous vehicles (AVs), connected vehicles (CVs), smartphones, roadside equipment, and related technologies.

The work's products provide ideas and equipment for more efficient intersection and network operations for connected, autonomous vehicle (CAV) operations, alongside a suite of behavioral and traffic-flow forecasts for Texas regions and networks under a variety of vehicle mixes (smart plus conventional, semi-autonomous versus fully autonomous, connected but not automated, passenger vehicles and heavy trucks). The work provides and then evaluates various strategies that the Texas Department of Transportation (TxDOT) as well as Texas cities and regions may pursue to bring smarter, safer, more connected, and more sustainable ground transportation systems to Texas, in concert with auto manufacturers, technologists, and the traveling public. The effort supports proactive policymaking on vehicle and occupant licensing, liability, and privacy standards, as technologies become available and travel behaviors change.

1.1.1. Organization of Report

This report's organization largely follows the project work plan, including a series of distinctive and meaningful tasks, from legal analyses to travel behavior and fleet forecasting, and from traffic simulations with smart and micro-tolled intersections and ramp controls to design and demonstrations of location-finding and CV applications for better traffic management, road condition monitoring, and safety improvements across Texas. The following sub-sections offer executive summaries of each chapter of this extensive report, to provide readers an overview of contents and findings.

1.2. Legal Analysis (Chapter 2)

During Phase 1 of this project, the research team conducted an in-depth review of Texas law to ascertain the major issues arising from the introduction of C/AVs and in particular issues that would impact TxDOT. During Phase 2 (September 2016 through writing of this chapter in March

2018), the research team continued to track legislation introduced at the state and federal levels. This section summarizes those legislative activities.

At the federal level, the 2016 election did create some delay as newly appointed agency heads developed new or amended policies. However, this is not to say that no measurable activity took place. In September and October of 2016 the National Highway Traffic Safety Administration (NHTSA) released policies on AVs and best practices for cybersecurity in modern vehicles (NHTSA, 2016 (a) and (b)). These were both *deliberately issued* as policy (rather than regulation) in recognition of the changing dynamic of technologies and public sector groups entering into this field. NHTSA's goal was to encourage technological development and creation of a consistent national framework (rather than a patchwork of state-specific laws that could be incompatible with one another). The AV policy issued by NHTSA reiterated the agency's previous stance regarding the role of the states and the federal government. Notwithstanding NHTSA's efforts in this area, an examination of state-level legislation does indicate that regulatory inconsistencies are potentially arising between the states. NHTSA also began to use a new term *highly automated vehicle* (HAV), re-aligned its set of levels to mirror SAE International's J3016 levels (the global industry reference for defining the six levels of automated/autonomous driving), and noted in the policy that it was specifically set for Levels 3 and above. The cybersecurity policy, however, was clearly aimed at all vehicles—human-driven, partially automated, connected, and HAVs.

Other federal agencies continued work in this area as well. The Federal Trade Commission and Federal Motor Carrier Safety Administration held multiple hearings on HAVs and CVs during 2017. Most importantly, on July 13, 2017, the Federal Communications Commission (FCC) announced it had unlocked new airwaves for vehicular radar use (FCC, 2017b). The FCC expanded the current 76–77 GHz spectrum allocation to include the entire 76–81 GHz band and transition radar devices out of the 24 GHz band. This move allows consistency with the internationally available spectrum, thus reducing the need for vehicle customization across different international markets. Allowing access to this additional spectrum enables these radar devices to better distinguish between objects in areas close to the vehicle, and improves performance for applications such as lane change warnings, blind spot detection, parking aids, stop and follow, stop and go, autonomous braking, and pedestrian detection.

The U.S. Congress also began to introduce legislation on HAVs. The Senate in June 2017 released bipartisan principles for self-driving vehicle legislation. In July 2017 the House Committee on Energy and Commerce's Subcommittee on Digital Commerce and Consumer Protection introduced an unnumbered bill regarding highway AV testing and deployment. The draft legislation proposed to clarify the federal and state roles for regulating HAVs; it would require NHTSA to publish new rules and a safety priority plan for HAVS, and HAV manufacturers to submit safety assessment certifications and develop written cybersecurity plans. The legislation called for a process for controlling access to automated driving systems and providing employee training and management. The draft legislation also required creating a federal advisory committee within NHTSA that would have subcommittees to examine various areas, including mobility access for communities underserved by traditional public transportation.

At the time of writing at the state level it is worth noting that 21 states have now implemented legislation or executive orders regarding C/AVs (four states have executive orders in place). These legislative and executive components run the gamut from commissioning studies, to creating commissions or boards, to drafting extremely detailed and nuanced legislation. The level of regulation similarly runs the gamut from almost completely hands-off integration of the technology immediately onto roadways to much more extensive regulatory oversight. Over 80 bills were introduced in front of state legislatures between fall 2016 and summer 2017.

What is noticeable within the states is that the various bills employ different terminologies within their definitions sections. As noted within Phase 1 report (0-6838-2), nomenclature and definitions are extremely important. While many of the state legislatures referred to the SAE J3016 definitions and indicated they should be used, many did not. As an example, various bills inconsistently employed an array of terminology for HAVs, including *self-driving vehicles*, *autonomous vehicles*, *automated vehicles*, and *highly automated vehicles*.

The consistent use of nomenclature in setting out definitions for driver, operator, technologies, applications, and the vehicles themselves will be extremely useful as these fleets begin to emerge within the U.S. The research team recommends that states begin to look at harmonization terminology within their statutes and regulations, to reduce confusion, set a standard, and reduce litigation anomalies.

1.3. Information Sharing for Connected and Autonomous Vehicles (Chapter 3)

This chapter describes use of a Markov decision process to model CAVs' routing behaviors under uncertain network conditions. We use node-states to describe the conditions of information reception for CAVs when they arrive at a node and prepare to make route decisions. The possible actions for each state include all possible links that will be traversed next based on this state; the transition functions depend on action and next state regardless of the current state.

1.4. Autonomous Intersection Management (Chapter 4)

This chapter proposes a Hybrid Autonomous Intersection Management protocol (H-AIM), which is a variant of the Autonomous Intersection Management protocol (AIM) presented in earlier project reports. Similar to AIM, H-AIM coordinates the right-of-way for CAVs through an intersection in a way that is far more efficient compared to traditional traffic signals. Unlike AIM, H-AIM is shown to outperform traditional traffic signals also when traffic is mostly composed of human-driven vehicles (HVs). For instance, at 10% CAV penetration rate, H-AIM achieves 3% more traffic throughput compared to traditional traffic signals. H-AIM builds on top of existing traffic signal infrastructure, it assumes that HVs stop at red signals and cross intersections on green signals. By assuming the ability to detect incoming HVs, H-AIM is able to safely direct AVs through the intersection even if they arrive on a lane that is assigned a red signal. Experimental results are provided showing that H-AIM can decrease traffic delay for AVs even at a 1%

technology penetration rate. Furthermore, the presented results suggest that restricting HVs' turning options in each lane—e.g., they can only turn right in the rightmost lane as opposed to also having the option to continue straight—is beneficial for AVs. Apart from presenting H-AIM, this chapter also provides general guidelines for assigning lanes with turning options for HVs and As.

1.5. Methods for Implementing Smart Intersections (Chapter 5)

As previously found, smart intersections using a tile-based reservation (TBR) system and assuming a first-come-first-serve (FCFS) priority function have sometimes shown paradoxical effects on road networks, worsening travel times and congestion. These retrogressive effects were seen due to increased queue spillback onto adjacent links of major intersections as well as links adjacent to other intersections within close proximity of the observed major intersection. It was concluded that some combination of signalized and smart-controlled intersections in a large-scale network would provide the most benefit to system-wide congestion and travel times. This chapter presents two methods to identify these subsets of intersections where smart, TBR-controlled intersections would provide not just local benefits of increased throughput in the intersection, but system-wide congestion benefits.

In the first method, a multilinear regression model is estimated to predict the differential impacts of individual intersection characteristics on travel times as a smart and as a signalized intersection. All intersections observed are from the downtown Austin and downtown Dallas networks, and intersection characteristics included relatively easy-to-obtain metrics such as cycle length, intersection capacity, and turning demands. The dependent variable estimated is an intersection's estimated difference in total system travel time (TSTT) between reservation control and signalized control to encapsulate the differential effect of the two controls on the system. To find this data, each eligible intersection is treated as an independent network involving only its directly adjacent links, and a demand table found from simulation of the larger, parent network. Intersection data is used to estimate a regression model, which is then applied to intersections in the downtown Austin network and intersections are ranked according to their difference in TSTT, with those having a larger benefit under reservation control assuming higher ranks. The raw, dependent variable data is tested in simulation using a dynamic traffic assignment (DTA) model to simulate the downtown Austin network by choosing a top percentage of the "best" smart-controlled performing intersections assuming TBR FCFS control, and the rest signal control. Raw results show great improvement in TSTT, beating the base network with 100% TBR control in TSTT by 4% with only 60% of intersections under smart control. TSTT values decrease as the proportion of smart intersections increases. According to ranking results, smart intersections are situated in corridors of consecutive smart intersections, typically in regions of the network away from large clusters of smaller intersections. Regression model results, however, did not show the same benefits, with TSTT values higher than randomized networks of the same smart-control proportions, showing that interdependencies between proximal intersections cannot necessarily be modeled by a linear trend.

The second method offers a genetic algorithm (GA), used in coordination with a DTA model solving for dynamic user equilibrium to find system-optimal (no limitation of the number of TBR intersections) and limited-TBR control optimal allocations of smart intersections. The GA randomly generates an initial population with each individual in the same downtown Austin network under a randomized combination of controls amongst intersections. The GA then uses the concept of natural selection to cross the best-performing individuals of a population to create new, better networks until a convergence is reached. Similar to the raw regression data, the results show TSTT values that fall below the base 100% smart-controlled network TSTT by 4% at just 40% TBR control. The system-optimal GA shows an optimal proportion of 86% of intersections as smart-controlled, with the proportion of smart intersections increasing along with iterations in the GA. Again, GA results show that smart intersections accumulate in "smart corridors" of consecutive TBR-controlled intersections along a roadway.

Overall results show that a completely smart-controlled network will not necessarily result in the best congestion and travel time benefits, and most likely that a placement of smart intersections in corridors of consecutive intersections in less congested areas of the network would work best. Future work must be done to generalize intersection interdependencies by possibly including clustering into a different non-linear regression to generalize intersection characteristics and their effects on system-wide congestion and travel times.

1.6. Road Pricing (Chapter 6)

Communication and computation capabilities are becoming increasingly common on vehicles. Such capabilities present opportunities for developing safer, cleaner, and more efficient road networks. One way of increasing road efficiency is to incentivize vehicles to travel via less congested routes.

It has been known for nearly a century that drivers seeking to minimize their private travel times need not minimize the total level of congestion. In other words, self-interested drivers may reach a user equilibrium (UE) that is not optimal from a system perspective. On the other hand, disincentivizing vehicles to traverse certain links (using tolls for instance) can lead to the system optimum (SO).

This chapter discusses the concept of micro-tolling, defined as the ability to set individualized and dynamic toll values for each link within a road network. Specifically, this chapter tackles a problem where, given current, observable traffic conditions (traffic volume, travel speed, travel time, etc.), the goal is to set a dynamic toll value to each link such that the UE aligns with the SO.

This chapter introduces a novel micro-tolling scheme denoted Delta-tolling. Delta-tolling assigns a toll to each link proportional to the difference between its current travel time and its free-flow travel time (denoted Delta). The constant of proportionality (denoted beta) requires tuning. Since Delta changes according to observed traffic, Delta-tolling is adaptive to traffic changes in real time. Since computing the toll value is done locally for each link, Delta-tolling is tractable for large networks.

This chapter conjectures that Delta-tolling leads to optimal system performance. Two types of supporting evidence are provided for this claim. From a theoretical standpoint, we show that under additional assumptions Delta-tolling is equivalent to marginal-cost tolling, which provably yields optimal system performance. From an empirical standpoint, using two different traffic simulation models, we show that Delta-tolling leads to a significant improvement in system performance, up to 33% and 32% improvement in social welfare and average travel time respectively. As the annual cost of traffic congestion in the United States alone is $160 billion, even small reductions in travel time can have dramatic benefits.

To the best of our knowledge Delta-tolling is the first tolling scheme that is adaptive in real time, able to scale to large networks, does not assume user homogeneity, and enhances system performance. Moreover, given appropriate communication capabilities Delta-tolling is practical to implement in real life.

1.7. Partial Compliance with Tolls (Chapter 7)

In multiagent systems, there are generally two paradigms of interaction. Centralized control paradigms assume that a single decision-making entity is able to dictate the actions of all the agents, thus leading them to a coordinated social optimum. Decentralized control paradigms, on the other hand, assume that each agent selects its own actions, and while it is in principle possible for them to act altruistically, they are generally assumed to be self-interested. A central theme of multiagent mechanism design is finding interaction mechanisms for self-interested agents that incentivize them to reach coordinated behavior that is as close as possible to the social optimum.

In this chapter, we consider a routing scenario in which a subset of agents is controlled centrally (compliant agents), while the remaining are self-interested agents. We model the system as a Stackelberg routing game in which the decision-maker for the centrally controlled agents is the leader, and the self-interested agents are the followers.

Additionally, we provide a computationally tractable methodology for 1) determining the maximum number of self-interested agents that a system can tolerate at optimal flow, 2) determining whether a given subset of centrally controlled agents are sufficient to achieve SO, and 3) computing the actions the leader should prescribe to a sufficient set of compliant agents in order to achieve SO.

A known fact in routing games is that agents seeking to minimize their private travel time need not minimize the total system's travel time. That is, self-interested agents may reach a UE that is not optimal from a system perspective. However, if all agents are assigned paths with minimum system marginal cost then the system will achieve optimal performance.

Therefore, from a system manager perspective, it is desirable that all agents traversing a network would strictly utilize minimal marginal cost paths, even if such paths are not of minimum travel time for an individual agent. However, in many important scenarios, it will not be possible to

enforce path assignment on all agents, but it may be possible to affect the behavior of a subset (the compliant agents).

As a motivating example, consider an opt-in tolling system where drivers are given positive incentives to enroll but, in exchange, they will be subject to tolls that affect their route choice.

We show that, in the general case, computing the optimal assignment of compliant agents is NP-hard. Therefore, we focus on the specific scenario where the portion of compliant agents is sufficiently large to achieve SO.

We present a novel linear program representation for computing the maximal portion of self-interested agents that allow the system to achieve SO and to determine whether a given set of compliant agents is sufficient to achieve SO.

Furthermore, we provide a method to tractably compute the flow assignment for the compliant agents such that SO performance is guaranteed. Next, we prove that in this case, assigning minimal marginal cost paths to all compliant users leads to the SO.

Experimental results, performed using a standard traffic simulator, are provided and demonstrate that the number of compliant agents necessary to achieve SO can be a relatively small percentage of total flow (between 13% and 53%).

1.8. Performance Guarantees for Micro-Tolling (Chapter 8)

Self-interested agents that are routed in a congestible network, such as vehicles in a road network, impose a UE that is often far worse than the SO flow. Charging marginal cost tolls (MCT), in which each agent is charged a toll equivalent to the damage it inflicts on all other agents, results in a UE that achieves SO performance.

Calculating the MCT for a given agent, *a*, on a given path, *p* (i.e., the damage *a* inflicts by traversing *p*) is very challenging without making specific assumptions (e.g., well-defined and known travel time functions) that do not hold in most traffic models and certainly not in real-life traffic. Recent work suggested a model free technique, denoted *Delta-tolling*, for evaluating MCT.

Since Delta-tolling, or any scheme that approximates MCT, is not guaranteed to result in the exact MCT, no optimality guarantees can be given regarding the system's performance. In fact, applying tolls different than MCT might result in a system performance that is worse than not applying tolls at all. This fact might deter public officials from implementing any tolling scheme that is not guaranteed to impose the exact MCT.

This chapter examines the impact of imposing inaccurate MCT on the system’s performance. Specifically, we provide conditions under which the system performance will not be worse than applying no tolls, i.e., the system will not be worse off by imposing the tolling scheme. This chapter establishes that charging a toll that is off by a factor, *r*, from the true MCT will not hurt the system's performance if $0 <= r <= 1$ (i.e., if MCT is underestimated by a constant factor).

Moreover, this chapter proves that the function mapping r to the system's performance (total travel time) has a single (global) minima and no local extrema. This fact implies that calibrating schemes for evaluating MCT, e.g., Delta-tolling, can be carried out by identifying a minimum, which is guaranteed to be the global optimum.

Finally, experimental results from a traffic simulator are presented for different traffic scenarios. The experimental results support our theoretical claims by showing that, across various traffic scenarios, a global optimal flow is achieved for r=1 and no extrema exist elsewhere.

1.9. CAVs' Impacts on Long-Distance Household Travel across the U.S. and Texas (Chapter 9)

This chapter investigates Americans' preferences in adopting and using AVs. Rapid advances in technologies have accelerated the timeline for public use of fully-automated and communications-connected vehicles. Public opinion on self-driving vehicles or AVs is evolving rapidly, and many behavioral questions have not yet been addressed. This study emphasizes AV mode choices, including Americans' willingness to pay (WTP) to ride with a stranger in a shared AV (SAV) fleet vehicle on various trip types and the long-distance travel impacts of AVs. The 2,588 complete responses to a stated-preference survey with 70 questions provide valuable insights on privacy concerns and crash ethics, safety and ride-sharing with strangers, long-distance travel, and preferences for smarter vehicles and transport systems. While the starting sample data were relatively demographically unbiased, Texans were purposefully over-sampled, and all statistics adjusted/corrected (via sample weights) to match U.S. demographics on gender, education, income, and age. Weighted results suggest that Americans are willing to pay, on average, $2073 to own AVs over conventional vehicles and an additional $1078 to maintain/include a manual driving option on such vehicles. Ride-sharing will be popular at 75¢ per mile, under most scenarios, and many Americans are willing to pay $1, on average, to anonymize their trip ends' addresses. Most are also willing to let children 16 years of age and older have unsupervised access to AVs (both privately owned and shared). Nearly 50% of long-distance travel appears captured by AVs and SAVs in the future, rather than airlines, at least for one-way trip distances up to 500 miles.

1.10. Survey Analysis to Determine CAVs' Impact on Travel (Chapter 10)

Two hurdle models (which allow for a high share of zero-value responses) were estimated in this chapter: one to predict WTP to share a ride and another to determine WTP to anonymize location while using AVs. The first two-part model shows how travel time delays, person and household attributes, and land use densities can significantly affect Americans' willingness to share rides. The second hurdle model suggests that traveler age, presence of children, household income, vehicle ownership, and driver's license status are major predictors of one's WTP to obscure pick-up and drop-off locations.

A binary logit was used to model current mode choice for long-distance (over 50 miles, one-way) travel (between one's private car and an airplane), with household income as the leading predictor. On average, older Americans and/or those with children prefer such travel by car. Finally, a multinomial logit anticipated mode shifts when AVs and SAVs become available and affordable. Everything else constant, private cars remain preferred by older people, but SAVs may be used in the future for more business travel.

1.11. Traffic and Trade Impacts of Automated Trucking (Chapter 11)

This chapter anticipates changes in U.S. highway and rail trade patterns following widespread availability of self-driving or autonomous trucks (Atrucks). It uses a random-utility-based multiregional input-output (RUBMRIO) model, driven by foreign export demands, to simulate changes in freight flows among 3109 U.S. counties and 117 export zones, via a nested-logit model for shipment or input origin and mode, including the shipper's choice between autonomous trucks and conventional or human-driven trucks (Htrucks). Different value of travel time and cost scenarios are explored, to provide a sense of variation in the uncertain future of ground-based trade flows.

Using the current U.S. Freight Analysis Framework (FAF4) data for travel times and costs—and assuming that Atrucks lower trucking costs by 25% (per ton-mile delivered)—truck flow values in ton-miles are predicted to rise 11%, due to automation's lowering of trucking costs, while rail flow values fall 4.8%. Rail flows are predicted to rise 6.6% for trip distances between 1,000 and 1,500 miles, with truck volumes rising for other distances. Introduction of Atrucks favors longer truck trades, but rail's low price remains competitive for trade distances over 3,000 miles. Htrucks continue to dominate in shorter-distance freight movements, while Atrucks dominate at distances over 500 miles. Eleven commodity sectors see an increase in trucking's domestic flows, and twelve see increased export flows. The total ton-miles across all 13 commodity groups rise slightly by 3.1%, as automation lowers overall shipping costs.

1.12. Agent-Based Population from Four-Step Data (Chapter 12)

This chapter introduces methodology to synthesize person-level data from traditional data sources. Car-sharing offers travelers an alternative method of transport in or between cities; the transformative implementation of CAVs will likely further promote the sharing. To provide decision-makers reasonable information about car-sharing strategies or shared CAVs, transportation planners and researchers are looking for advanced travel modeling approaches. Activity-based modeling (ABM) is one of the most promising approaches, modeling travel demand at the person-level and offering great temporal and spatial details about individuals' travel patterns. Currently, the four-step travel demand modeling process is the most commonly used approach that is trip-based, modeling travel demand at an aggregated level of traffic analysis zones. However, this approach is unable to track individuals' travel patterns with great spatial details. ABM can be used to estimate the impacts of car-sharing in transportation systems and evaluate the

policies/strategies related to the CAV operations. ABM takes the individual's daily activities chained by a series of travel trips (also called "tour" if the last trip ends where the first trip starts, e.g., home) as the travel demand input. The input can be simply summarized into "4Ws": Who this individual is, where this individual lives and works, what daily activities this individual person does, and when this individual plans to perform activities. This study delivers a methodological framework to prepare the "4W" inputs, taking advantage of existing travel model data (including the travel survey data) and open-source data (e.g., Open Street Maps). This chapter presents a programming-based tool composed of a series of algorithms that output synthetic population, synthetic locations for activities, travel tours (i.e., chained trips and activities), and travel schedules for performing activities, respectively. The tool is particularly useful for planning practitioners from state agencies and regional planning organizations who already have the data (e.g., regional travel models and travel survey data) and seek to convert their existing trip-based models to activity-based models that may be more suitable for simulating the individuals' travel patterns.

1.13. Potential for Dynamic Ride-Sharing with SAVs: Study with Cellphone Data (Chapter 13)

This chapter discusses potential for dynamic ride-sharing by using a cellphone dataset to simulate trip-matching as probable by a fleet of SAVs. Transportation network companies are regularly demonstrating the economic and operational viability of dynamic ride-sharing (DRS) to any destination within a city (e.g., uberPOOL or Lyft Line), thanks to real-time information from smartphones. In the foreseeable future, fleets of SAVs may largely eliminate the need for human drivers, while lowering per-mile operating costs and increasing the convenience of travel. This may dramatically reduce private vehicle ownership and deliver extensive use of SAVs. Using AirSage's cellphone-based trip tables across 1,267 zones over 30 consecutive days, this study anticipates DRS matches (by assigning independent travelers with overlapping routes in time and space to the same SAV) and simulates SAV travel across the Orlando network to determine optimal SAV fleet size. Those results suggest significant opportunities for DRS-enabled SAVs: nearly 60% of the single-person trips can be shared with other persons traveling solo and with less than 5 minutes added travel time (to arrive at their destinations). This value climbs to 80% and 86% for 15 and 30 minutes of added wait or travel time, respectively. The results indicate that 120,000 SAVs will be required to meet less than 45% of seats in Orlando's 2.8 million single-traveler trips. In other words, just 1 SAV per 20 person-trips, on average, is able to serve almost half the region's demand, helping reduce congestion while filling up passenger vehicle seats.

1.14. Pricing Strategies with CAVs in the Mix (Chapter 14)

This chapter employs different pricing strategies in an agent-based model to assess a future with AVs, SAVs, and traditional modes of transport and how mode-shares will look. The introduction of autonomous (self-driving) vehicles and SAVs will affect travel destinations and distances, mode choices, vehicle-miles traveled, and congestion. Although some congestion reduction may be achieved (thanks to fewer crashes and tighter headways, long-term), car-trip frequencies and VMT are likely to rise significantly in most settings, compromising the benefits of driverless vehicles.

Congestion pricing (CP) and road tolls are both key tools for moderating demand and incentivize more socially optimal travel choices. This work develops multiple CP and tolling scenarios and investigates their effects on Austin network conditions and traveler welfare, using the agent-based simulation model MATSim. Results suggest that, although all different CP schemes manage to reduce congestion and improve modal share, their impacts on social welfare differ from each other. More complex and advanced schemes may considerably improve traffic efficiency, but they need not necessarily bring higher economic benefits. The possibility to refund users by reinvesting toll revenues can play a crucial role in the overall efficiency of each CP strategy.

1.15. Technologies for Congestion Pricing (Chapter 15)

Congestion pricing of high-demand roadways seeks to influence travelers' route choices, trip timing, modes, and destination choices, to keep vehicles moving and avoid excessive congestion. This chapter describes the use of various technologies to enable more advanced and cost-effective congestion pricing applications.

Video-based systems require cameras to capture the state of traffic, plus some form of communication back to users. Both DSRC and cellular-based systems use GPS data to price roads and toll users based on traffic conditions. DSRC employs roadside units (RSUs) to receive and send messages to in-vehicle DSRC units. A cellular-based system could use communications from cellular towers in combination with a smartphone, on-board diagnostics port (OBD-II), or pre-installed cellular chip. DSRC is a recommended technology to pilot congestion pricing at highly congested locations, such as bridges and major highways, while cellular communications enable congestion pricing across entire networks.

VMT taxes can be relatively simple, or variable in space and time, facilitating transportation-agency cost recovery. A next step for roadway management is CBCP, which can better reflect the marginal delay costs of one's travel choices and enable a more equitable distribution of each community's scarce roadway assets.

1.16. Traffic Flow Estimation Using Fast-Algorithms for Fast-Forward Simulations of Macroscopic Traffic Models (Chapter 16)

CV applications of this work are discussed in this chapter and involve the use of IMUs (inertial measurement units) to solve a variety of sensing problems associated with vehicle operations. We focused on the use of IMUs to determine the condition of the pavement (PSR, or present serviceability rating) by monitoring vertical acceleration timeseries and comparing some features of these timeseries to human-determined PSR data. We examined the positioning accuracy of GPS-IMU systems, and the possibility of using high-resolution positioning data to determine the state of traffic on multiple lanes of a highway. This requires both the investigation of the accuracy levels of current GPS-IMU combinations, and the possibility of computing solutions to macroscopic multilane flow models quickly and efficiently. Finally, we investigated the use of acceleration and

rotation rate data to detect potentially dangerous areas in the transportation network that result in frequent abrupt user inputs (braking or steering).

1.17. Development of an IMU-Based Traffic and Road Condition Monitoring System (Chapter 17)

This chapter presents a new type of wireless platform designed for real-time traffic and road surface monitoring. The sensor platform is built around a 32-bit ARM Cortex M4 microcontroller and a LSM9DS0 IMU module, as well as a short-range Bluetooth transceiver. This platform is designed to enhance the performance of probe vehicles and can be easily installed or retrofitted to a vehicle using a USB car charger. A self-calibration unit is designed to improve the applicability and accuracy through automatically computing the relation between the coordinates of the device and the coordinates of the vehicle after a short calibration period. This system is multipurpose, and can be used to obtain an estimate of the trajectory of the vehicle, relative to a fixed sensor network on the ground. It can also be used to monitor the vehicle speed, as well as additional information regarding traffic—for example the presence of accidents or of stop-and-go waves. The same device can also be used to monitor pavement condition through vertical acceleration measurements.

This chapter also details the main versions of this platform, including first, second, and third generations; their design principles (including microcontroller and peripherals); and their measurement capabilities.

1.18. Cybersecurity Analysis of Connected Vehicles Using Deep Learning (Chapter 18)

This chapter describes the research team's efforts to meet two objectives: 1) solve an inverse modeling problem (predicting the vehicle dynamics from vehicle input commands, and from vehicle measurement data) and 2) use this dynamical model to detect input faults or spoofing, or sensor faults or spoofing (particularly for GPS position sensors). We used deep neural networks to obtain the most accurate representation of the vehicle dynamics. We used this dynamical model in conjunction with the initial state of the vehicle (determined by the vehicle sensors) and with the inputs to the vehicle system (which can be the output of the AV guidance computer in an AV, or the positions of the throttle, brake pedals, and steering wheel in a conventional vehicle), to simulate the vehicle trajectory. If this simulated trajectory significantly deviates from the actual position of the vehicle, then either the model is incorrect, or one or multiple sensors and actuators are spoofed or faulty. We tested this on a simulated GPS spoofing scenario involving fake positional data.

1.19. Prototype Development and Limited Deployment of CAV Technologies on Texas Roadways (Chapter 19)

In Phase 1 of this project, research team members from the Southwest Research Institute (SwRI) demonstrated vehicle-to-vehicle applications for emergency vehicle alert, emergency electronic brake lights, and intelligent message propagation and vehicle-to-infrastructure applications for

static and dynamic wrong-way driver detection and road condition monitoring. Leveraging these systems and applications for Phase 2, SwRI extended the wrong-way driver detection system and the porting and transition of UT Austin's Autonomous Intersection Management (AIM) system onto physical vehicles and infrastructure for demonstration at the SwRI facilities. This chapter describes this work in some detail.

Chapter 2. Legal Analysis

2.1. Federal Update

This section details federal activities. It runs in reverse chronological order, from the date of writing this final report back to the start of Phase 2 of this project in September 2016.

2.1.1. U.S. Congress

The U.S. Senate Committee on Commerce, Science and Transportation held a hearing on June 14, 2017 regarding the release of bipartisan principles for self-driving vehicle legislation. Those principles can be seen in Figure 2.1.

"**Prioritize Safety:** As with conventional vehicles, federal standards will be important to self-driving vehicle safety.

Legislation must consider both the near-term and long-term regulatory oversight of these vehicles, recognizing that new safety standards governing these vehicles should eventually be set.

Promote Continued Innovation and Reduce Existing Roadblocks: Currently, there is a body of regulations governing conventional vehicles, developed over decades, that does not directly address self-driving vehicles. Developing new standards takes significant time.

Legislation must allow the life-saving safety benefits of self-driving vehicle technology to move forward as new standards development is underway.

Legislation must find ways to preserve and improve safety while addressing incompatibility with old rules that were not written with self-driving vehicles in mind.

Remain Tech Neutral: Self-driving vehicles are likely to take different forms, use diverse technologies, serve consumers with varying capability levels, and follow multiple business models.

Legislation must be technology neutral and avoid favoring the business models of some developers of self-driving vehicles over others.

Reinforce Separate Federal and State Roles: Traditionally, the federal government has regulated the vehicle itself, while states have regulated driver behavior.

Legislation must clarify the responsibilities of federal and state regulators to protect the public and prevent conflicting laws and rules from stifling this new technology.

Legislation must be based on the existing relationship between federal and state regulators and their current separation of authority, but make necessary targeted updates for new challenges posed by the current regulatory environment with respect to self-driving vehicles.

Strengthen Cybersecurity: Cybersecurity should be a top priority for manufacturers of self-driving vehicles and it must be an integral feature of self-driving vehicles from the very beginning of their development.

Legislation must address the connectivity of self-driving vehicles and potential cybersecurity vulnerabilities before they compromise safety.

Educate the public to encourage responsible adoption of self-driving vehicles: Government and industry should work together to ensure the public understands the differences between conventional and self-driving vehicles.

Legislation must review consumer education models for self-driving vehicles and address how companies can inform the public on what self-driving vehicles can and cannot do based on their level of automation and their individual capabilities."

Source: US Congress, 2017

Figure 2.1 Principles for bipartisan legislation on self-driving vehicles

Twenty bills were introduced into the 115th U.S. Congress regarding automated vehicles, or ancillary and tangential issues. These include thirteen bills that directly regulate automated vehicles. H.R. 3401, H.R. 3388, H.R. 3416, S. 1885, H.R. 3404, H.R. 3407, H.R. 3430, H.R. 3405, H.R. 3412, H.R. 3411, H.R. 3414, H.R. 3413, and H.R. 3408.

The SEVEN ancillary/tangential bills include H.R. 2778, S.1809, H.R. 3901, H.R. 346, H.R. 3421, H.R. 4625 and S. 2217

None of these bills as at time of drafting this chapter in this Phase 2 report has passed both chambers in the U.S. Congress.

- H.R. 2778 – Less Traffic with Smart Stop Lights Act 2017.
- S.1809 – Smart grants to cities/states re connected and automated vehicles.
- H.R. 3901 – smart grant
- H.R. 3406 – partial exemption on number of vehicles with reduced bumper safety – but for automated.
- H.R. 3421, another exemption bill creating database of exemptions and noting AV type vehicles.
- H.R. 4625 and S2217 – artificial intelligence

On July 28, 2017, the U.S. House introduced H.R. 3401 to amend chapter 301 of subtitle VI of title 49, United States Code, to update or provide new motor vehicle safety standards for HAVs and other purposes. The bill defines automated driving system(s), dynamic driving task(s), HAV and operational design domain. The bill requires the Secretary to issue rules within 24 months requiring the submission of safety assessment certifications regarding how safety is being addressed by each entity developing HAVs or ADS (U.S.C., 2017 (d)). In the interim, the bill would require safety assessment letters are submitted to NHTSA under its policy issued in September 2016 or under any successor guidance. If this bill moves forward, it would require amendment to NHTSA's September 2017 guidance, which now only has voluntary safety self-assessment guidance, where entities can choose to submit or not submit.

On July 25, 2017, the U.S. House introduced the Safely Ensuring Lives Future Deployment and Research in Vehicle Evolution Act (Self Drive Act) H.R. 3388 (U.S.C., 2017(b)). H.R. 3388 was passed by the House as amended and agreed to by voice vote on September 6, 2017. It was received in the Senate and read twice before it was referred to the Senate Committee on Commerce, Science and Transportation on September 6, 2017. HR 3588 clarifies federal and state roles. It preempts states (and political sub divisions) from creating laws or regulations regarding the design, construction or performance of HAVs, automated driving systems or components of automated driving systems, unless the laws are identical to federal laws. HR 2588 would require NHTSA to issue within 18 months rules on submission of safety assessment certifications on how safety is

addressed by manufacturers of highly automated vehicles or automated driving systems. NHTSA should identify elements in HAVs that may require performance standards including human machine interface, sensors, and actuators, and consider process and procedure standards for software and cybersecurity as necessary

The act also amends Chapter 3001 of Subtitle VI of Title 49 USC by adding a new section on rear seat occupant alert systems. The Secretary of Transportation is required to issues rules, within two years of enactment, requiring all new passenger motor vehicles weighing less than 10,000 pounds gross vehicle weight to be equipped with an alarm system to alert the operator to check rear designated seating positions after the vehicle motor or engine is deactivated by the operator.

H. R. 3888 also requires manufacturers to develop written cybersecurity plans. H.R. 3888 requires the Secretary to also create within six months of enactment a Highly Automated Vehicle Advisory Council Membership that is diverse, and will be determined by the USDOT Secretary.

On July 26, 2017 the U.S. House introduced H.R. 3416 to establish in the National Highway Traffic Safety Administration a Rural and Mountainous Advisory Council. This council would make recommendations regarding the testing and deployment of HAVs and automated driving systems in areas that are rural, remote, mountainous, insular, or unmapped (U.S.C., 2017 (e)). The council would be convened by NHTSA within six months of enactment. Council members are to be appointed for three-year terms by the Secretary. The Council will undertake information gathering, develop technical advice, and present best practices or recommendations to the Secretary. The council will terminate six years after enactment. Within Section 1 automated driving system, dynamic driving task, highly automated vehicle and operation design domain are defined. H.R. 3416 notes that if SAE revises definitions, it must notify the Secretary who is required to publish these within the federal register for comment. If the Secretary determines that this new definition does not meet the need for motor vehicle safety or is otherwise inconsistent with United States Code, the existing Section 1 definition shall remain in effect. If the Secretary does not reject a definition revised by SAE it will amend regulations and standards as necessary.

On September 8 the Senate Commerce Committee circulated the American Vision for Safer Transportation through Advancement of Revolutionary Technologies Act, (AV START Act – S.1885). The bill has similarities to the House's SELF Drive Act (H.R. 3588), but also some major departures within specific sections. Most notably including addressing trucking. The definitions section includes brackets pertaining to a vehicles weight, so inclusion of trucks and buses is considered within this bill. S. 1885 makes major differences in the approach to preemption, with AV laws and regulations enacted by states considered to be pre-empted if they pertain to any of nine subject areas of the Safety Evaluation Report that this bill requires (U.S.C., 2017 (a)).

H.R. 3404 provides for the establishment of a NHTSA HAV Advisory Council (U.S.C., 2017 (f)). The proposed act would require NHTSA to establish a HAV Advisory Council within six months of enactment. Council members are to be appointed for year 3-year terms by the Secretary. Membership is to be diverse and include business, academia, state and local representation, labor organization, environmental experts and other members. Any subcommittee of this Council shall

be composed of not less than 15 and not more than 30 members appointed by the Secretary. The Council may form subcommittees as needed to undertake information-gathering activities, develop technical advice, and present best practices or recommendations to the Secretary regarding:

> (1) labor and employment issues that may be affected by the deployment of HAVs;
> (2) the impact of the development and deployment of HAV on the environment;
> (3) protection of consumer privacy and security of information collected by HAVs; and
> (4) cabin safety for highly automated vehicle passengers, and how automated driving systems may impact collision vectors, overall crashworthiness, and the use and placement of airbags, seatbelts, anchor belts, head restraints, and other protective features in the cabin.

The council's recommendations are to be reported to Congress. The council shall terminate six years after the Acts enactment.

H.R. 3407 was introduced in September 26, 2017 to require a cybersecurity plan for highly automated vehicles, and for other purposes (U.S.C., 2017 (g)). It has not yet been introduced into a committee. H.R. 3407 would amend Chapter 301 of Subtitle VI of title 49 United States Code to insert a new section (as added by section 4) after Section 30129 as a new Section 30130 Cybersecurity of Automated Driving Systems. A manufacture cannot sell, introduce or deliver any HAV that performs partial driving automation or automated driving unless it has a cybersecurity plan that includes a written policy on detection and response. The policy must include process for identifying, measuring and responding to attacks, processes for preventative and corrective actions to mitigate against vulnerabilities that includes incident response and intrusion detection. The plan must identify a company point of contact, processes for limiting access to automated driving systems, employee training and processes to restrict employee access. H.R. 3407 also includes the provision seen in H.R. 3416 regarding SAE definition revisions.

In July 2017 the house introduced H.R. 3430 Highly Automated Information Sharing Advisory Council (Shares Act) that would establish a council to make recommendations on the development of a framework to allow manufacturers of HVS to share information relating to testing and deployment. The act as at writing this chapter was still referred to a subcommittee (U.S.C. 2017 (h).).

H.R. 3405 was introduced on July 28, 20017 to the Subcommittee on digitation commerce and protection (U.S.C. 2017 (i)). It is called the MORE Act. It would expand the exemption from motor vehicle safety standards for testing or evaluation purposes to cover manufacturers of highly automated vehicles and automated driving system components, and for other purposes. The bill has not moved out of committee.

H.R. 3412 was introduced on July 28, 2017 to the Subcommittee on Digital Commerce and Consumer Protection as is called the LEAD'R Act (USC 2017 (c)). H.R. 3412. It would amend Section 3013 of Title 49 United States Code to establish sole authority for NHTSA over regulation of HAVs. The bill would preempt state and other local jurisdictions from creating, maintaining or enforcing any law/regulation that prescribes the design, construction, or performance of highly

automated vehicles, automated driving systems, or components of automated driving systems unless such law or regulation is identical to a standard prescribed under this chapter.

On July 28, 2017, the U.S. House introduced H.R. 3411 to amend chapter 301 of subtitle VI of title 49, United States Code, to update or provide new motor vehicle safety standards for highly automated vehicles, and for other purposes (U.S.C. 2017 (j)). The bill proposes an automated driving system cybersecurity council that will be convened within six months of the bill's enactment. Set by the Secretary, representation will be diverse and capped at 30 members. In the same fashion as HR 3416, this bill notes that if SAE revises definitions, it will notify the Secretary who is required to publish these within the federal register for comment. The secretary will then notify SAE that if it has determined that the definition does not meet the need for motor vehicle safety or is otherwise inconsistent with United States Code, the existing Section 1 definition shall remain in effect. If the Secretary does not reject a definition revised by SAE it will amend regulations and standards as necessary (U.S. Congress, 2017 (j).

H.R. 3413 was introduced on July 28, 2017 and referred to the Subcommittee on Highways and Transit in the House Transportation and Infrastructure committees. The bill is titled the ACCESS Act – Addressing Community Challenges Emerging from Self-Driving Systems. This bill directs NHTSA to create an Advisory Council on Improving Mobility Access for Underserved Populations and Senior Citizens. The council shall undertake information gathering activities, develop technical advice, and present best practices or recommendations to DOT regarding mobility access for senior citizens and populations underserved by traditional public transportation services and educational outreach efforts with respect to the testing and distribution of highly automated vehicles in commerce. Members of the Council will serve for a term of three years, and shall include a diverse group that represents, the private and public sector, nonprofits, safety and consumer advocates, labor organizations, environmental experts and engineers. The council shall be comprised of no less than 15 and not more than 30 members appointed by the Secretary (U.S.C. 2017 (l)).

H.R. 3408 would amend section 30113 of title 49, United States Code to establish new exemptions for motor vehicle safety standards, and for other purposes. The amendments would make easier the development or field evaluation of— a feature of a highly automated vehicle providing a safety level at least equal to the safety level of the standard for which exemption is sought; or a HAV providing an overall safety level at least equal to the overall safety level of nonexempt vehicles (U.S.C., 2017 (m).

2.1.2. Other Bills

- H.R. 2778
- S.1809
- H.R. 3901

- H.R. 346
- H.R. 3421
- H.R. 4625
- S. 2217

2.2. Federal Agencies

2.2.1. National Highway Traffic Safety Administration (NHTSA)

The most important development from the standpoint of Texas is NHTSA's September 2016 Policy on Automated Vehicles (NHTSA, 2016(a)), its October 2016 *Policy on Best Practices for Cyber Security Best Practices in Modern Vehicles* (NHTSA, 2016 (b)), and the September 2017 *Automated Driving Systems: A Vision for Safety* policy (NHTSA, 2017 (a)), which replaced the previous 2016 policy.

The September 2016 policy publication, which was deliberately issued as guidance and not as rulemaking to facilitate best-practice guidance within a preliminary framework, set out the roles and responsibilities for NHTSA and the states. The guidance also set out the USDOT's expectations of industry for the immediate short term to test and deploy HAVs. Unlike the 2013 policy, this new 2016 policy is aimed at Level 3-and-above vehicles (under SAE J3016 definitions). A vehicle performance section set out best practices for safe pre-deployment, design, development, and testing of HAVs, and defined deployment as the operation of a HAV by members of the public who were not agents or employees of the designer, developer, or manufacturer of the HAV (NHTSA, 2016 (a) p. 8). The September 2016 policy also confirmed the model state policy articulated in the 2013 NHTSA policy: state responsibilities will include licensing of drivers (human) and motor vehicle registration, law and traffic enforcement, inspections, and motor vehicle liability and insurance rules. NHTSA noted that this was to ensure the creation of a consistent national framework, rather than a patchwork of laws that could be incompatible with one another.

In October 2016 NHTSA (NHTSA, 2016 (b)) released a new policy on best practices for cybersecurity in modern vehicles. This policy, which covers all motor vehicles, recommended a layered approach to cybersecurity, with the goal to reduce the probability of a cyber-attack's success and diminish unauthorized access ramifications. NHTSA stated that the approach should

- Be built upon risk-based prioritization
- Provide for timely detection and rapid response
- Create methods to 'design-in' rapid recovery from an incident
- Institutionalize methods for adopting lessons learned.

NHTSA encouraged use of the ISO 2700 series of standards and other best practices used in other technology sectors for developing protocols and approaches (NHTSA, 2016 (b)). The cybersecurity policy also recommended "penetration testing and documenting," which has stages that employ qualified testers who were not involved in development and are incentivized to unearth vulnerabilities. In summary, NHTSA set out at pages 17–20 a series of fundamental vehicle cybersecurity protections that it recommended.

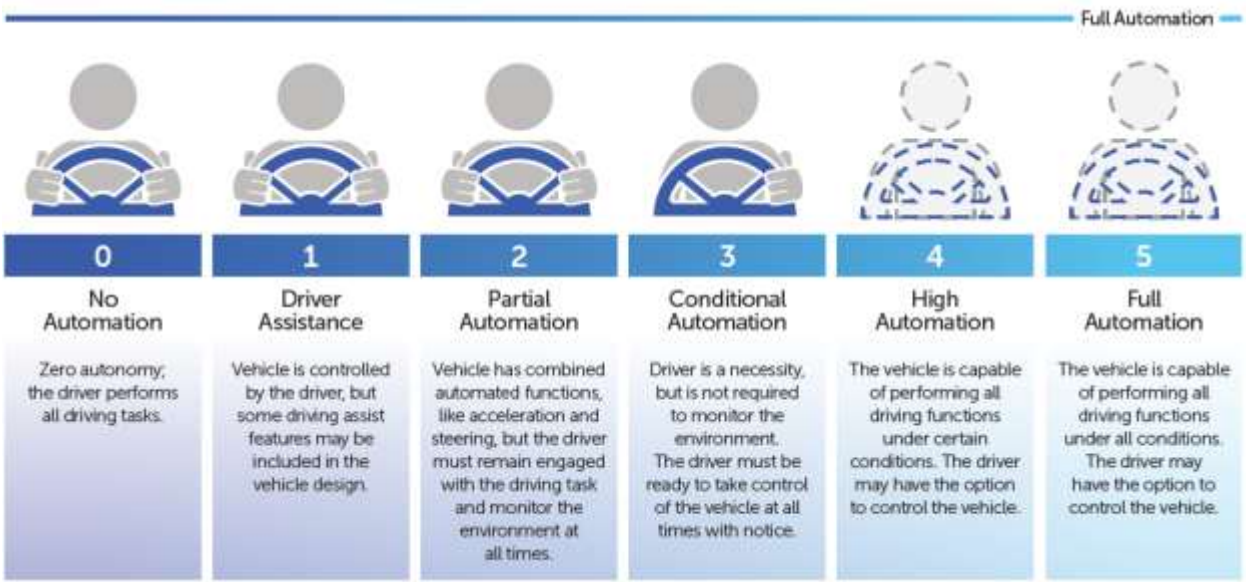
In September 2017 the new administration under Secretary Chao issued its policy for Automated Driving Systems. The policy was again introduced as policy and not as rulemaking.

This policy framework notes that it offers a path forward for safe deployment of automated vehicles by

- Encouraging new entrants and ideas that deliver safer vehicles
- Making the department's regulatory process nimble to help match the pace of private sector innovation
- Supporting industry innovation and encouraging communication with the public and stakeholders.

The policy also continues to adopt SAE's automation levels and created a new iconography to convey to the general public the various levels of automation (Figure 2.2).

As part of this new policy's development, the agency also responded to comments sought on the 2016 policy. The policy document is split into two sections. Section One has voluntary guidance that details ADS safety elements and ends with a voluntary safety self-assessment component. Section Two incorporates what it calls common safety-related components and significant elements regarding ADSs that states should consider incorporating into legislation.



Source: NHTSA 2017 (a)

Figure 2.2 NHTSA's SAE automation levels

Voluntary guidance for automated driving systems in Section One, contains twelve priority safety design elements for consideration. Table 2.1 provides brief descriptions of the ADS safety elements. The policy encourages entities that are engaged in testing and deployment to publicly disclose their voluntary safety self-assessments to demonstrate varied approaches to achieving safety. The difference between this requirement and the previous NHTSA 2016 policy is that entities will not be required to submit these safety assessments. These are now entirely voluntary. This includes some new elements such as fallback minimal risk condition, data recording, human machine interface, and post-crash ADS behavior.

Table 2.1 NHTSA's 2017 Policy ADS Safety Elements

Safety Element		Brief Descriptor
1	System safety	A robust design and validation process based on systems engineering approach to design ADSs free of unreasonable safety risks. Including a hazard and safety risk assessment for overall vehicle design integration. Design decisions should be linked to assessed risk that impact safety-critical system functionality.
2	Operational design domain	Define and document ODD for each ADS available on their system including: road types, geographic area, environmental conditions, speed range and domain constraints.
3	Object and event detection and response	Detection by driver or ADS circumstances relevant to immediate driving task and implementation of driver system response. Document process for assessment, testing and validation, crash avoidance and variety of behavioral competencies for ADSs.
4	Fallback minimal risk condition	Process for transitioning to a minimal risk condition when a problem is encountered and ADS cannot operate safely. At higher automation, where human driver is not available, ADS must fall back into minimal risk condition without driver intervention.
5	Validation methods	As scope, technology, and capabilities widen, entities are encouraged to develop validation methods to appropriately mitigate safety risks associated with ADS approach.
6	Human machine interface	At minimum the ADS should be capable of informing the human operator/occupant through indicators that the ADS is ADS functioning properly, i.e., is currently engaged or unavailable, experiences malfunction and/or requests control from ADS to the operator.
7	Vehicle cybersecurity	Encouraged to follow a robust product development process based on systems engineering approach to minimize safety risks due to cybersecurity threats and vulnerabilities. Documentation encouraged, including changes, design choices, analysis, and testing. Groups involved with ADSs should consider adopting a coordinated vulnerability reporting/disclosure policy
8	Crashworthiness	As vehicle mix may be operating (those with/without ADS), entities should consider scenarios of non-DS vehicle crashing into ADS-equipped vehicle and how to protect.
9	Post-crash ADS behavior	In testing or deployment consider how to return ADS to a safe state immediately after an incident, e.g., moving to a safe spot. If vehicle is a CV, communication with a relevant entity is encouraged to share and reduce harm resulting from a crash.
10	Data recording	Entities engaged in testing/deployment are encouraged to establish a process for data collection and validation to establish crash causes leading to fatalities/injuries. ADS data recommended to be stored and available for retrieval for crash reconstruction.
11	Consumer education and training	Develop, document, and maintain employee, dealer, distributor and consumer education and training programs to address anticipated differences in use and operation of ADS vehicles.
12	Federal state and local laws	Document how federal, state, and local traffic laws and updates will be integrated in vehicle design and ADSs.

Source: NHTSA, 2017

Section 2 incorporates notes common safety-related components and significant elements regarding ADSs that states should consider incorporating into legislation.

Section 2's technical assistance to the states notes that:

> "The purpose of this Voluntary Guidance is to help designers of ADSs analyze, identify, and resolve safety considerations prior to deployment using their own, industry, and other best practices. It outlines 12 safety elements, which the Agency believes represent the consensus across the industry, that are generally considered to be the most salient design aspects to consider and address when developing, testing, and deploying ADSs on public roadways. Within each safety design element, entities are encouraged to consider and document their use of industry standards, best practices, company policies, or other

> methods they have employed to provide for increased system safety in real-world conditions. The 12 safety design elements apply to both ADS original equipment and to replacement equipment or updates (including software updates/upgrades) to ADSs" (NHTSA, 2017).

A framework of best practices for highway safety officials is also provided within this section. This includes the traditional areas of licensing, registration and testing, liability and insurance and working with law enforcement. The policy highlights that as part of development best practices legislatures should:

- Provide a technology neutral environment
- Provide licensing and registration procedures
- Provide reporting and communications for public safety official
- Review traffic laws and regulations that could serve as barriers to operation of ADS.

2.2.2. Federal Trade Commission (FTC)

The FTC held a workshop on June 28, 2017 in which they examined consumer privacy and security issues posed by AVs (FTC, 2017). Workshop attendees, which included public and private sector stakeholders and consumer advocates, discussed these topics:

- the types of data vehicles with wireless interfaces collect, store, transmit, and share;
- potential benefits and challenges posed by such data collection;
- the privacy and security practices of vehicle manufacturers;
- the role of the FTC, NHTSA, and other government agencies regarding privacy and security issues related to CVs; and
- self-regulatory standards that might apply to privacy and security issues related to CVs (FTC, 2017a).

2.2.3. Federal Communications Commission (FCC)

On July 13, 2017 the FCC announced it had unlocked new airwaves for vehicular radar use (FCC, 2017b). According to the Commission's Press Release, *"The Commission's action expands the current 76-77 GHz spectrum allocation to include the entire 76-81 GHz band and transitions radars out of the 24 GHz band. This is consistent with the spectrum that is available internationally, avoiding the need to customize the radars in vehicles for different markets."* According to the FCC, access to this additional spectrum will enable innovation; allow these radar devices to better distinguish between objects in areas close to the vehicle; and improve performance for applications such as lane change warnings, blind spot detection, parking aids,

"stop and follow," "stop and go," autonomous braking, and pedestrian detection. The FCC Order also permits the use of this band for fixed and mobile radars at airports (2017). The order amends Amendment of Parts 1, 2, 15, 90 and 95 of the Commission's Rules to Permit Radar Services in the 76-81 GHz Band (FCC, 2017a).

2.2.4. Federal Motor Carrier Safety Administration (FMCSA)

The FMCSA held a public listening session on HAVs on April 24, 2017 (FMCSA, 2017). The public listening session was held in Atlanta, Georgia, and was accompanied by an online portal for stakeholders to listen in and to make comments. The notice in the federal register stated

> "FMCSA seeks information on issues that need to be addressed to ensure that the Federal safety regulations provide appropriate standards for the safe operation of HACVs from design and development through testing and deployment. Specifically, FMCSA welcomes comments and information on the application of the following regulatory provisions in title 49 CFR to HACVs: Part 383 (Commercial Driver's Licenses); part 391 (Qualifications of Drivers); sections 392.80 and 392.82 (use of electronic devices); part 395 (Hours of Service of Drivers); and part 396 (Inspection, Repair, and Maintenance). The FMCSA also requests public comments on how enforcement officials could identify CMVs capable of various levels of automated operation and the types of HACV equipment that can be effectively inspected at roadside" (FR Doc 2017-07723, 2017).

2.3. Texas Legislative Developments

The 85th Regular Texas Legislative Session saw a few bills pass out of committee with regard to CVs and AVs. All have been signed by Governor Greg Abbot. First, House Bill (HB) 1791 amends the Transportation Code (TC) regarding platooning vehicles that are using connected braking systems. Section 545.062 of TC is amended so that an operator of a vehicle that is equipped with a braking system, and is following another vehicle equipped with the same system, can use the first system to maintain an "assured clear distance or sufficient space as required by this section." Connected braking system is defined here as "a system by which the braking of one vehicle is electronically coordinated with the braking system of a following vehicle."

Senate Bill (SB) 2205 regarding automated vehicles amends TC Section 545 to add a new subchapter J. Definitions for automated motor vehicles include:

> *"Automated driving system" means hardware and software that, when installed on a motor vehicle and engaged, are collectively capable of performing, without any intervention or supervision by a human operator:*
>
> *(A) all aspects of the entire dynamic driving task for the vehicle on a sustained basis; and*
>
> *(B) any fallback maneuvers necessary to respond to a failure of the system.*
>
> *"Automated motor vehicle" means a motor vehicle on which an automated driving system is installed.*

"Entire dynamic driving task" means the operational and tactical aspects of operating a vehicle. The term:

(A) includes:

(i) operational aspects, including steering, braking, accelerating, and monitoring the vehicle and the roadway; and

(ii) tactical aspects, including responding to events, determining when to change lanes, turning, using signals, and other related actions; and

(B) does not include strategic aspects, including determining destinations or waypoints.

"Human operator" means a natural person in an automated motor vehicle who controls the entire dynamic driving task.

"Owner" has the meaning assigned by current TC at Section 502.001.[1]

The new chapter governs exclusively automated motor vehicles and automated driving systems, unless an exception is called out.[2] Political subdivisions of the state cannot impose either regulations, or franchise on the operation of an automated motor vehicle or automated driving system.[3]

When the automated driving system is installed on a motor vehicle or is engaged, the owner of the automated driving system is considered the *"operator of the automated motor vehicle solely for assessing compliance with applicable traffic or motor vehicle laws*." This is regardless of whether a person is physically present in the vehicle when the vehicle is operating and the automated driving system is considered to be licensed to operate the vehicle. A licensed human operator is not required to operate a motor vehicle if an automated driving system is installed and engaged[4].

Automated motor vehicles are authorized to operate in the state with the automated system engaged, with or without a human physically present in the vehicle.[5] However, an AV may not operate on a highway in the state with the automated system engaged unless the vehicle:

1) can operate in compliance with traffic and motor vehicle laws,
2) is equipped with a data recording device installed by the manufacturer of the vehicle or automated driving system,
3) has an automated driving system that complies with all federal laws and federal motor vehicle standards,
4) is registered and titled in accordance with current laws, and
5) has vehicle liability coverage or self-insurance required under current laws.[6]

[1] SB 2205 Section 545.451 subsections (1) through (5)
[2] SB 2205 Section 545.452 Subsection (a)
[3] SB 2205 Section 545.452 Subsection (b)
[4] SB 2205 Section 545.453
[5] SB 2205 Section 545.454 (a)
[6] SB 2205 Section 545.454 (b) (1) through (5)

Section 545.455 sets out the duties required following an accident with an AV. In the event of an accident that involves an AV, the automated motor vehicle (or the human operator of this vehicle) must comply with TC 550.[7]

Section 545.456 provides that an owner identified within TC 502.001 (31) may identify the vehicle as an AV or having an automated driving system to the Department of Motor Vehicles (DMV). The bill does not provide any specific authority for specific rulemaking by any state agencies (TxDOT, DMV, or DPS),

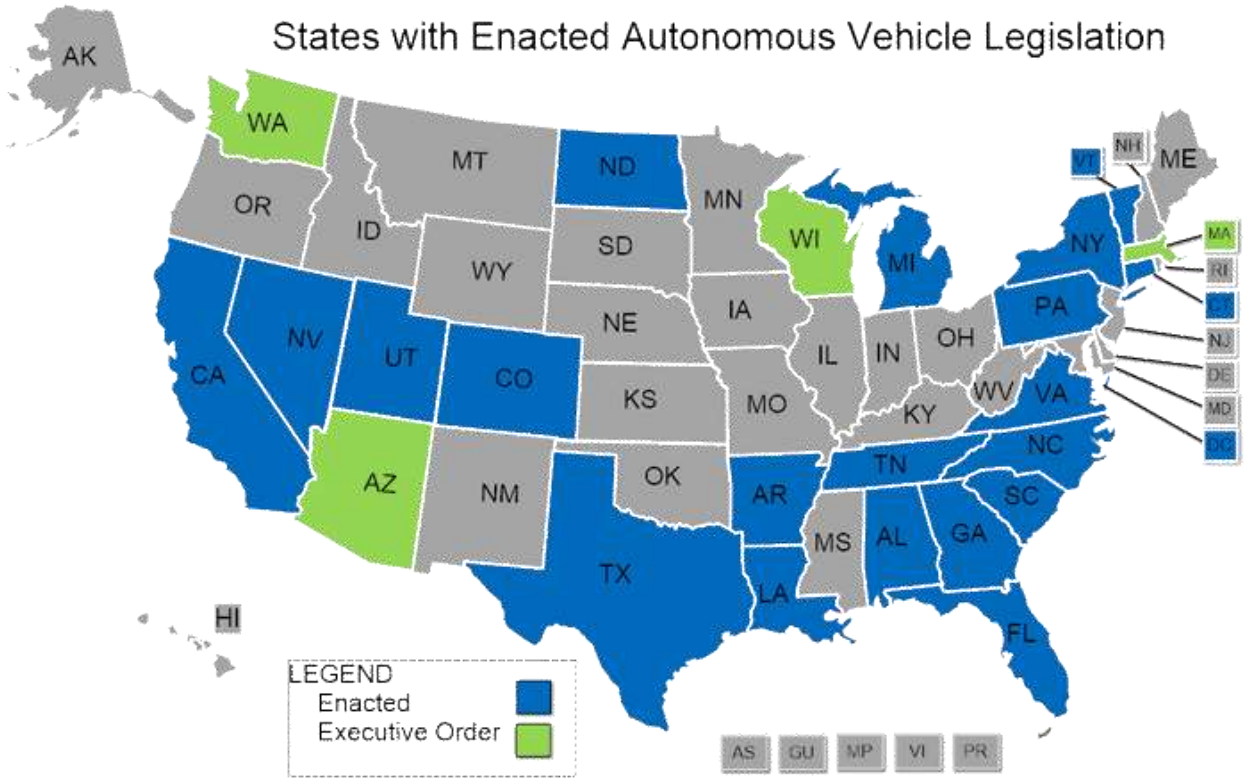
HB 8 relating to cybersecurity for state agencies requires State agencies, through an amendment to Government Code Section 2054.515, to conduct an information security assessment of the agency's information resources systems, network systems, digital data storage systems, digital data security measures, and information resources vulnerabilities at least once every two years. The Information Services Department may create by rule requirements for the type of information security assessment and report.

The Sunset review bill (HB 100) did not make any provisions for AVs.

2.4. Legal Developments within the States

In the United States, legal oversight of AV technologies has been initiated primarily at the state level. At the time of writing, 21 states have enacted legislation that governs the operation of C/AVs in the state, four states have executive orders, and there were over 80 bills in state houses across the U.S. during the spring 2017 legislative sessions. The NCSL is tracking the progress of bills in state legislatures and has a database of legislation; Figure 2.3 shows NCSL's current map of enacted legislation (NCSL, not dated).

[7] SB 2205 Section 545.455



Source: NCSL.org, as of August 2017

Figure 2.3 Map of states with enacted legislation for self-driving vehicles

2.4.1. States Developing Legislation

This section lists in alphabetical order the states that have developed and/or passed laws since the last review of state activity in late October 2016. We've indicated instances where a bill has not passed into law. Note in particular that Tennessee was added to this list, as they passed legislation in April 2016 that created a per-mile tax structure for AVs, and then amended and added to their legislation in 2017.

The section details activities that leading-edge states for AVs have undertaken, including amending their regulations on HAVs, CVs, platooning, and related matters.

2.4.1.1. Arizona

In February 2017 Arizona introduced HB 2434, which had amended language regarding handheld device prohibition. HB 2434 would have deemed a person not to be operating a motor vehicle if the motor vehicle is driven autonomously through the use of artificial intelligence software and the autonomous operation of the motor vehicle is authorized by law. A companion bill was also introduced in the Senate (SB 1135). Neither of these bills were enacted.

2.4.1.2. Arkansas

Arkansas enacted HB 1754 that regulates the testing of vehicles with autonomous technology and specifically added provisions regarding DATP systems and reduced the following distances of such systems.

The Act defines "*driver-assistive truck platooning system*" as technology that integrates sensor array, wireless communication, vehicle controls, and specialized software to synchronize acceleration and braking between two or more vehicles while leaving the designated vehicle's steering control and systems command in the control of its human operator. It additionally defines "*autonomous technology*" as technology installed on a motor vehicle that has the capability to drive the vehicle without the active physical control or monitoring by a human operator for any duration of time. Finally, "*autonomous vehicle*" is defined as a vehicle equipped with autonomous technology that can drive the vehicle without the active physical control or monitoring of a human operator for any duration of time.

The Act amended Arkansas Code §27-51-305 regarding following too closely to not prevent overtaking and passing of vehicles equipped with DATP systems. Under the Act at Section 1 (c) vehicles equipped with DATP systems may follow other vehicles closer than allowed under subsection (a) and (b) (1). These previously required a motor vehicle to follow not more closely than reasonable prudent having due regard for speed, and for a motor truck on a roadway outside of a business or residence district could not follow within 200 feet of another vehicle.

DATP is defined as technology that "integrates sensor array, wireless communication, vehicle controls, and specialized software to synchronize acceleration and braking between 2 or more vehicles while leaving and designated vehicle's steering control and systems monitoring in the control of its human operator."

Section 2 of the bill amends Arkansas Code Title 27, Chapter 51 at Subchapter 15, (§27-51-1408) to add an additional section that authorizes DATP truck platooning systems on a street or highway if a plan for general platoon operations is filed with the State Highway Commission. A person may operate a DATP system upon approval of the plan by the State Highway Commission, or 45 days after the plan is not rejected by the State Highway Commission. The bill was enacted as at April 1, 2017.

2.4.1.3. California

California enacted SB 1 in April 2017 (SB 1, April 28, 2017). As part of the transportation funding program, it created at Chapter 2 the Road Maintenance and Rehabilitation Program to address deferred maintenance on the state highway and local street and road systems. Chapter 2 §2030 §(d) provides that to the extent possible and cost effective, and where feasible, the department and cities and counties receiving funds under the program shall use advanced technologies and communications systems in transportation infrastructure that recognize and accommodate advanced automotive technologies that may include, but are not necessarily limited to, charging

or fueling opportunities for zero-emission vehicles, and provision of infrastructure-to-vehicle communications for transitional or full autonomous vehicle systems.

2.4.1.4. Colorado

Colorado's SB 213, effective on August 1, 2017, provides definitions for "automated driving system," "dynamic driving task," and "human operator." The Act notes that the use of motor vehicles with Level 0 through 3 automation as defined by SAE J3016 is legal under Colorado law with a human driver in the vehicle, and is not addressed in the Act.

"Automated driving system" is defined as hardware and software that are collectively capable, without intervention or supervision by a human operator, of performing all aspects of the dynamic driving tasks for a vehicle on a part-time or full-time basis, described under J3016 as Levels 4 and 5. "*Dynamic driving task"* is defined to include all of the following:

I. Operational aspects, including steering, braking, accelerating, and monitoring the vehicle and the roadway;

II. Tactical aspects, including responding to events, determining when to change lanes, turning, using signals, and other related actions.

Dynamic driving task does not include strategic aspects of driving, including determining destinations or way points.

The Act allows a person to use an automated driving system to drive or control a function of a motor vehicle if the system is capable of complying with every state and federal law that applies to the function that the system is operating. If the vehicle cannot comply with every relevant state and federal law, it must be submitted for approval via vehicle testing. The department must submit a report on the testing of the automated driving systems by September 1, 2018. The Act preempts state agencies and local jurisdictions from adopting or enforcing a policy, rule, or ordinance that sets standards for an automated driving system different from standards set for a human driver.

2.4.1.5. Connecticut

In SB 260—which was enacted on June 27, 2017—Connecticut defined the terms "*fully autonomous vehicle*," "*automated driving system*," and "*operator*." The bill requires the development of a pilot program for up to four municipalities for the testing of fully autonomous vehicles on public roads in those municipalities. It specifies the requirements for testing, including having an operator seated in the driver's seat and providing proof of insurance of at least $5 million. A task force is to be established to study autonomous vehicles. The study must include an evaluation of NHTSA's standards regarding state responsibility for regulating AVs; an evaluation of laws, legislation, and regulations in other states; recommendations on how Connecticut should legislate and regulate AVs; and an evaluation of the pilot program.

2.4.1.6. Georgia

In HB 472—enacted on May 9, 2017—Georgia provides for an exception for following requirements for vehicles following in a procession when speeds of the non-leading, participating vehicles are coordinated automatically and repealed conflicting laws. HB 472 specifies that the law prohibiting following too closely does not apply to the non-leading vehicle in a coordinated platoon. It defines coordinated platoon as a group of motor vehicles traveling in the same lane utilizing vehicle-to-vehicle communication technology to automatically coordinate the movement of the vehicles.

2.4.1.7. Louisiana

Louisiana's HB 511 (from the 2017 regular session) regulates and provides for the operation of AVs, including definitions, insurance requirements, registration/title, accident reporting, and additional/related matters and rules. The bill was left pending in committee.

The bill provides definitions for "automated driving system," dynamic driving task, fully autonomous vehicle, human driver, minimal risk condition (which is defined as a low-risk operating mode in which a fully autonomous vehicle operating without a human driver achieves a reasonably safe state, such as bringing the vehicle to a complete stop upon experiencing a failure of the vehicle's automated driving system that renders the vehicle unable to perform the entire dynamic driving task), on-demand autonomous vehicle network, and operational design domain. The bill then specifies under which specific conditions fully autonomous vehicles may be operated on public roads. These conditions include:

> (1) if the failure of the automated riving systems occurs that renders the system unable to perform the entire dynamic driving task relevant to its intended operation design domain, the vehicle will achieve a minimal risk condition…
>
> (2) the AV must be capable of complying with traffic and motor laws of Louisiana, and
>
> (3) the AV must bear the manufacturer's certification label indicating that it was in compliance with federal law at the time of its manufacture.

The bill additionally provides for insurance requirements; the AV must be insured in compliance with current statues and regulations at (R.S. 32:861 and R.S. 32:900) and a person must submit proof of financial responsibility for this to the Department of Public Safety and Office of Motor Vehicles. Any accident with an AV must be reported in accordance with 32:871. A person may operate an on-demand AV network, and provide transportation for multiple passengers. Additionally, registration and title of an AV must be completed in accordance with general registration and title laws. Finally, this bill would be exclusively governed by said chapter of the law.

2.4.1.8. Nebraska

Nebraska's LB 627 (which was left pending in committee) provides for operation of AVs and harmonizes with previous motor vehicle laws. The bill provides definitions for autonomous motor

vehicle, and autonomous technology. The bill provides that a person is considered the operator of an AV when they cause the technology to engage. A person may operate an AV when they have a valid operator's license.

Per safety of the vehicle, the AV must meet all federal and state regulations. In addition, the AV must have a safety alert system that alerts the operator to technology failure, at which point the operator must take control of the AV or bring the AV to a stop in the event that they cannot take control. The operator must also have a means of visually indicating when the AV is in autonomous mode. The bill further provides that prohibitions against using handheld written communications do not apply to a person operating an AV.

2.4.1.9. New Jersey

New Jersey's A3745 (which was reported out of Assembly Comm., with amendments on second reading in December 2016) would permit the testing and use of AV's under certain circumstances. The bill provides definitions for autonomous mode, autonomous technology, autonomous vehicle, commission, manufacturer, operator, and sensors.

The bill provides that an AV may be operated on public roads for testing purposes provided that (1) it is being operated solely by persons designated by the manufacturer, (2) the operator is inside the vehicle, capable of taking control, and (3) the manufacturer obtains insurance in the amount of $5 million.

Before public road operation, the manufacturer of an AV must apply for authorization by the commission. The application must contain certification that the AV may be disengaged from the operator, may visually indicated when it is in autonomous mode and contain a safety alert system of technology failure, at which point the operator will be required to take action to control or stop the vehicle. The manufacturer must provide certification that the AV has been tested on private property. The AV must also have the capability of recording and storing sensor data before and after a collision. Finally, the manufacturer must provide a disclosure to a purchaser of an AV of what personal information is collected by the technology of the AV.

2.4.1.10. North Dakota

North Dakota's HB 1202, which was enacted on April 13, 2017, provides for a Department of Transportation Study by creating a new section for Chapter 39-06 of the North Dakota Century Code. The Act provides that the North Dakota DOT shall collaborate with the AV technology industry to study the use of, and data collected by, AVs on state highways. The North Dakota DOT must review current laws of licensing, registration, insurance, and data ownership to be applied to AV use. North Dakota's DOT would report this study to the 66th legislative assembly of North Dakota.

2.4.1.11. Oklahoma

Oklahoma's SB 202 (which was left pending in committee in February 2017) provides for the Department of Public Safety to adopt regulations relating to autonomous vehicles. The bill would add a new section of law in Oklahoma States at Section 12-103 of Title 47. The act defines autonomous vehicles as a motor vehicle that uses artificial intelligence sensors and global system coordinates to drive itself without the active intervention of a human operator. It established that the Department of Public Safety would adopt regulations regarding operation of autonomous vehicles on the highway. These regulations include setting the minimum safety standards for AVs, as well as the requirements for operation and insurance, and providing for testing of AVs.

2.4.1.12. South Carolina

South Carolina's HB 3289, enacted on May 31, 2017, relates to the distance that must be maintained between vehicles traveling along a highway, and provides that this section does not apply to the operator of any non-leading vehicle traveling in a procession of vehicles if the speed of each vehicle is automatically coordinated.

The Act revised the term "driver" to "operator" in regard to these vehicles. At section (b) it notes that "the operator of a truck or motor vehicle that is drawing another vehicle traveling upon a roadway outside of a business or residence district and which is following another truck or motor vehicle drawing another vehicle shall, whenever conditions permit, leave sufficient space so that an overtaking vehicle may enter and occupy such space without danger, except that this shall not prevent a truck or motor vehicle drawing another vehicle from overtaking and passing any vehicle or combination of vehicles."

For motor vehicles operated upon roadway outside of a business or residence district in a caravan or motorcade—whether or not towing other vehicles—shall be operated as to allow sufficient space between each vehicle or combination of vehicles to enable any other vehicle to enter and occupy such space without danger.

This Act does not apply to the operator of any non-leading commercial motor vehicle subject to federal motor carrier safety regulations and traveling in a series of commercial vehicles using cooperative adaptive cruise control or any other automated driving technology.

2.4.1.13. Tennessee

Tennessee's SBN 1561, which was enacted in 2016 established certification program through its department of safety for manufacturers of AVs before such vehicles may be tested, operated, or sold. The law was enrolled and chaptered on April 27, 2016 at Pub.Ch 927. It also created a per mile tax structure for AVs (with a "use tax" that is in addition to the traditional gas tax). The Act distinguishes between a non-operator-required autonomous vehicle (NORAV) and an operator-required autonomous vehicle (ORAV).

A NORAV is defined an autonomous vehicle that may have operational controls for a human operator, including a steering wheel, accelerator, or brake, but does not require a human operator to be present in the vehicle during vehicle operation. There are two special license requirements for operators of NORAVs appropriate to the class of vehicle based on weight rating or number of passengers. An ORAV is defined as an autonomous vehicle equipped with operational controls for a human operator, including steering wheel, accelerator, and brake, and requires a human operator to be present in the vehicle for vehicle operation.

Tennessee's SB 2333, which was enrolled and chaptered on March 22, 2016, allows a motor vehicle to be equipped with an integrated electronic display visible to the operator while the motor vehicle's autonomous technology is engaged.

2.4.1.14. Washington

The governor signed an executive order in June 2017 (Washington State Governor: EO 17-02, 2017) to set up an autonomous vehicle work group and to begin to address autonomous vehicle testing and enabling pilot programs within the state. The working group is to have at least one representative from the Governor's office, and from other state agencies (that are listed). Pilot programs are authorized within the state in partnership with entities developing autonomous vehicle technology equipment. Pilot programs conducting testing and operation of autonomous vehicles with human operators physically present in the vehicle shall comply with these requirements:

> "Vehicles shall be operated or monitored only by a trained employee, contractor, or other person authorized by the entity developing autonomous technology.
>
> Vehicles shall be monitored, and an operator must have the ability to direct the vehicle's movement if assistance is required.
>
> Individuals able to exercise operational control of an autonomous vehicle during operation shall possess a valid U.S. driver license.
>
> Vehicle owners shall attest to proof of financial responsibility as required by RCW 46.30.020.
>
> Developing entities shall self-certify to DOL that they are compliant with the above requirements before beginning a pilot program."

In addition, the pilot programs that are conducting testing without a human operator present in the vehicle shall comply with these requirements:

> "Vehicles shall be equipped with an automated driving system that performs all aspects of the driving task on a part- or full-time basis within the vehicle's operational design limits, and it must be capable of bringing the vehicle to a safe condition in the event of a system failure.

> Vehicles shall be capable of being operated in compliance with Washington State motor vehicle laws relevant to the vehicle's operational design limits.
>
> Vehicle owners shall attest to proof of financial responsibility as required by RCW 46.30.020.
>
> Developing entities shall self-certify to DOL that they are compliant with the above requirements before beginning a pilot program."

2.4.1.15. Wisconsin

The Governor signed an executive order in May 2017 that will create a Steering Committee on Autonomous and Connected Vehicle Testing and Deployment (Wisconsin, 2017).

2.4.2. States Amending Their Regulations

States that could be considered pacesetting or leading-edge states, such as California, Nevada, Florida, and Michigan, have also begun to make amendments to current legislation and regulations.

2.4.2.1. California

California legislation and regulation provides oversight for AV testing and deployment. California DMV requirements for manufacturer testing include registering the AV with the DMV, completing previous AV testing under controlled conditions, using qualified test drivers who sit in the driver's seat with the ability to take control of the AV, and a $5 million insurance or surety bond maintained by the manufacturer (CA Vehicle Code 38570(A)(5)). Currently, in order to deploy a vehicle in California after testing, the vehicle must be approved by the California DMV.

In March 2017, the California DMV proposed regulations to amend Article 3.7 on Testing of Autonomous Vehicles, and to add in a new section Article 3.8 on Deployment of Autonomous Vehicles (California DMV, 2017 (a) and (b)). The department notes that it is adding Article 3.8 to specify the requirements to deploy autonomous vehicles, including the stipulation that these vehicles do not require a driver inside the vehicle. However, the deployment of these vehicles still necessitates obtaining a permit and certification by the manufacturer that it will maintain an instrument of insurance and that the vehicle meets any federal safety standards. The manufacturer also has to certify that the autonomous technology has a mechanism to engage and disengage the autonomous technology, and that a communication link with a remote operator allows continuous two-way communication with any passengers in the vehicle. Finally, the permit requires a process to display or communicate the vehicle owner or operator information in the event there is a collision or if there is a need to communicate with law enforcement.

In February 2018 California's Office of Administrative Law approved driverless testing regulations. The department posted a notice of this approval on its website and will begin approving applications after April 2, 2018 (California DMV, 2018b). California DMV's website had posted application requirements for the driverless autonomous vehicle tester program as of writing this report, including the required application forms to be submitted.

Within the definitions section of the testing of autonomous vehicles at Article 3.7 under Title 13, Division 1, Chapter 1, autonomous mode, autonomous test vehicle, autonomous test driver, convention mode, dynamic driving task, minimal risk condition, remote operator and operational design domain are specifically defined. The new regulations are linked to SAE's taxonomy under J3016 (California DMV, 2018 (a)). At section 227.041 manufacturers can conduct testing of autonomous vehicles on public roads. Proof of insurance is required to be kept in the test vehicles at all time, and these must be identified in writing to the department under Section 227.16 with make, model, year, vehicle identification number and license plate and state of issuance. Section 227.18 details that a driverless vehicle cannot be tested on a public road without a permit to conduct such testing issued by the department.

Manufacturers shall not test autonomous vehicles (including driverless autonomous vehicles) on public roads unless they have tested them under controlled conditions that have simulated, as closely as practicable, each operation design domain where the manufacturer intends the vehicles to operate on public roads. The manufacturer's testing permit for a driverless vehicle will be valid for two years (§227.23); the fee for a manufacturer's testing permit application is $3600 for processing (§227.22). For this permit, the manufacturer under Section 227.38 must certify that local authorities where the vehicle will be tested have provided written notification that contains all of the following:

(1) The operational design domain of the test vehicles
(2) A list of all public roads in the jurisdiction where the vehicles will be tested.
(3) The date that testing will begin.
(4) The days and times that testing will be conducted on public roads.
(5) The number of vehicles to be tested and the types of vehicles to be tested.
(6) Contact information, including name, telephone number, address, and email for contact person for the manufacturer conducting the testing.

In addition, the application must state that the manufacturer also complies with elements such as:

- Ensuring a communication link between the vehicle and remote operator to provide two-way communication and location information.
- Ensure that communication between the remote operator and any passengers occurs if the vehicle experiences any failures or would endanger the passengers' safety or other road users
- Descriptions of how the manufacturer will monitor the link, and also that the communication link will be continuously monitored.
- There is a process to display or communicate vehicle owner or operator information if the vehicle is involved in a collision and this information must be provided to a law enforcement officer for any reason (§227.38 (b (1-3)).
- The manufacturer certifies the vehicle is capable of operating without the presence of a driver and meets SAE's J3016 taxonomy descriptions for level 4 or 5 operating system ((§227.38 (c)).
- The vehicle complies with all federal motor vehicle safety standards (§227.38 (d)).
- Manufacturer provides a copy of a law enforcement interaction plan (§227.38 (e)).

- Manufacture maintains a training program for its remote operators (§227.38 (f)).
- Under Section 227.42, a testing permit can be suspended or revoked and there is a mechanism to appeal this within the regulations.
- All collisions must be reported under Section 227.48, as well as disengagement of autonomous mode under §227.50.
- The vehicle must be licensed and titled, and transfer of title can only be conducted by manufacturers under §227.54.

California has also now authorized the deployment of autonomous vehicles. Under new Modified Express Terms within Title 13, Division 1, Chapter 1 a new Article 3.8 Deployment of Autonomous Vehicles is also added to the DMV's adopted regulatory text (California DMV 2018 (a) at page 20). Manufacturers cannot deploy an autonomous vehicle unless it submits, and receives approval for, an application for a permit to deploy autonomous vehicles on public streets. The permit application is on form OL 321 and has a fee of $3,275 for application processing (§228.06 (a) (4)). The manufacturer shall certify in the application that the autonomous vehicles are equipped with an autonomous technology data recorder that captures and stores autonomous technology sensor data for all vehicle functions that are controlled by the autonomous technology at least 30 seconds before a collision with another vehicle, person, or other object while the vehicle is operating in autonomous mode. The data captured and stored by the autonomous technology data recorder, in a read only format, must be capable of being accessed and retrieved by a commercially available tool (§228.06 (a) (6)). Under (§228.06 (a) (8) the manufacturer must certify it meets all federal motor vehicle safety standards, and under (§228.06 (a) (9)) the manufacturer must certify that it can detect and respond to roadway situations in compliance with California Vehicle Code and local regulations that are applicable to the dynamic driving task in the vehicle's operational design domain, except when necessary to enhance the safety or the vehicle's occupants or other road users.

Under (§228.06 (a) (10)) the manufacturer must certify that the autonomous vehicles meet appropriate and applicable current industry standards to help defend against, detect, and respond to cyber-attacks, unauthorized intrusions, or false vehicle control commands. Under subsection 11 they manufacturer must certify it has conducted test and validation methods and is satisfied, based on the results of the tests and validations, that the vehicles are safe for deployment on public roads in California.

Section (§228.06 (b) requires that the manufacturer also certify that:

(1) A communication link between the vehicle and the remote operator, if any, to provide information on the vehicle's location and status and allow two-way communication between the remote operator and any passengers, if applicable, should the vehicle experience any failures that would endanger the safety of the vehicle's passengers or other road users while operating without a driver.

(2) The ability to display or transfer vehicle owner or operator information as specified in Vehicle Code section 16025 in the event that the vehicle is involved in a crash, and collision, or accident or if there is a need to provide that information to a law enforcement officer for any reason

(3) For any vehicle that is not equipped with manual controls for the completing the dynamic driving task, that it complies with all applicable federal motor vehicle safety standards and has an exemption approved by NHTSA.

2.4.2.2. Michigan

Michigan initially allowed C/AV testing so long as the vehicle is operated by an authorized agent of the manufacturer, and an individual is present in the vehicle and able to take control immediately if necessary. The State did specifically ban operation of AVs for non-testing purposes (Mich. Comp. Laws §§ 257.663, 665) as of 2015. However, in December 2016 Michigan passed legislation (SBs 995, 996, 997, and 998) that authorized driverless cars to be driven for any of the following purposes: personal use; road testing; as part of a SAVE program or "on-demand automated vehicle network;" and as part of a platoon. For example, under SB 995-998, the list of eligible drivers will expand to include people driving for personal use, university researchers who are conducting road testing, and Michigan DOT employees who are conducting road-testing. This means HAVs, or driverless cars, will operate without a human driver or any human present in the car. The news laws authorize operation of an automated motor vehicle without any control or monitoring by a human operator.

Michigan also had SB 927 drafted and introduced in 2016. This would amend Public Act 53 of 1979, which prohibits access to computers, computer systems, and computer networks for certain fraudulent purposes, to prohibit a person from intentionally gaining access, or causing access to be made, to an electronic system of a motor vehicle in order to willfully destroy, damage, impair, alter, or gain unauthorized control of the vehicle. It was referred to the Committee on Communications and Technology in November 2016 and was not introduced in the 2017 legislative session.

2.4.2.3. Tennessee

In May 2017 Tennessee enacted SB 151, which moved the state beyond the existing statute enacted in 2016. SB 151 establishes requirements for AVs to operate on public roads and highways. It defines an automated driving system (ADS) and authorizes motor vehicle manufacturers to commence a SAVE project. SAVE is an initiative by a manufacturer that makes ADS-operated vehicles available to the public for operation on the public roads and highways as determined by the manufacturer. As outlined in SB 151, a SAVE project also includes making an on-demand ADS-operated vehicle network available to the public.[8] The Act establishes the following procedures for manufacturers to operate ADS-operated vehicles. [Note: in the following list,

[8] This bill was obviously crafted with input from transportation network companies and the major car manufacturers who are looking to bring AV fleets into major metro areas in the next 3 to 4 years.

asterisks denote subsections later amended by Tennessee's SB 676, as discussed later in this section.]

1) ***** Only motor vehicle manufacturers are eligible to participate in a SAVE project, and each manufacturer is responsible for the safe operation of its participating fleet. The manufacturer must submit a letter to the department of revenue that includes the geographical areas in which the fleet will operate and a certification that:

    (A) The vehicles in the fleet are owned or controlled by the manufacturer and are equipped with an automated driving system, automatic crash notification technology, and a data recording system that has the capability of recording the automated driving system's status and other vehicle attributes, including speed, direction, and location, during a specified time period before an accident;

    (B) The fleet complies with all applicable state and federal laws; and

    (C) Vehicles in the fleet are capable of being operated in compliance with applicable traffic and motor vehicle laws of this state;

2) ***** A manufacturer must maintain incident records and provide periodic summaries related to the safety of the fleet to the department of revenue, the transportation and safety committee of the senate, the transportation committee of the house of representatives, and the National Highway Traffic Safety Administration (NHTSA);

3) ***** Prior to commencement and during the operation of a SAVE project, the manufacturer must make a privacy statement publicly available that discloses its data-handling practices in connection with the fleet;

4) While the ADS is in control of the vehicle, the manufacturer will assume liability for incidents where the ADS is at fault. A manufacturer is immune from any liability for damages from any modification made to an ADS-operated vehicle or an ADS by another person without the manufacturer's consent; and

5) The department of revenue may charge the manufacturer a fee for the operation of a SAVE project, but the fee must not exceed an amount necessary to implement this bill.

Under this Act, it is an offense for any person to knowingly operate a motor vehicle on Tennessee public roads or highways without a human driver in the driver's seat of the vehicle and without satisfying the requirements of the Act. A violation will be a Class A misdemeanor. The Act prohibits the following persons from operating an ADS-operated vehicle:

(1) Any person who operates/has operated an ADS-operated vehicle on a public road or highway without satisfying the eligibility requirements of the applicable jurisdiction; and

(2) Any person who was cited or found by law enforcement, a court, a state agency, or other applicable governing body to have violated a statute or regulation requiring prior notification or authorization to operate a vehicle equipped with an ADS.

The Act prohibits political subdivisions, by ordinance, resolution, or any other means, from banning or regulating the use of an ADS-operated vehicle or SAVE project that is operating under the Act's authority and otherwise complies with all laws of the political subdivision. The Act revised various laws regarding motor vehicles to reflect the existence of ADS's, such as child seat belt laws and accident-reporting laws.

On May 4, 2017, the Tennessee Senate adopted SB 676, which made amendments to SB 151 sub-sections one through three (highlighted with asterisks [*****] in the earlier text detailing SB 151) and a fourth amendment regarding preemption of local control of ADS.

Amendment number one removed the requirements regarding the commencement of a SAVE project and all SAVE-project related provisions and instead provides that an ADS-operated vehicle may drive or operate on Tennessee streets and highways with the ADS engaged without a human driver physically present in the vehicle if the vehicle meets the following conditions:

1) Unless an exemption has been granted under applicable federal or state law, the vehicle is capable of being operated in compliance with applicable provisions of Tennessee's motor vehicle safety and traffic, and has been, at the time it was manufactured, certified by the manufacturer as being in compliance with applicable federal motor vehicle safety standards;
2) In the event of a failure of the automated driving system that renders that system unable to perform the entire dynamic driving task relevant to its intended operational design domain, the vehicle is capable of achieving a minimal risk condition;
3) Is registered, and if registered in Tennessee, the vehicle is identified on the registration as an ADS-operated vehicle; and
4) The manufacturer that owns the vehicle maintains primary automobile liability insurance providing at least $5 million for death, bodily injury, and property damage; or the non-manufacturer owner maintains primary automobile liability insurance providing at least $50,000 for death or bodily injury, per person; $100,000 for death or bodily injury, per incident; and $30,000 for property damage.

This amendment also:

(1) Removes references to ADS's having automatic crash notification technology and specifies, for purposes of this state's seat belt laws, that a passenger or human operator required to be restrained by a safety belt will be solely responsible for the passenger's or human operator's compliance with such requirement; and

(2) Adds that liability for accidents involving an ADS-operated vehicle will be determined in accordance with product liability law, common law, or other applicable federal or state law. When the ADS is fully engaged, operated reasonably and in compliance with manufacturer instructions and warnings, the ADS will be considered the driver or operator of the motor vehicle for purposes of determining:

(A) Liability of the vehicle owner or lessee for alleged personal injury, death, or property damage in an incident involving the ADS-operated vehicle; and

(B) Liability for non-conformance to applicable traffic or motor vehicle laws; and

(3) Requires that no later than February 1, 2021, the commissioner of safety and the commissioner of commerce and insurance submit a report to the transportation and safety committee of the senate and the transportation committee of the house. The report must make recommendations with appropriate rationale as to whether the insurance and bonding coverages and coverage amount requirements of this bill should be increased, decreased, extended, or otherwise amended.

Amendment number 2 rewrote SB 151's requirement that a manufacturer owner of an ADS-operated vehicle maintain primary automobile liability insurance providing at least $5 million for death, bodily injury, and property damage; and that a non-manufacturer owner maintain primary automobile liability insurance providing at least $50,000 for death or bodily injury, per person; $100,000 for death or bodily injury, per incident; and $30,000 for property damage. This amendment instead requires that such vehicles be:

(1) Covered by a single-limit primary automobile liability insurance policy that provides at least $5 million for death, bodily injury, and property damage and that satisfies the requirements of the law governing uninsured motor vehicle coverage;

(2) Covered by a surety bond executed and filed with the commissioner of safety in the amount of $5 million for death, bodily injury, and property damage; or

(3) Self-insured for at least $5 million for death, bodily injury, and property damage, by a person certified to be a self-insurer by the commissioner of safety.

This amendment will expire on July 1, 2022, unless re-enacted, extended, or amended prior to such date. The amendment states that it is the legislative intent that any such proposed legislation to re-enact, extend, or amend be referred to the transportation and safety and transportation committees of the Tennessee house and senate. The amendment requires that the Commissioners of Safety and Commerce and Insurance submit a joint report to the transportation and safety and transportation committees of the house and senate no later than February 1, 2021. The Commissioner's Joint Report is to make recommendations that shall include:

- the appropriate rationale for reenactment, extension, or amendment and any proposed legislation thereto,

- whether the insurance and bonding coverages and coverage amount requirements of this amendment should be increased, decreased, extended, or otherwise amended.

Amendment number 3 rewrote SB 151's requirements regarding primary automobile liability insurance coverage of ADS-operated vehicles and requires vehicles are covered by primary automobile liability insurance in at least $5 million per incident for death, bodily injury, and property damage, and the automobile liability insurance satisfies the requirements of the law governing uninsured motor vehicle coverage. Again, this provision will expire on July 1, 2021. This amendment also changed the dates for the report required in amendment two from February 1, 2021 to February 1, 2020, which aligns this with the date of the Commissioners' Joint Report.

A fourth amendment also extends this bill's prohibition on political subdivisions, by ordinance, resolution, or any other means, banning or regulating the use of an ADS-operated vehicle to include a motor vehicle operated at any level of autonomous technology. It specifies that for motor vehicles operated at any other level of autonomous technology, the motor vehicle and driver will be held to the same laws as conventionally operated motor vehicles, including the financial responsibility requirements, unless an exemption is specifically set out for a vehicle operated with any level of autonomy.

On May 5, 2017 the House substituted SB 151 for House Bill 381. It adopted amendment number four and passed SB 151 as amended. Amendment number four incorporated amendments one and three, and amendment two was rewritten by Senate amendment number three.

The SAVE project in the final enrolled bill was also made manufacturer-specific and has a network initiated by the manufacturer (§55-54-102).

In April 2017 Tennessee also enacted SB 676, which permits the operation of a platoon on streets and highways in the state after the person provides notification to the department of transportation and the department of safety. Vehicles are not caravan; and operator controls the lead vehicle. Platoon is defined as a group of individual motor vehicles that are traveling in a unified manner at electronically coordinated speeds.

2.5. Conclusions

As this chapter notes, there has been much statewide activity regarding C/AVs, with states such as California entering into new phases of regulation amendment and development. At the federal level, no bills have yet passed out of the U.S. Congress. NHTSA, FTC, FCC, and other federal agencies are continuing to develop regulations and opportunities within this area, and this is expected to continue as this new technology evolves.

One critical component that we note is still problematic is nomenclature/terminology. The review of draft legislation within U.S. state legislatures shows a patchwork of definitions using different nomenclature/terminology, with some referring to SAE-defined levels, and some not. Harmonization of terminology in this area will be helpful for future policymakers entering into the field, and for the general public in understanding how these vehicles work. The use of standardized nomenclature will reduce confusion for the general public in this emerging area, set clear definitions, make reciprocity and other interstate compact agreements easier to manage for multiple public agencies and the federal government, and most importantly, reduce litigation anomalies within circuits due to terminology definitions.

Another critical component that will need to be reviewed by TxDOT as well as other public agencies and local governments is the role of preemption. While NHTSA has outlined the traditional areas of federal and state roles, they may well be grey areas where jurisdictional authority may not always be clear between federal and state, and within the states themselves between the state, state agencies, and local jurisdictions.

Critical preemption issues also arise between the state and local levels. Some local jurisdictions are taking a keen interest in state legislative activities and expect to serve as partners as legislation and regulations are developed. State legislators may need to address this shared responsibility more directly since no one wants to see a patchwork of legislative and regulatory authorities emerge through a haphazard process. Indeed, across the United States state legislatures are beginning to preempt local jurisdictions from instigating competing regulations while preserving the flexibility to address critical local concerns that may emerge as these vehicles are introduced.

A final recommendation for TxDOT staffers—particularly those within the policy realm—is to continue to monitor activities by key agencies and states who have been at the forefront of this realm. Monitoring activities within NHTSA, FCC, California, Michigan, and Nevada would be prudent.

Chapter 3. Information Sharing for Connected and Autonomous Vehicles

3.1. Introduction

Connected vehicles (CVs) can communicate with their internal and external environments (Lu et al., 2014), and provide a two-way wireless communication environment enabling vehicle-to-vehicle and vehicle-to-infrastructure communications (Lee and Park, 2012). Thus, CVs and infrastructure can collect high-fidelity traffic data, such as vehicles' origins or destinations and their trajectories, and also share such information with other CVs and infrastructure managers. This chapter describes the impacts of information sharing on the routes chosen by CVs, and the resulting impacts on traffic. This adaptive routing carries both potential benefits (allowing drivers to avoid non-recurring congestion) as well as risks ("overreaction" to transmitted information can overload alternative routes). To model this, we consider a scenario where an incident happens in the network, increasing the travel costs on those affected roads, using a stochastic shortest path approach with recourse to model re-routing and diversion.

Stochastic shortest path problems with recourse were first mentioned in the early 1990s (Andreatta and Romeo, 1988; Bertsekas and Tsitsiklis, 1991; Polychronopoulos and Tsitsiklis, 1993). This research drew on similar concepts including shortest or optimal path in probability or stochastic network. These shortest path problems are defined on networks with random link costs, under two different sets of assumptions. First, that information on link cost values is accumulated as the network is being traversed, with the objective to find a policy that leads from an origin to destination node with minimal expected cost; and second, that link costs become known only after a path is chosen, so the objective is to find a path with minimal expected value of the link lengths. For the first class of problems, one should not look for a best path, but rather for an optimal policy, the rule for deciding where to go next given the currently available information. Miller-Hooks and Mahmassani (2000) extended the problem to time-varying networks, with their concept of least expected time paths comparing to the stochastic shortest paths. Link costs are random variables with probability distribution functions that vary with time. Two procedures are presented in the paper, the first procedure determines the a priori least expected time paths from all origins to a single destination for each departure time in the peak period, and the second procedure determines lower bounds on the expected times of these a priori least expected time paths.

Online shortest paths (OSP) and user equilibrium with recourse (UER) are related areas of work. OSP is the sub-problem to UER, in the same manner that the traditional shortest path forms the sub-problem to the static user equilibrium (UE) problem (Unnikrishnan and Waller, 2009). Recourse can be viewed as the opportunity for a decision-maker to reevaluate his or her remaining path at each node based on knowledge obtained en route (Waller and Ziliaskopoulos, 2002). The UER definition implies that all used routing policies will have equal and minimum expected cost. Unnikrishnan and Waller (2009) developed a convex mathematical program for static UE under uncertain link states and update their route choice in an online manner. Gao (2005) presented an

algorithm when link costs are time-dependent, stochastic and have general dependency, with the concepts of adaptive routing. This same terminology is used in optimal information location for adaptive routing (Boyles and Waller, 2011) and adaptive transit routing in stochastic time-dependent networks (Rambha et al., 2016).

Generally speaking, vehicles can get incident information one of two ways: from an operator, or merely by observing traffic conditions such as long queue, low travel speed and spill back (this kind of information may cause poor judgments). But for connected autonomous vehicles (CAVs), their perception of the incident state in the network may evolve while traveling in response to received information or from experiencing abnormal congestion on links affected by an incident. Consequently, the perception of the incident state will influence routing decisions and diversions when available. However, we only use the concept of incident perception of CAVs in this chapter; that is, we ignore the technological details about how the CAVs perceive the incident, but take the incident perception of CAVs as uncertain behaviors. Markov decision processes (MDP) are applied to model the routing behavior of a single CAV when it obtains information.

The following section presents a freeway instance for a single CAV and introduces the concept of an MDP. The generalized MDP formulation is provided, followed by numerical experiments that explore the impacts of CAVs' real-time congestion awareness on minimum expected-cost routing policies.

3.2. Motivating Example

This section demonstrates the model's capacity for information sharing and rerouting. To facilitate explanation, this model and notation are introduced in the context of a freeway corridor; the following section generalizes to an arbitrary network. For the purposes of this section, an incident may happen on the freeway, which would increase travel time on an affected road segment. However, upstream of the affected road segment, vehicles may divert into alternative routes to avoid congestion. As shown in Figure 3.1, the freeway includes four nodes and six links.

A node-state, denoted by (*j,Info,Inc*), includes three components: *j* is the node that a vehicle arrives at; *Info* indicates whether the vehicle perceives information about an incident (there are two ways that the vehicle could perceive incident information—receiving information from the system's side, or inferring the incident by its own observation); and *Inc* indicates whether an incident occurs in the first place. Let p be the probability of an incident and q the probability that a vehicle learns about the incident; in the example below, $p = 0.1$.

Thus, the state space is

$$\mathcal{X} = \{ (A,0,0), (A,0,1), (A,1,1), (B,0,0), (B,0,1), (B,1,1), (C,0,0), (C,0,1), (C,1,1), (D,0,0), (D,0,1), (D,1,1)\}$$

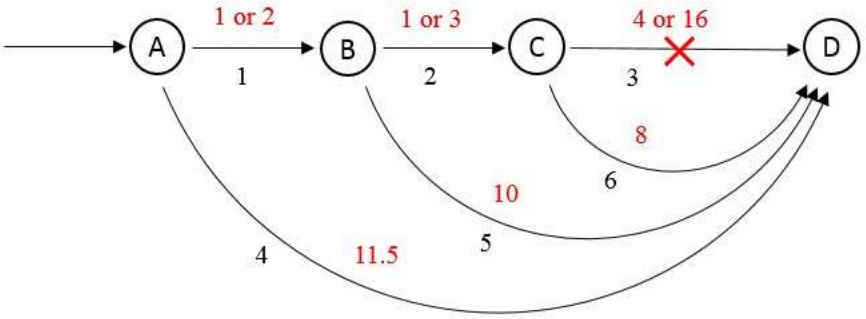


Figure 3.1 Freeway network with affected and alternative links (red numbers are link costs)

An action for a state, denoted by u, is the link a vehicle will travel next from this state. Action space for a state x = (*j,Info,Inc*), denoted by $U(x)$, contains all possible links that the vehicle may next travel based on this state. For instance, in Figure 3.1, if a vehicle makes a route choice at node *A*, then it can travel link 1 or link 4, regardless of the state at this node, so the action space for states (*A*,0,0), (*A*,0,1) and (*A*,1,1) is:

$$U(x = (A, 0, 0)) = U(x = (A, 0, 1)) = U(x = (A, 1, 1)) = \{1, 4\}$$

Similarly,

$$U\big(x = (B, 0, 0)\big) = U\big(x = (B, 0, 1)\big) = U\big(x = (B, 1, 1)\big) = \{2, 5\}$$
$$U\big(x = (C, 0, 0)\big) = U\big(x = (C, 0, 1)\big) = U\big(x = (C, 1, 1)\big) = \{3, 6\}$$

For the node *D* is the destination and the vehicle does not need to make route choices, so there is no action for states at node *D*.

Given a vehicle's current state and the action it will take, there are a set of probabilities for each of the next states that the vehicle may get to, and this set of probabilities are called the **transition probabilities** from current state to the next possible states. For example, if a vehicle arrives at node *A*, there are three possible states that the vehicle may get to, i.e., (*A,* 0, 0), (*A,* 0, 1) and (*A,* 1, 1). If the vehicle is in the state (*A*,0,0) and decides to travel link 1, there are three possible next states: (*B*,0,0), (*B*,0,1), and (*B*,1,1). If the next state is (*B*,0,0), which indicates that no incident happens, then the transition probability from (*A*,0,0) to (*B*,0,0) with the action of link 1 is given by,

$$P((B, 0, 0)|(A, 0, 0), 1) = 1 - p$$

Likewise,

$$P((B, 0, 1)|(A, 0, 0), 1) = p(1 - q)$$
$$P((B, 1, 1)|(A, 0, 0), 1) = pq$$

If the state is (*A*,0,1) and the vehicle decides to travel link 1, there are two possible next states: (*B*,0,1), (*B*,1,1). Then the transition probability is given by,

$$P((B, 0, 1)|(A, 0, 1), 1) = 1 - q$$
$$P((B, 1, 1)|(A, 0, 1), 1) = q$$

If the vehicle is in the state (A,1,1) and decides to travel link 1, according to the assumption, the next state can only be (B,1,1); thus the transition probability is one, i.e.:

$$P\big((B,1,1)\big|(A,1,1),1\big)=1$$

The probability distribution for transition from one state to the next remains the same for other states and actions.

Generally, the vehicle will get a reward after one-step transition from one state x to next state x' with an action u. Our MDP model replaces reward with cost, denoted by $c(x'|x,u)$, which gives one-step expected cost of transitioning from the current state to the next state based on an action.

A *policy* is a mapping from states to actions, which can be viewed as a rule to decide where to go next given the current available information. Optimal policy gives an optimal action, corresponding to one state when a vehicle travels to this state.

The objective of the MDP model for a single CAV is to minimize the sum of the expected cost for each state among all possible policies,

$$\min_{\pi}\mathbf{E}\left[\sum_{x,u}c(x,u)\middle|\pi\right]$$

where the expected cost for each state represents the expected cost from the node in this state to the destination.

3.2.1. Backward induction

MDPs can be solved by dynamic programming. Backward induction is one of the main methods for solving the Bellman equation in dynamic programming, which is defined by the equation below when the algorithm converges,

$$V_{i+1}(x):=\max_{u}\sum_{x'}P\,(x'|x,u)\cdot[-c(x'|x,u)+V_i(x')]$$

where i is the iteration number.

We apply backward induction to calculate the minimal total expected cost (which can be viewed as the Bellman equation in this model) in the objective function. Starting from the destination node D, determine the optimal policy and the corresponding minimal cost traveling from node C to node D.

Given current state (C, 0, 0)

$$
\begin{aligned}
c((C,0,0),3) = \;& P(no\_Inc)\cdot c((D,0,0)|(C,0,0),3) + \\
& P(Inc)\cdot P(perceived)\cdot c((D,1,1)|(C,0,0),3) + \\
& P(Inc)\cdot P(notPerceived)\cdot c((D,0,1)|(C,0,0),3) \\
= \;& 4(1-p)+16p(1-q)+16pq \\
= \;& 4\times 0.9+16\times 0.1 \\
= \;& 5.2 \\
c((C,0,0),6) = \;& P(no\_Inc)\cdot c((D,0,0)|(C,0,0),6) + \\
& P(Inc)\cdot P(perceived)\cdot c((D,1,1)|(C,0,0),6) + \\
& P(Inc)\cdot P(notPerceived)\cdot c((D,0,1)|(C,0,0),6) \\
= \;& 8(1-p)+8p(1-q)+8pq \\
= \;& 8
\end{aligned}
$$

Since $5.2 < 8$, $u^*_{(C,0,0)} = u^*(x=(C,0,0)) = 3$, the corresponding minimal expected cost is

$$c^*\big((C,0,0),u^*_{(C,0,0)}\big) = 5.2$$

If the current state is (*C*,1,1), then

$$
\begin{aligned}
c((C,1,1),3) = \;& c((D,1,1)|(C,1,1),3) = 16 \\
c((C,1,1),6) = \;& c((D,1,1)|(C,1,1),6) = 8
\end{aligned}
$$

Since 8 < 16, then u_((C,1,1))^*=u^* (x=(C,1,1))=6,

$$c^* ((C,1,1),u\_((C,1,1))^*)=8$$

The state (*C*,0,1) is special, for the vehicle does not perceive any information that an incident indeed happens in this state. The vehicle will follow the policy as if it were in the state (*C*,0,0). Thus, $u_{(C,0,1)} = u^*_{(C,0,0)} = 3$,

$$
\begin{aligned}
c\big((C,0,1),u^*_{(C,0,0)}\big) = \;& c((C,0,1),3) \\
= \;& P(perceived)\cdot c((D,1,1)|(C,0,1),3) + \\
& P(notPerceived)\cdot c((D,0,1)|(C,0,1),3) \\
= \;& 16q+16(1-q) \\
= \;& 16
\end{aligned}
$$

Next, we need to determine the optimal policy and corresponding minimal expected cost for node *B*. Given current state (*B*,0,0), the expected costs with different actions are given by,

$$
\begin{aligned}
c((B,0,0),2) = \;& P(no\_Inc)\cdot\big[c((C,0,0)|(B,0,0),2)+c^*\big((C,0,0),u^*_{(C,0,0)}\big)\big] + \\
& P(Inc)\cdot P(perceived)\cdot\big[c((C,1,1)|(B,0,0),2)+c^*\big((C,1,1),u^*_{(C,1,1)}\big)\big] + \\
& P(Inc)\cdot P(notPerceived)\cdot\big[c((C,0,1)|(B,0,0),2)+c\big((C,0,1),u^*_{(C,0,0)}\big)\big] \\
= \;& (1-p)\cdot[1+c^*((C,0,0),u^*)]+pq\cdot[3+c^*((C,1,1),u^*)] + \\
& p(1-q)\cdot\big[3+c\big((C,0,1),u^*_{(C,0,0)}\big)\big] \\
= \;& (1-p)(1+5.2)+pq(3+8)+p(1-q)(3+16) \\
= \;& 6.2+12.8p-8pq \\
= \;& 7.48-0.8q \quad (\because p=0.1)
\end{aligned}
$$

$$
\begin{aligned}
c((B,0,0),5) = \quad & P(no\_Inc)\cdot c((D,0,0)|(B,0,0),5) + \\
& P(Inc)\cdot P(perceived)\cdot c((D,1,1)|(B,0,0),5) + \\
& P(Inc)\cdot P(notPerceived)\cdot c((D,0,1)|(B,0,0),5) \\
= \quad & 10(1-p) + 10p(1-q) + 10pq \\
= \quad & 10
\end{aligned}
$$

Since $7.48 - 0.8q \le 7.48 < 10$, then $u^*_{(B,0,0)} = u^*(x = (B,0,0)) = 2$,

$$c^*\big((B,0,0), u^*_{(B,0,0)}\big) = 7.48 - 0.8q$$

Note that if we change the values of p and q, the optimal policy may turn into link 5.

If the current state is (B,1,1), then

$$
\begin{aligned}
c((B,1,1),2) = \quad & c((C,1,1)|(B,1,1),2) + c^*\big((C,1,1), u^*_{(C,1,1)}\big) = 3 + 8 = 11 \\
c((B,1,1),5) = \quad & c((D,1,1)|(B,1,1),5) = 10
\end{aligned}
$$

Since 8 < 11, then $u^*_{(B,1,1)} = u^*(x = (B,1,1)) = 5$,

$$c^*\big((B,1,1), u^*_{(B,1,1)}\big) = 10$$

If a vehicle is in the state (B,0,1), it will follow the optimal policy as if it were in the state (B,0,0). Thus, $u_{(B,0,1)} = u^*_{(B,0,0)} = 2$,

$$
\begin{aligned}
c\big((B,0,1), u^*_{(B,0,0)}\big) = \quad & c((B,0,1),2) \\
= \quad & P(perceived)\cdot\big[c((C,1,1)|(B,0,1),2) + c^*\big((C,1,1), u^*_{(C,1,1)}\big)\big] + \\
& P(notPerceived)\cdot\big[c((C,0,1)|(B,0,1),2) + c\big((C,0,1), u^*_{(C,0,0)}\big)\big] \\
= \quad & q(3+8) + (1-q)(3+16) \\
= \quad & 19 - 8q
\end{aligned}
$$

However, $c\big((B,0,1), u^*_{(B,0,0)}\big) = 19 - 8q \ge 11 > 10 = c\big((B,0,1), u^*_{(B,1,1)}\big)$, which indicates that when the incident indeed happens but the vehicle neither receives the incident information nor perceives the information at former states by itself, following the optimal policy as if the incident would not happen is never the optimal policy.

Finally, we determine the optimal policy and minimal expected cost for node A. Similarly, given current state (A,0,0), the expected costs with different actions when p = 0.1 are given by

$$
\begin{aligned}
c((A,0,0),1) = \;& P(no\_Inc)\cdot\left[c((B,0,0)|(A,0,0),1) + c^*\left((B,0,0),u^*_{(B,0,0)}\right)\right] + \\
& P(Inc)\cdot P(perceived)\cdot\left[c((B,1,1)|(A,0,0),1) + c^*\left((B,1,1),u^*_{(B,1,1)}\right)\right] + \\
& P(Inc)\cdot P(notPerceived)\cdot\left[c((B,0,1)|(A,0,0),1) + c\left((B,0,1),u^*_{(B,0,0)}\right)\right] \\
= \;& (1-p)\cdot[1 + c^*((B,0,0),u^*)] + pq\cdot[2 + c^*((B,1,1),u^*)] + \\
& p(1-q)\cdot\left[2 + c\left((B,0,1),u^*_{(B,0,0)}\right)\right] \\
= \;& (1-p)(1+7.48-0.8q) + pq(2+10) + p(1-q)(2+19-8q) \\
= \;& 0.8q^2 - 2.42q + 9.732 \\
c((A,0,0),4) = \;& P(no\_Inc)\cdot c((D,0,0)|(A,0,0),4) + \\
& P(Inc)\cdot P(perceived)\cdot c((D,1,1)|(A,0,0),4) + \\
& P(Inc)\cdot P(notPerceived)\cdot c((D,0,1)|(A,0,0),4) \\
= \;& 11.5(1-p) + 11.5p(1-q) + 11.5pq \\
= \;& 11.5
\end{aligned}
$$

Since $0.8q^2 - 2.42q + 9.732 \leq 9.732 < 11.5$, then $u^*_{(A,0,0)} = u^*(x = (A,0,0)) = 1$, and

$$c^*\left((A,0,0),u^*_{(A,0,0)}\right) = 0.8q^2 - 2.42q + 9.732$$

Note that if we change the values of p and q, the optimal policy may turn into link 4.

Moreover, if the current state is (A,1,1), then,

$$
\begin{aligned}
c((A,1,1),1) = \;& c((B,1,1)|(A,1,1),1) + c^*\left((B,1,1),u^*_{(B,1,1)}\right) = 2 + 10 = 12 \\
c((A,1,1),4) = \;& c((D,1,1)|(A,1,1),4) = 11.5
\end{aligned}
$$

Since 11.5 < 12, then $u^*_{(A,1,1)} = u^*(x = (A,1,1)) = 4$,

$$c^*\left((A,1,1),u^*_{(A,1,1)}\right) = 11.5$$

If a vehicle is in the state (A,0,1), it will follow the policy as if it were in the state (A,0,0). Thus, $u_{(A,0,1)} = u^*_{(A,0,0)} = 1$,

$$
\begin{aligned}
c\left((A,0,1),u^*_{(A,0,0)}\right) = \;& c((A,0,1),1) \\
= \;& P(perceived)\cdot\left[c((B,1,1)|(A,0,1),1) + c^*\left((B,1,1),u^*_{(B,1,1)}\right)\right] + \\
& P(notPerceived)\cdot\left[c((B,0,1)|(A,0,1),1) + c\left((B,0,1),u^*_{(B,0,0)}\right)\right] \\
= \;& q(2+10) + (1-q)(2+19-8q) \\
= \;& 8q^2 - 17q + 21
\end{aligned}
$$

However, $c\left((A,0,1),u^*_{(A,0,0)}\right) = 8q^2 - 17q + 21 \geq 12 > 11.5 = c\left((A,0,1),u^*_{(A,1,1)}\right)$.

3.2.2. Properties for Minimal Expected Costs

We will further discuss about the properties of minimal expected cost calculated by backward induction below.

1. Monotonicity of the expected costs with respect to *q*

Recall the expected cost for state (*B*,0,0) with action of link 2 and the expected cost for state (*A*,0,0) with the action of link 1,

$$c\big((B,0,0),2\big) = (-8p)q + (6.2 + 12.8p) \coloneqq c_1(q,p)$$

$$\begin{aligned} \mathrm{c}\big((A,0,0),1\big) &= (1-p)(1+7.48-0.8q) + pq(2+10) + p(1-q)(2+19-8q) \\ &= (8p)q2 - (0.8+16.2p)q + (12.5p + 8.48) \coloneqq c_2(q,p) \end{aligned}$$

Because

$$\partial c_1(q,p)/\partial q = -8p < 0,$$

$$\partial c_2(q,p)/\partial q = 16p - (0.8 + 16.2p) = -0.2p - 0.8 < 0$$

i.e., if we keep p unchanged, these two specific expected costs are all non-increasing functions with respect to q, the expected cost for individual state will decrease when q goes up. The conclusion still holds when $c\big((B,0,1),u^*_{(B,0,0)}\big)$ changes from $19 - 8q$ to 10.

2. Changes of optimal policies with different values of *p* and *q*

Recall the expected cost for state (*B,* 0*,* 0),

$$\begin{aligned} c((B,0,0),2) &= 6.2 + (12.8 - 8q)p \\ c((B,0,0),5) &= 10 \end{aligned}$$

when $6.2 + (12.8 - 8q)p < 10$, i.e., $q + 0.475/p > 1.6$,Because

$$\partial c_1(q,p)/\partial q = -8p < 0,$$

$$\partial c_2(q,p)/\partial q = 16p - (0.8 + 16.2p) = -0.2p - 0.8 < 0$$

i.e., if we keep p unchanged, these two specific expected costs are all non-increasing functions with respect to q, the expected cost for individual state will decrease when q goes up. The conclusion still holds when $c\big((B,0,1),u^*_{(B,0,0)}\big)$ changes from $19 - 8q$ to 10.

3. Changes of optimal policies with different values of *p* and *q*

Recall the expected cost for state (*B,* 0*,* 0),

$$\begin{aligned} c((B,0,0),2) &= 6.2 + (12.8 - 8q)p \\ c((B,0,0),5) &= 10 \end{aligned}$$

when $6.2 + (12.8 - 8q)p < 10$, i.e., $q + 0.475/p > 1.6$, the best action for this state is link 2, i.e., $u^*_{(B,0,0)} = u^*(x = (B,0,0)) = 2$, then according to the assumption, the action for state (*B,* 0*,* 1) is also link 2, so $u_{(B,0,1)} = u^*_{(B,0,0)} = 2$. Correspondingly, the expected costs are

$$\begin{aligned} c\big((B,0,1),u^*_{(B,0,0)}\big) &= 19 - 8q \\ c\big((A,0,0),1\big) &= (8p)q^2 - (0.8 + 16.2p)q + (12.52p + 8.48) \end{aligned}$$

In contrast, when $q + 0.475/p \leq 1.6$, the best action for this state is link 5, i.e.,

$$u^*_{(B,0,0)} = u^*(x = (B,0,0)) = 5,$$

then

$$\begin{aligned} u_{(B,0,1)} = u^*_{(B,1,1)} &= 5, \\ c\big((B,0,1), u^*_{(B,1,1)}\big) &= 10 < 19 - 8q \\ c((A,0,0),1) &= (1-p)(1+7.48-0.8q) + pq(2+10) + p(1-q)(2+10) \\ &= (0.8p - 0.8)q + (3.52p + 8.48) \\ &\leq (8p)q^2 - (0.8 + 16.2p)q + (12.52p + 8.48) \end{aligned}$$

Therefore, different values of p and q have varying impacts on optimal policy and further affect total expected cost.

3.3. Model Generalization

Consider a traffic network $\mathcal{G} = (N, A)$ with set of nodes N and set of links A.

Let $\mathcal{I}$ be the set of possible incidents, which represent decreases in capacity due to temporary events such as construction or vehicle collisions. Each incident $i \in \mathcal{I}$ increases travel times on one or more links. Denote by $\diamond \in \mathcal{I}$ the state of no known incident occurring. If a traveler believes the incident state is $\diamond$, then the traveler believes that no incidents are active and will choose routes accordingly.

Let $\tau_a(i)$ be the travel time on link $a \in A$ when the incident state is i. The travel time without any incidents is $\tau_a(\diamond)$.

We make the following assumptions:

1. Vehicles know ahead of time (*a priori),* the probabilistic description of the network, and could get personalized information (e.g., by Google Maps).
2. Once a vehicle gets the incident information from the operator, it has the knowledge from then on.
3. If the vehicle gets the incident information, then an incident has occurred (the information is reliable); otherwise, an incident may or may not have occurred.
4. If the vehicle neither receives any incident information from the operator, nor perceives any incident, it will follow the optimal policy as if there was no incident happening.

We further assume that incidents are mutually exclusive, i.e. if incident i occurs then incident $i' \neq i$ does not occur. Travelers are aware of this mutual exclusion. The assumption that incidents are mutually exclusive is not limiting; because any incident $i \in \mathcal{I}$ may affect multiple links, multiple distinct events may be coded as one "incident" in $\mathcal{I}$. However, note that including combinations of many distinct capacity reductions in $\mathcal{I}$ will greatly increase its' size. Therefore, it may be reasonable to restrict incidents in $\mathcal{I}$ to singular causes in capacity reductions. This is likely fairly realistic: the probability of multiple distinct causes of capacity reductions occurring

simultaneously is low, and furthermore travelers may not react to every distinct capacity reduction in their adaptive routing.

3.3.1. State Space

A state consists of the vehicle's location in the network as well as the perception of the incident state. Let $\mathcal{X} \triangleq N \times \mathcal{I}$ be the state space. A state $x(k) \triangleq (n(k), i(k))$ consists of a location $n(k) \in \mathcal{N}$ and the incident perception $i(k) \in \mathcal{I}$. Note that the step k is distinct from time. A step consists of traversing a link in the network.

The vehicle also has a destination, s. The states $(s,\cdot)$ are all termination states that the traveler will remain in after reaching one.

The network location is deterministic and controlled by the vehicle. The incident perception, however, is stochastic from the perspective of the vehicle. If an incident occurs, $i(k)$ may update if the system informs the vehicle. Both the occurrence of the incident, and whether a vehicle is informed, are stochastic. We assume that the system does not falsely inform vehicles of incidents. In other words, if the actual incident state is i_{net}, the vehicle will either receive information that the incident state is i_{net}, or that the incident state is $\Diamond$, but not anything else.

3.3.2. Action Space

At each location, the traveler has the option to proceed on any of the downstream links. Let $\Gamma_n^+ \subseteq$ denote the set of links outgoing from node n. Let $U(x)$ denote the action space when the state is $x = (n, i)$. If $n = s$, $U(x) = \{\mathcal{P}\}$, where $\mathcal{P}$ is the action to park or remain parked, because $(s,\cdot)$ is a termination state (the traveler's destination). Otherwise, $U(x) = \Gamma_n^+$. The traveler can choose any downstream links, and will be able to traverse that link deterministically.

3.3.3. Transition Function

As there are two components of the state, there are two components to the transition. The vehicle's location in the network is deterministic and depends entirely on the choice of action. On the other hand, the perception of the incident state is stochastic as it depends on information propagation. Let $f(x,u)$ define the next state when the state is $x = (n, i)$ and the action taken is u. $f(x,u)$ is defined in two components as

$$f(x,u) = (f_{\mathrm{N}}(x,u), f_{\mathcal{I}}(x,u))$$

The location transition is deterministic.

$$f_{\mathrm{N}}(x,u) = \begin{cases} s & \text{if } n = s \\ \gamma^+(u) & \text{else} \end{cases}$$

where $\gamma^+(a) \in N$ is the downstream end of link a. Recall that if u 6= P, then $u \in \Gamma_n^+$ is the downstream link.

The incident perception is stochastic. Let p_i be the probability that incident $i \neq \emptyset$ occurs. Let q_i be the probability that the system informs the vehicle that incident i is occurring. The probability of receiving information about i is $p_i q_i$. However, the vehicle can learn about i another way. If the vehicle enters an affected link, the higher travel times will be noticed and cause the vehicle to infer that incident i is occurring. Therefore, the transition in the incident perception is

$$f_j(x,u) = \begin{cases} \emptyset & \text{if } u = \mathcal{P} \\ i & \text{if } i \neq \emptyset \\ j & \text{w. p. } p_j q_j \text{ if } \tau_u(\emptyset) = \tau_u(j) \\ j & \text{w. p. } p_j \text{ if } \tau_u(\emptyset) \neq \tau_u(j) \\ \emptyset & \text{else} \end{cases}$$

The incident perception updates through observation because traveling through a link affected by an incident will be noticeable both to travelers and autonomous vehicles. Travelers will notice the congestion and may be able to visually identify the incident itself. Although AVs may not visually recognize the incident, they will recognize any discrepancy in the travel time from what is expected. Furthermore, they can compare the experienced travel time with those expected for each possible incident.

From a modeling standpoint, updating the incident perception with observation ensures that vehicle perceptions of travel times remain accurate.

3.3.4. One-step Costs

If $u = \mathcal{P}$, then there is not any associated cost. The cost of traveling along a link is the associated travel time. Let $c(x,u)$ be the cost when the state is $x = (n,i)$ and the action is u. $c(x,u)$ is defined as

$$c(x,u) = \begin{cases} 0 & \text{if } u = \mathcal{P} \\ \tau_u(i) & \text{else} \end{cases}$$

Based on this definition, $u = \mathcal{P}$ is the termination state. After reaching s, the cost-to-go is 0.

3.3.5. Cost-to-go and Solution Algorithm

This is a non-discounted infinite horizon MDP, which can be solved by value iteration (Bellman, 1957). Pseudocode for value iteration is shown below:

Algorithm 1 Value Iteration for MDP

Require: S, state space; $u(x)$, actions; P, transition probability matrix; C, cost functions; ϵ, the maximum error

Ensure: A utility function

repeat

$V' \leftarrow 0, V \leftarrow V', \delta \leftarrow 0$

for each state x in S **do**

for each action in $u(x)$ **do**

$V'(x) \leftarrow \max_u \sum_{x'} P\,(x'|x,u)[-c(x'|x,u) + V(x')]$

end for

if $|V(x) - V'(x)| > \delta$ **then**

$\delta \leftarrow |V(x) - V'(x)|$

end if

end for

until $\delta < \epsilon$

3.4. Numerical Experiments

In this section, we conduct experiments on two networks, representing the city of Sioux Falls, and downtown Austin, Texas. Numerical results are shown based on the following two aspects: the adaptive routing behavior of a single CAV across multiple origins and destinations in both networks, including average expected link costs and the impacts of incidents and CAVs' perception of incident on the minimum expected cost and optimal routing policy by changing the probability of an incident and CAVs' incident perception.

In addition to running base scenarios, we perform sensitivity analysis with respect to three parameters: the probability of incident occurrence, the incident severity, and CAVs' perception of incident information. This analysis focuses on the following questions:

1. how will a CAV make its routing decisions under different incident environments, such as regular congestion with high incident probability and car accidents with fairly low incident probability?
2. how does a CAV react facing different severity of incidents?
3. how does a CAV's capability of information gaining such as incident perception, impact its' routing behavior?

The results should somehow demonstrate the overall value of receiving information for CAVs.

Since incidents may happen at each location of the network, it is complicated to show the entire probabilistic description of each incident throughout the network. Thus, for convenience of presentation, we show results for a single incident instance, rather than probabilistically describing the incidents over the whole network. For example, in Sioux Falls network, we assume that link

46 is the only potential location for an incident, and will affect links 41, 57, 28, 32, 48, 25, 13, 21, by increasing their cost if an incident occurs.

3.4.1. Sioux Falls Network

Primarily, we explore the difference in expected travel times across multiple origins and destinations (Figure 3. 2). Based on the incident instance above, we arrive at a quick conclusion: the optimal policies remain the same for the state with the same node regardless of the probability of an incident and CAVs' perception of incident, if node 1 to node 6 is taken as the destination, because for these six nodes, the shortest paths do not include affected links.

In the base scenario, we assume that the probability of an incident is 0.1 and the probability of CAVs' incident perception takes a value of 0.6. In Sioux Falls network, there are 24 nodes. We take one node as the destination and other nodes as the origins at one time, then calculate the average travel times across 23 origin-destination pairs.

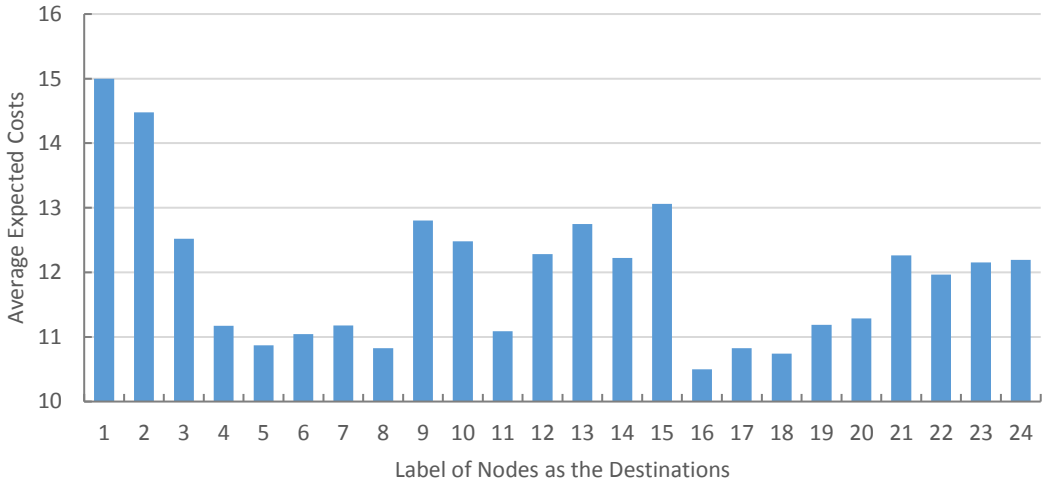


Figure 3.2 Average expected costs across multiple origins and destinations (p = 0.1, q = 0.6)

Secondly, we examine how much the expected costs depend on the probability of an incident. Starting from the base scenario, we keep the probability of incident perception for CAVs, q, constant, and change the incident probability, p. The average expected link costs for each destination and the increase of expected costs are shown in Figure 3.3:

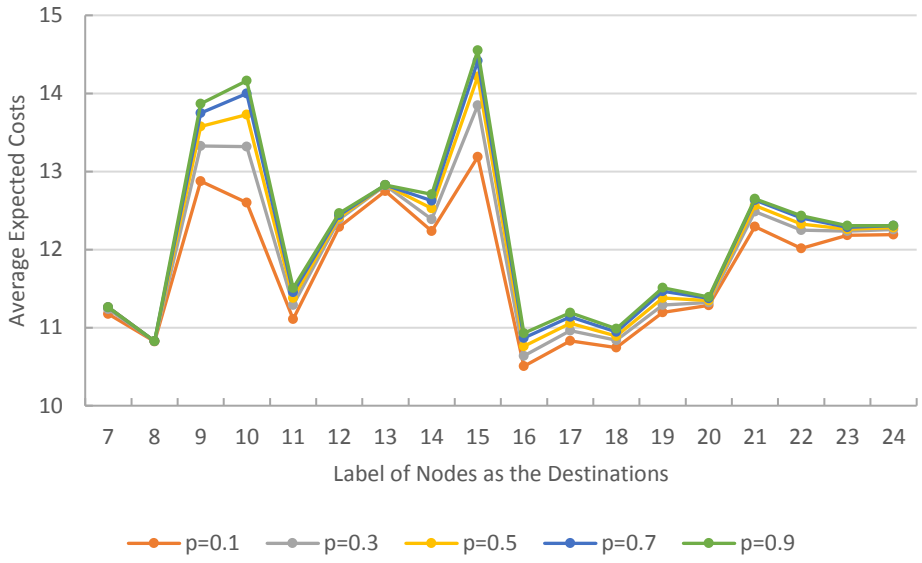


Figure 3.3 Average expected costs with different incident probabilities (q = 0.6)

The optimal policies and average expected costs remain the same for the state with the same node if node 1 to node 6 is taken as the destination, this is because the shortest paths do not include affected links for these six nodes.

Besides those seven nodes, the expected link costs will increase when the probability of an incident increases. In addition, the expected costs for incident-affected nodes as destinations are higher than those nodes which are not affected by the incident, and also the costs will increase more with the probability of the incident increases.

Thirdly, we explore how much the expected costs depend on the probability of CAVs' incident perception (Figure 3.4).

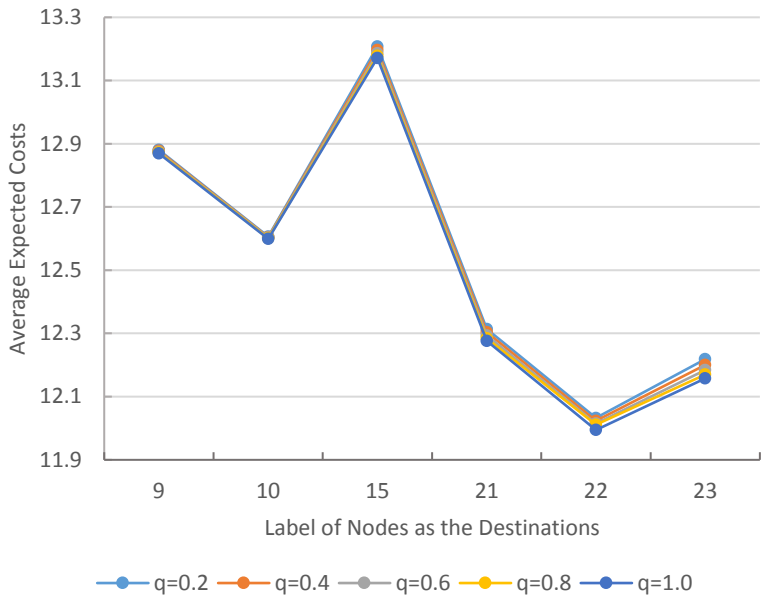


Figure 3.4 Average expected costs with different perception probabilities ($p = 0.1$)

The main results from this analysis are as follows:

1. Out of 24 nodes as destinations, only 7 destinations witness the slight decrease in average expected costs with the probability of CAVs' incident perception increasing. However, since the average expected costs merely decreased by 0.037% for node #19 as the destination, this node is not included in figure 4. For other nodes, the average expected cost remains invariant for all perception levels with the same destination.
2. The expected costs will decrease when CAVs' perception probability increases. This is simply because on average, the more incident information the vehicle gets, the wiser routing decisions it will take, which directly cause decreases in travel costs. Therefore, in spite of the fact that the minimal expected cost for an individual state increases monotonically with respect to q, higher perception probability will result in lower expected costs.
3. The average expected costs are more sensitive to change when the probability of the incident changes than when the probability of CAVs' perception of incidents changes.

3.4.2. Downtown Austin Network

We choose node #5469, which is located on the center of the downtown, as the destination to examine the extent to which both incident probability and CAV's perception probability have impact on the average expected costs.

Similar results are shown in Figure 3.5: Two average expected costs change in opposite directions for incident probability and CAV's perception probability. Along with increase of the incident

probability, the average expected costs would rise, but the rate of increase would lower. The average expected costs will decrease when the perception probability increases, with an exception of q = 0.8. The possible reason is that this value of q = 0.8 is a threshold for vehicle to choose optimal links, when q goes up, the vehicle will choose a link with lower expected cost when there is no incident and with high cost when there is an incident based on a certain state, then increase the expected cost of its previous states.

However, two rates of change are almost negligible, at 0.03% and -0.004%, respectively. So for a single CAV, both the incident and perception of incident have slight impacts on the expected travel times in Austin network.

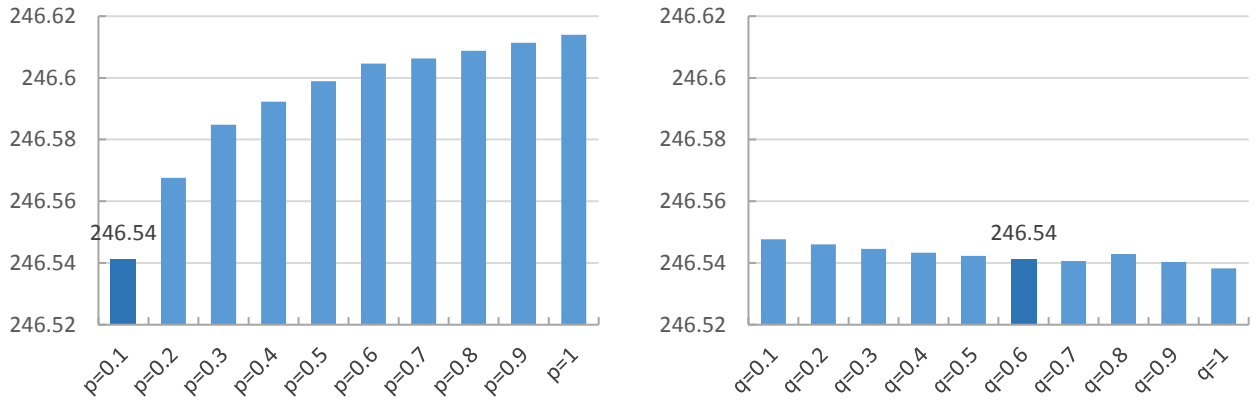


Figure 3.5 Average expected costs for downtown Austin. Left: Average expected costs with different incident probabilities (q = 0.6) ; Right: Average expected costs with different perception probabilities (p = 0.1)

Overall, for CAV's perception probability, the rate of change from the lowest to the highest is almost negligible, with the largest rate of 2% in Sioux Falls network and less than 0.02% in Austin network. Thus, we can conclude that the perception probabilities have minute impact on the expected costs for a single CAV. However, intuitively, the situation would be utterly different if it comes to the scenario with multiple CAVs.

3.5. Conclusions

Value iteration is employed to solve the non-discounted infinite horizon MDP. Similar results are shown in different networks in numerical experiments—if the vehicle gets the incident information, the best actions are always to travel the alternative routes to avoid the increased link cost. While for the uncertain states, without receiving incident information, the best actions are always to travel on the direct links.

Considerable future work remains. There are two possible extensions of the MDP model described in this chapter, including the activation and deactivation of incidents, which would get more stochastic link costs based on the time the vehicle receives the incident information, as well as the simulation for multiple CAVs, which would apply the UE model to solve the problems.

Chapter 4. Autonomous Intersection Management

This chapter presents Hybrid-AIM (H-AIM), an efficient intersection management protocol for early CAV penetration stages. H-AIM builds on the FCFS+Signals policy, which is part of the Autonomous Intersection Management (AIM) protocol developed by Dresner and Stone (2008). This chapter provides an overview of both AIM and the FCFS+Signals policy as well as surveying other relevant work.

4.1. Autonomous Intersection Management

AIM is a reservation-based protocol in which CAVs request to reserve trajectories crossing an intersection. The AIM protocol assumes that computer-controlled vehicles attempt to obtain the right to pass through the intersection by sending a reservation request message to the *intersection manager*. When using a "first come, first served" (FCFS) policy, the intersection manager approves reservation requests that do not conflict with any previously approved reservation or potential HVs. In brief, the protocol proceeds as follows (see Figure 4.1 for an illustration of the process).

A green signal is assigned to all northbound lanes while all other lanes are assigned a red signal. Green trajectories marked with solid or dashed green lines across the intersection. Active green trajectories marked only by dashed green lines.

Figure 4.1 Four-way intersection

1. An approaching CAV, *v*, sends a message to the intersection manager requesting a reservation. The *request-reservation message* contains data such as the vehicle's size, predicted arrival time, velocity, acceleration, and arrival and departure lanes.

2. The intersection manager processes the request message by simulating the trajectory of *v* through the intersection; the simulated trajectory is denoted by *path*(*v*).

3. If *path*(*v*) does not conflict with any previously approved reservations or potential HVs then the intersection manager issues a new reservation based on *path*(*v*) and sends an *approve message* containing the new reservation details back to *v*.

4. If *path*(*v*) does conflict with a previously approved reservations or potential HVs then the intersection manager sends a *reject message* to *v* which, after a predefined time period, may request a new reservation.

5. After receiving an approve message, it is the responsibility of v to arrive at, and travel through, the intersection as specified in $path(v)$ (within a range of error tolerance). A CAV may not enter the intersection unless it successfully obtained a reservation.

6. Upon leaving the intersection, the CAV informs the intersection manager that its passage through the intersection was successful.

The AIM protocol does not rely on communication capabilities between vehicles (V2V) but only between vehicles and the intersection manager (V2I). The protocol is robust to communication failures: if a message is lost, either by the intersection manager or by the CAV, the system's efficiency might be reduced, but safety is not compromised. Safety is guaranteed also when considering a mixed scenario where both HVs and CAVs are present. To address such a scenario, Dresner and Stone (2008) introduced the FCFS+Signals policy.

4.1.1. FCFS+Signals

Dresner and Stone's (2008) FCFS+Signals policy is a combination of AIM and traditional traffic signals. Whenever the traffic signal is green for a given lane, all vehicles arriving at that lane have the right to pass. However, when the traffic signal shows a red signal, only CAVs that were granted a reservation may drive through the intersection.

Since the protocol is not assumed to know the location and trajectory of HVs, such vehicles are assumed to occupy all trajectories that are approved by the traffic signal. In this report we define such trajectories as *green trajectories*.

> **Definition 1** (Green trajectories). A trajectory through the intersection is green if its incoming lane is assigned a green signal.

Figure 4.1 shows an example of green trajectories across an intersection (both the solid and dashed lines represent green trajectories). Note that green trajectories are dynamically changing; once the signal changes, the green trajectories will also change. The signal's timing is assumed to be known to the intersection manager, so it is able to predict green trajectories in advance.

The FCFS+Signals policy prohibits CAVs from obtaining reservations that conflict with green trajectories. In our example from Figure 4.1, all reservation requests will be automatically denied except those made by southbound or eastbound CAVs that are requesting to turn right.[9]

4.1.2. Experimental Results for AIM

Dresner and Stone (2008) reported *average delay* for a mixture of CAVs and HVs obtained from the AIM simulator running the FCFS+Signals policy.

> **Definition 2** (Delay). Delay is defined as the increase in travel time for a vehicle caused by red traffic signals or other vehicles. In other words, it is the difference between the

[9] This report assumes driving on the right side of the road. However, the ideas can trivially be generalized to a left-side driving policy.

vehicle's observed travel time and its theoretical travel time in free-flow conditions (no congestion) with full right-of-way (green signals).

For CAV penetration of 90% and below, FCFS+Signals yielded a mild improvement over traditional traffic signals. The improvement is attributed to CAVs that make right turns on red. If HVs are assumed to be able to turn right on red (as is common in the U.S.) or right-turning vehicles have a designated lane bypassing the intersection, then this policy would likely result in no improvement at all.

For CAV penetration greater than 90% the *one-lane signal policy* was suggested, which yielded a significant reduction in average delay. In the one-lane signal policy, the right to pass for HVs (i.e., green signal) is given to a single lane at a time instead of an entire road (all lanes arriving from the same direction). The one-lane signal policy results in a significant reduction in green trajectories at the cost of increased delay for HVs. As a result, the one-lane signal policy proved to be inefficient when considering lower CAV penetration (less than 90%).

4.1.3. Other Related Work

In recent years, different variants and enhancements to the basic AIM protocol were suggested. A line of work developed techniques for ordering reservation requests in ways that are more efficient than FCFS. The intuition behind such work is that, in some cases, approving a reservation that conflicts with several other requests is inefficient, even if that request was submitted first. Au et al. (2011) presented the notion of batch reservations where a batch of vehicles arriving from the same incoming road is granted the right to pass as a group. This approach was shown to be superior to FCFS for imbalanced intersections where an arterial road intersects with a low capacity road. Zhu et al. (2009) suggested a protocol named LICP that uses a look-ahead approach where the intersection manager optimizes reservation allocation within a defined moving time window (the look-ahead). LICP presented up to 25% reduction in average delay compared to the traditional FCFS approach. In contrast to our work, LICP assumes that all vehicles are connected and autonomous.

Market-inspired approaches for ordering reservation requests were also presented (Vasirani and Ossowski, 2009; Carlino et al., 2013). In this line of work, automated agents bid for the right to pass through intersections and auction mechanisms are used to determine the winners. Such auction-based approaches were shown to apply for traditional intersections using stop signs and traffic signals, as well as to intersection management protocols. These studies focus on fairness issues and network wide efficiency where vehicles travel through a network that is composed of several intersections. By contrast, our work focuses on minimizing delay and maximizing throughput in a single intersection.

Another line of work assumed that the intersection manager is able to control the speed of incoming vehicles (Lee and Park, 2012; Bento et al., 2012). Controlling the vehicle's speed allows the intersection manager to precisely coordinate the vehicles' time of arrival and crossing schedule in a way that dramatically reduces the vehicles' need to stop, which, in turn, reduces emissions and

delays. Work covering this approach usually assumes that all vehicles are connected and autonomous, and that the intersection manager is able to manipulate the speed and trajectory of incoming vehicles prior to entering the intersection. Our work makes none of these assumptions.

VanMiddlesworth et al. (2008) presented a protocol for coordinating CAVs through an *unmanaged* intersection. In the presented protocol, vehicles negotiate the right to pass amongst themselves. On the one hand, this protocol is cheap to implement as it doesn't require any road side equipment—on the other hand, it was shown to be less effective than AIM except in very low traffic volumes. To date, no version of this protocol that can handle a mixture of CAVs and HVs has been presented.

Bento et al. (2013) presented an intersection management protocol for mixed traffic named *legacy early method for intelligent traffic management* (LEMITM). When LEMITM detects an incoming HV, it computes an upper and lower bound for its arrival time. LEMITM then tries to reserve all possible trajectories through the intersection for the given time interval. If the reservation does not conflict with any previous reservation, it is approved and a designated traffic signal will turn green, indicating the right to pass for the HV. Similar to FCFS+signals, LEMITM was shown to be efficient only for high CAVs penetration rates ($\geq 90\%$). The focus of our work, by contrast, is early adoption stages where most of the traffic is composed of HVs.

4.1.4. Autonomous Intersection Management Protocol for Mixed Traffic

CAVs are expected to penetrate the automobile market gradually over many years. Reaching 90% AV penetration rates will probably not happen in the near future (Bansal and Kockelman, 2016). Hence, a new intersection management protocol is required for managing traffic that is comprised mostly of HVs.

4.1.5. Assumptions and Desiderata

When compared to traditional traffic signals, the new intersection management protocol should provide the following:

- Reduce the average delay experienced by vehicles crossing the intersection.
- Reduce queue length on incoming lanes. Once the vehicle queue is longer than the length of the incoming link, a phenomenon known as queue spillback occurs (Abu-Lebdeh and Benekohal, 1997).
- Prevent queue spillbacks, which have a negative cascading effect and should be avoided as much as possible (Liu and Chang, 2011).
- Increase throughput. Higher intersection throughput helps reduce congestion accumulated on links leading to the intersection.

- Provide a relative advantage to CAVs over HVs so as to incentivize drivers to transition to CAVs, which are assumed to be safer (Furda and Vlacic, 2011) and more efficient (Regele, 2008).

- Preserve safety guarantees. Similar to traditional traffic signals, the suggested protocol must guarantee that vehicles on conflicting trajectories are not given right-of-way simultaneously. This guarantee must hold also for cases of faulty communication and dropped messages.

The protocol presented in this report makes these same assumptions that were made by the original AIM protocol:

- CAVs can communicate with the intersection manager through a commonly known message protocol.

- A CAV may not enter the intersection without a fitting reservation.

- When crossing the intersection, a CAV precisely follows its reserved trajectory.

- An HV may not enter the intersection while its incoming lane is assigned a red signal (by a traditional traffic signal).

In addition to these assumptions, H-AIM also makes the following assumptions:

- Using a sensor (loop detector, camera, or radar), the intersection manager is able to detect approaching vehicles on each lane (sensing speed and heading is not assumed).

- A CAV may not pose as an HV. Even if a CAV is arriving on a lane with a green signal, it may not enter the intersection unless it follows an approved reservation.

- HVs may not change incoming lanes within sensing distance. That is, it is safe to assume that once an HV is detected on an incoming lane, it will occupy the same lane until it enters the intersection.

4.1.6. Hybrid AIM

Next, we present the Hybrid-AIM (H-AIM) protocol. Similar to FCFS+Signals, H-AIM grants reservations in a FCFS order. However, while FCFS+Signals automatically rejects reservation requests that conflict with green trajectories, H-AIM rejects reservation requests that conflict with *active green trajectories*.

> **Definition 3** (Active green trajectories). A green trajectory (see Definition 1) is active if an HV is present on it or on its incoming lane.

Figure 4.2 illustrates active green trajectories shown as dashed green lines across the intersection (notice vehicle 1 on the incoming lane). Active green trajectories are a subset of the green

trajectories making H-AIM at least as efficient as FCFS+Signals; there can be no reservation that is approved by FCFS+Signals and denied by H-AIM. However, the reverse is possible. As an example, consider the setting depicted in Figure 4.2. Assume vehicle 2 is a CAV and is heading north. Under the FCFS+Signals policy, vehicle 2 would be automatically denied a reservation as it crosses a green trajectory. H-AIM, on the other hand, would consider such a reservation as it doesn't cross an active green trajectory.

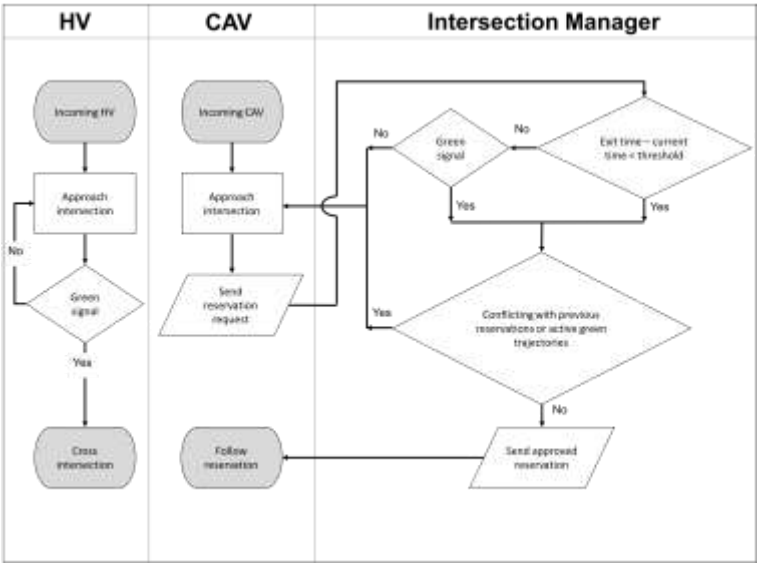


Figure 4.2 Flowchart presenting the working principle of H-AIM

Note that the existence of a CAV on an incoming lane does not create an active green trajectory. As a result, the system is required to be able to identify whether an approaching vehicle is of type CAV or HV. For doing so we suggest the following procedure:

1. Let v = the number of vehicles detected on a given lane, l.
2. Let r = the number of reservation requests from unique vehicles seeking to enter the intersection from lane l. Reservations are considered only if the specified exit time is greater than the current time.
3. If $v > r$ then assume a human vehicle on lane l.

Note that the above procedure is safe in the sense that it will never misidentify an HV as a CAV. In the case of faulty communication this procedure might misidentify a CAV as an HV but doing so does not pose a safety issue. It might, however, hurt efficiency since a green trajectory might, mistakenly, be considered active. Safety can be compromised, however, if HVs are allowed to change lanes in close proximity to the intersection. For this reason HVs must be prohibited from changing lanes within detection range.

Figure 4.3 illustrates the H-AIM protocol. As in traditional traffic signals, HVs approaching the intersection may cross it only if a green signal is given to their incoming lane. Else, they may continue to approach the intersection but may not cross it. An approaching CAV, on the other

hand, sends a reservation request to the intersection manager. The intersection manager checks if the reservation request's exit time minus the current time is larger than a given threshold. The threshold represents the minimal duration of time taken between the identification of an approaching HV and the time that the same HV reaches the intersection. Considering requests that are within the threshold guarantees that all potentially threatening HVs are identified. If the exit time is beyond the given threshold, the intersection manager inquires whether the reservation request's entrance time and entrance lane align with a green signal. If this is the case, it is still safe to consider the reservation, as no conflicting green trajectories can exist. Once the intersection manager determines that it is safe to consider the reservation request, it examines whether the request conflicts with any previously approved reservations or active green trajectories. If this is not the case, the reservation request is approved and an approve message is sent to the CAV, which, in turn, must precisely follow the reservation (or risk losing the right-of-way).

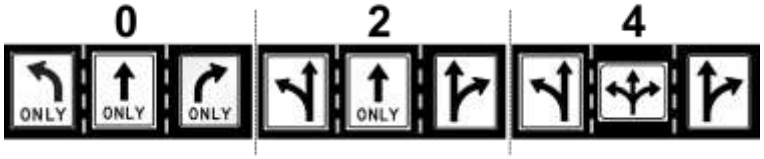


Figure 4.3 Three turning assignment policies for a three-lane road approaching a four-way intersection

4.2. Reducing the Number of Green Trajectories

Green trajectories (as a super-set of active green trajectories) can limit CAVs from obtaining reservations. As such, CAVs benefit from reducing the number of green trajectories to a minimum. On the other hand, HVs cannot cross the intersection unless traveling on a green trajectory. Thus, HVs generally benefit from an increased number of green trajectories.

Dresner and Stone (2008) presented the one-lane signal policy (see Section 4.1.3). This policy results in green trajectories that originate from a single lane at a time, which significantly reduces the number of green trajectories. On the other hand, the one-lane signal policy was shown to have a dramatic negative effect on HVs.

We suggest a more conservative approach for reducing the number of green trajectories. Revisiting Figure 4.1, assume vehicle 3 is autonomous and is heading west. When applying H-AIM, vehicle 3 is automatically denied a reservation since the requested reservation crosses an active green trajectory. Currently, the lane on which vehicle 1 approaches the intersection allows continuing straight or turning right. If the turning policy on that lane is changed to "right only," the dashed straight green trajectory will no longer exist, allowing vehicle 3 to obtain a reservation.

4.2.1. Turning Assignment Policy

As was shown in the previous section, the performance of a managed intersection is affected by the allowed turning options in each lane. When considering a four-way intersection, each incoming lane has between one and three turning options from the set {left, straight, right}. The turning assignment policy assigns each incoming lane with allowed turns.

Assuming three incoming lanes, this study considers three representative turning assignment policies that are depicted in Figure 4.3. The policies are ordered and labeled according to degrees of freedom.

> **Definition 4** (Turning policy degree of freedom). Define degree of freedom for a lane as the number of turning options minus one. Define degree of freedom for a turning assignment policy as the sum of degrees of freedom over all lanes.

A restrictive turning policy is one that has a low degree of freedom, which, in turn, translates to fewer green trajectories. Policy 0 is an extreme case, representing the most restrictive turning policy (0 degrees of freedom). On the other hand, policy 4 is an extreme case of a liberal turning policy.

> **Definition 5** (Consistent turning policy). A turning assignment policy is said to be consistent if trajectories originating from the same road never cross each other.

In our representative policy set, turning policy 4 is not consistent, while 0 and 2 are. When considering more than one type of vehicle, different turning policy combinations might be considered. For instance, we might choose to assign one turning policy for HVs and a different one to CAVs (illustrated in Figure 4.4).

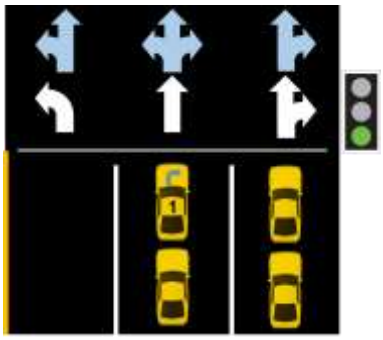


Figure 4.4 An inconsistent policy combination. Top: AV policy (blue arrows); Bottom: HV policy (white arrows)

> **Definition 6** (Consistent turning policy combination). A set of turning assignment policies are said to be a consistent combination if no trajectory from one policy crosses any trajectory from any other policy when both originate from the same road.

In our representative policy set, {0, 4} is a consistent turning policy combination (even though 4 is not a consistent policy on its own), while {2, 4} is not a consistent turning policy combination.

For safety reasons we don't consider assigning an inconsistent policy to HVs. On the other hand, assigning such a policy to CAVs is reasonable since conflicting reservations are automatically denied by the intersection manager. During our empirical study, we observed that assigning inconsistent policy combinations for CAVs and HVs is counterproductive from an efficiency standpoint and should be avoided. Figure 4.4 demonstrates the inefficiency that stems from an inconsistent turning policy combination. The figure presents a single road approaching a four-way

intersection. CAVs are assigned the turning policy shown on the top level (checkerboard texture) while HVs are assigned the bottom turning policy (plain texture). Vehicle 1 is autonomous. It is located in the middle lane and would like to turn right. Assuming a green signal for this incoming road and that HVs are arriving on the rightmost lane, vehicle 1 will not be able to obtain a reservation as it crosses an active green trajectory. Vehicle 1 will thus be stuck and will jam all the vehicles behind it despite having a green signal.

4.3. Empirical Study

This section presents results from a comprehensive empirical study. The goals of these experiments are two-fold:

1. Study the effectiveness of H-AIM for mixed traffic with an emphasis on low CAV ratios.
2. Indicate which turning policy should be assigned to HVs and CAVs in different traffic and CAV penetration levels.

Similar to the experiments presented by Dresner and Stone (2008), our experiments assume that a CAV may communicate with the intersection manager starting at a distance of 200 meters. Following Dresner and Stone, results are presented as averages over 20 instances per setting where each instance simulates one hour of traffic. Unlike Dresner and Stone's, our experiments assume a speed limit of 15 meters/second and a safety distance of 0.5 second between CAVs' trajectories. Dresner and Stone considered a speed limit of 25 meters/second, which is uncommonly high for signaled intersections, and a safety distance of 0.1 second, which might cause discomfort among passengers.

In line with our desiderata (presented in Section 4.1.5), we present average results for the following measurements:

- Average delay – see Definition 2.

- Maximal queue length – the maximal number of vehicles that simultaneously occupy a single incoming lane. Note that 29 vehicles is the maximal queue length for any lane in the simulator; no new vehicles will be generated on a lane as long as this limit is reached. When high traffic volumes are considered, the maximal queue length is often reached and queue spillbacks occur. In such cases it is hard to compare different policies as they all return similar results, making the maximal queue length measurement less valuable. Hence, we also report throughput.

- Throughput – the number of vehicles that cross the intersection in one hour. When low traffic volumes are considered, the maximal throughput is often reached since all approaching vehicles eventually cross the intersection. At high traffic volumes, when queue spillbacks occur, throughput can give evidence on the severity of spillbacks, i.e., the degree to which the spillbacks block new vehicles from entering the system.

The experiments presented in this section were obtained using the AIM4 simulator http://cs.utexas.edu/ ~aim/. Several adaptations were required in order to run these experiments.

4.3.1. Modifications to the AIM Simulator

Below is a list of changes introduced to the AIM simulator in order to simulate H-AIM. The reader is encouraged to view a video presenting the modified simulator at: http://youtube.com/watch?v=79UwpfD0u6s

- Vehicles are spawned with equal probability on all roads, and are generated via a Poisson process governed by the probability that a vehicle will be generated at each time step. Each vehicle is randomly assigned a type (HV or CAV) and destination. Given the assigned destination, a vehicle is placed on an incoming lane from which it can continue to its destination (the incoming lane must allow turning to the vehicle's destination). If several such lanes exist, it will be placed on the lane with the least number of vehicles currently on it. For instance, considering the example of Figure 4.1, a vehicle arriving at the intersection from the south heading north would be assigned the middle lane since the left lane does not allow continuing north and the right lane already has one vehicle (versus zero in the middle lane).

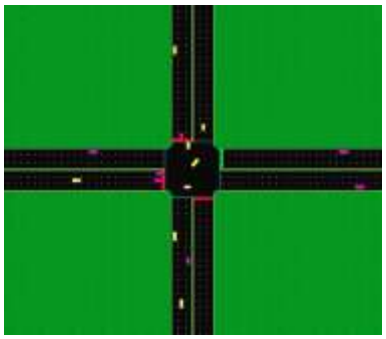
- Reservation requests for exiting the intersection more than 6 seconds in the future are queued but not processed. Only once the exit time drops below the 6-second threshold, a request is processed and might be approved. We add this constraint in order to allow the system enough time to detect all relevant approaching HVs. This restriction is not enforced for reservations entering the intersection on a green signal (HVs cannot be in conflict with such a reservation). At maximal speed (15 meters/second) a vehicle can travel a maximal distance of 90 meters in 6 seconds, resulting in it still being about 100 meters or 6.6 seconds away from the intersection, which is a considerable safety distance.

- A reservation is not necessarily denied if it conflicts with a green trajectory.

- A reservation is necessarily denied if it conflicts with an active green trajectory.

- HVs may turn right on red if the path is clear. This practice is a common case in the U.S.

Figure 4.5 presents a snapshot from the modified AIM simulator. HVs (in purple) wait at the entrance of the intersection for a green signal while CAVs (in yellow) are allowed to enter the intersection as long as they are following an approved reservation.



HVs (in purple) may not enter the intersection on a red signal. CAVs (in yellow) may enter the intersection on a red signal when following an approved reservation.

Figure 4.5 A screenshot from the modified AIM simulator

4.3.2. Four-way Intersection

Following Dresner and Stone (2008), we start by presenting results from simulating a four-way intersection with three lanes on each of the incoming roads (similar to the intersection presented in Figure 4.5). Thirty percent of the vehicles turn right at the intersection, 20% turn left, and 50% continue straight regardless of the incoming road and vehicle type.[10] A fixed six-phase traffic signal timing was used (the signal timing is presented in Table 4.1).

Table 4.1 Six-phase traffic signal timing

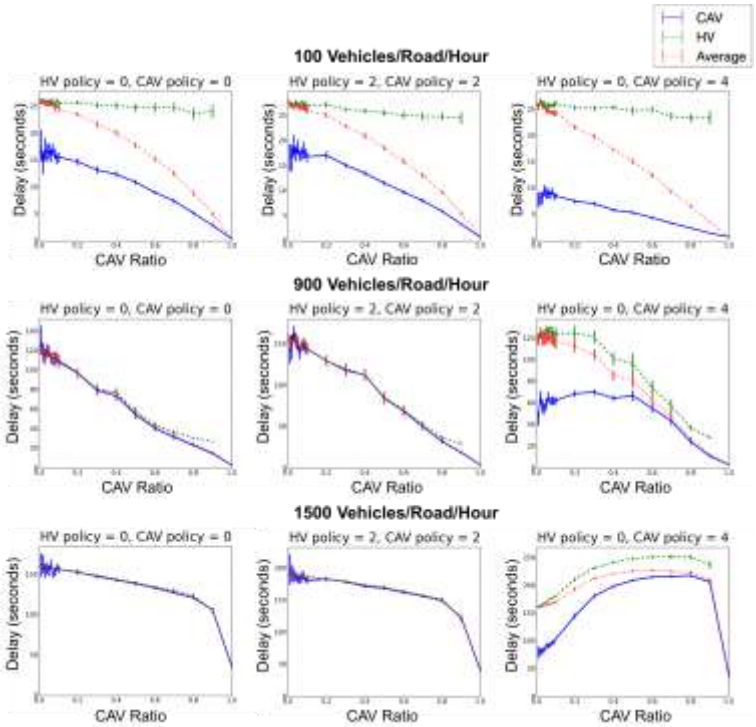
Phase	Direction	Green	Yellow
1	East-west	30	0
*2	Westbound	15	3
*3	Southbound	15	0
4	North-south	30	0
*5	Northbound	15	3
*6	Eastbound	15	0

Recall that under our assumption that HVs can turn right on red, the FCFS+Signals protocol has no advantage over traditional traffic signals (unless using the one-lane signal policy; see Section 4.1.3 for more details). Since FCFS+Signals using the one-lane signal policy was found to be helpful when considering 90% CAVs or more, it is not relevant to our current study, which focuses on early CAV adoption stages.

Results are presented for low, medium, and heavy traffic demand scenarios where 300, 900, and 1500 vehicles are spawned per incoming road per hour. The top part of Figure 4.6 presents three graphs for the four-way intersection scenario with low traffic demand. Each graph presents average delay for CAVs and HVs in seconds (y-axis) versus CAV penetration rates (x-axis). Each graph

[10] Dresner and Stone (2008) do not report the turning ratios for their mixed traffic experiment. Our turning ratio was chosen since it results in a good balance between the incoming queues when 100% of the vehicles are HVs.

refers to a different consistent turning policy combination based on the policies presented in Figure 4.3.



Note: The 95% confidence intervals are provided for each data point.

Figure 4.6 Average delays (y-axis) for different CAV penetration rates (x-axis) according to vehicle type in a four-way intersection scenario with low, medium, and heavy traffic demands (100, 900, 1500 vehicles/road/hour)

For low traffic demand, assigning a restrictive turning policy to HVs (policy 0) combined with a liberal turning policy (policy 4) for CAVs results in reduced delay for CAVs (especially at the early adoption stages) while having no significant negative effect on HVs' delay. The average delay over all vehicles (the "Average" line) is lower than the base case (where all vehicles yield to traffic signals, i.e., 100% HVs) and is decreasing as the CAV ratio increases.

The second row in Figure 4.6 presents results for a similar scenario with medium traffic demand. The trends are somewhat similar to those observed in the low traffic demand scenario in the sense that policy combination {HV-0, CAV-4} is most beneficial for CAVs while not hurting the performance of HVs. Unlike ({HV-0, CAV-0} and {HV-2, CAV-2}), the low traffic demand scenario, we see that other policy combinations present no advantage for CAVs over HVs. Nonetheless, the total delay is still clearly decreasing, which gives evidence that H-AIM is effectively improving the intersection's performance with regard to delays.

Finally, the bottom part of Figure 4.6 presents results for heavy traffic demand. Similar to the low and medium traffic demand cases, policy combination {HV-0, CAV-4} is most beneficial for CAVs at early adoption stages (until 0.2). However, as CAVs ratio increases, we observe an anomaly in the behavior of policy combination {HV-0, CAV-4}: the delay imposed on both HVs

and CAVs increases with the CAVs ratio. We explain this anomaly through the example depicted in Figure 4.7. In this example HVs are assigned a strict turning policy (policy 0) while CAVs are assigned a liberal policy (policy 4). Vehicle 1 is a CAV and would like to turn left from the middle lane. Assuming that a green signal is assigned to the east and westbound roads (phase 1 in Table 4.1), vehicle 1 is blocked from obtaining a reservation due to an active green trajectory. This active green trajectory is caused by continually arriving eastbound HVs (vehicle 2 for instance). Vehicle 1, being unable to obtain a reservation, blocks all vehicles behind it from entering the intersection. Imagine vehicle 3 is an HV and would like to continue straight. As long as vehicle 1 blocks the way, it is unable to cross the intersection despite having the right to pass (green signal).

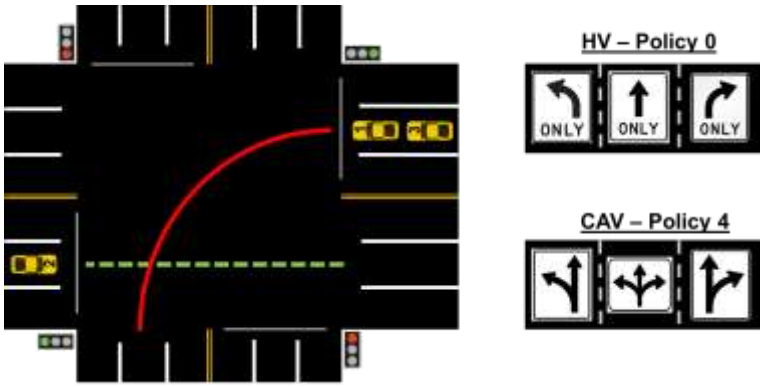


Figure 4.7 An example where a combination of strict turning policy for HVs and liberal policy CAVs is counterproductive [vehicle 1 (CAV) blocks vehicle 3 (HV) from passing the intersection]

Table 4.2 presents average results for maximal queue length and throughput for the four-way intersection scenario. Results are presented for different CAV penetration and traffic demand levels. An asterisk in front of a value indicates that it is significantly better (lower queue or higher throughput) compared to the values of the other two policies. There is no turning policy combination that is globally better for avoiding congestion (minimizing queue length or maximizing throughput). The best performing turning policy combination is a function of the traffic demand levels and CAV penetration levels. For most cases, policy {HV-0, CAV-4} performs best. A significant exception is observed at high traffic levels (500 vehicles/road/hour with> 0.1 but lower than 1). This result is consistent with the anomaly that is medium and high CAV ratios (discussed above) preventing policy {HV-0, CAV-4} from performing well in such cases.

Table 4.2 Results for a four-way intersection scenario using different turning policy combinations and different CAV penetration levels (CAV ratio)

CAV ratio	Maximal Queue			Throughput		
	HV-0, CAV-4	HV-0, CAV-0	HV-2, CAV-2	HV-0, CAV-4	HV-0, CAV-0	HV-2, CAV-2
	300 vehicles/road/hour					
0	9.33	**8.80**	9.78	**1,162**	1,157	1,157
0.01	9.30	**9.05**	9.70	1,150	**1,165**	1,152
0.05	**8.85**	8.90	9.70	**1,172**	1,170	1,171
0.1	**8.75**	8.93	9.63	1,154	1,151	**1,157**
0.3	* **7.10**	8.50	9.65	1,164	1,147	**1,164**
0.5	* **6.15**	8.30	8.70	**1,182**	1,161	1,174
0.7	* **4.60**	8.00	8.95	**1,179**	1,170	1,167
1	* **2.00**	3.75	3.60	* **1,194**	1,159	1,163
	900 vehicles/road/hour					
0	28.03	**27.93**	28.10	**3,103**	3,097	2,869
0.01	28.05	**28.00**	28.05	**3,111**	3,107	2,891
0.05	**27.95**	28.00	28.15	3,109	**3,122**	2,897
0.1	27.95	**27.90**	28.05	**3,144**	3,139	2,945
0.3	27.55	**27.50**	28.00	* **3,266**	3,233	3,089
0.5	**24.20**	25.55	27.90	* **3,384**	3,287	3,259
0.7	* **16.20**	20.30	27.25	* **3,465**	3,340	3,377
1	***4.90**	9.70	10.25	* **3,540**	3,351	3,358
	1500 vehicles/road/hour					
0	28.43	28.28	28.28	3,758	**3,763**	3,257
0.01	28.30	28.40	**28.25**	3,762	**3,777**	3,259
0.05	28.35	28.40	* **28.05**	* **3,818**	3,797	3,281
0.1	28.40	28.43	**28.33**	* **3,879**	3,821	3,314
0.3	28.55	28.40	28.40	3,887	* **3,978**	3,422
0.5	28.40	28.30	**28.25**	3,848	* **4,143**	3,564
0.7	28.45	**28.20**	28.35	3,880	* **4,339**	3,757
1	* **25.40**	27.90	27.95	* **5,756**	5,129	5,203

*An asterisk in front of a value indicates that it is significantly better (lower queue or higher throughput) compared to the values of the other two policies.

4.3.3. Three-way Intersection

Next, we present results from simulating a three-way intersection with two lanes in each of the incoming roads (similar to the intersection presented in Figure 4.8). Sixty percent of the eastbound or westbound vehicles continue straight while the rest (40%) turn (either right or left depending on the incoming road). Fifty percent of the northbound vehicles turn right and the rest (50%) left. We used a three-phase fixed traffic signal timing that is presented in Table 4.3.

Figure 4.8 also depicts three representative turning policies (with 0, 3, and 6 degrees of freedom). Since a three-way intersection is not symmetrical, each turning policy is broken into three policies (one per origin).

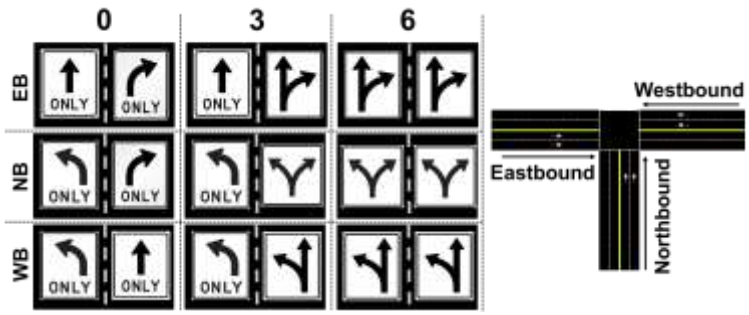


Figure 4.8 Three different turning assignment policies for a two-lane road approaching a three-way intersection

Values represent maximal queue length and throughput. An asterisk represents a significant advantage for one policy over the two others using a single tale unpaired t-test with 95% confidence. We chose these three policies as they resemble the ones used in the four-way intersection experiment. Policy 0 is the most restrictive policy, similar to policy 0 in the four-way case. Policy 3 has the highest degree of freedom among the consistent policies, similar to policy 2 in the four-way case. Policy 6 has the maximal degrees of freedom overall, resembling policy 4 in the four-way case.

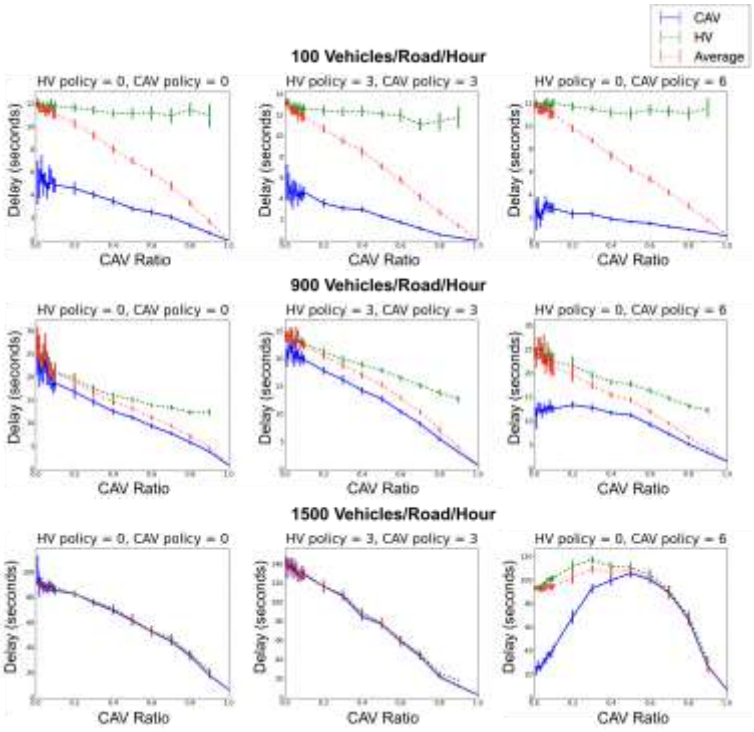
Figure 4.9 presents nine graphs for the three-way intersection case. The layout of these graphs is similar to those presented for the four-way case (Figure 4.6). Each graph is affiliated with one of the three consistent turning policies combinations shown in Figure 4.8. Results show a general trend that is similar to the one observed in the four-way intersection scenario. For low and medium traffic demand, assigning a restrictive policy to HVs (policy 0) and a liberal one to CAVs (policy 6) is most beneficial for reducing delays as well as giving CAVs the biggest relative advantage over HVs. At high traffic demand, on the other hand, policy combination {HV-0, CAV-6} is counterproductive, similar to the anomaly observed in the four-way case). (Figure 4.7 can be easily adapted to apply for a three-way intersection.)

Similar to Table 4.2, Table 4.3 presents maximal queue length and throughput but for the three-way intersection scenario. For this scenario we observe that policy {HV-0, CAV-0} is never significantly superior to the two other policies. This result seems to be in contradiction to the results presented in Figure 4.9 where, for the case of heavy traffic demand, policy {HV-0, CAV-0} seems to outperform the others.

Table 4.3 Three-phase traffic signal timing

Phase	Bound	Green	Yellow
1	East-west	30	0
*2	Westbound	15	3
*3	Northbound	15	3

(Green and yellow duration are given in seconds. An asterisk next to a phase number means that left turns are allowed during that phase.)



Note: The 95% confidence intervals are provided for each data point.

Figure 4.9: Average delays (y-axis) for different CAV penetration rates (x-axis) according to vehicle type in a three-way intersection scenario with low, medium, and heavy traffic demands (100, 900, 1500 vehicles/road/hour)

This discrepancy is due to the fact that north and westbound vehicles that request to turn left can do so only from the leftmost lane under policy {HV-0, CAV-0}. As such, the leftmost lane becomes congested and, once the queue reaches 29 vehicles, no more left-turning vehicles are spawned. In such cases the throughput and queue length decrease and increase, respectively, while the average delay decreases, since left-turning vehicles that suffer from more delays are less abundant. The same explains the fact that policy {HV-3, CAV-3} produces shorter queues in many cases, allowing left-turning vehicles to use both lanes alleviates the congestion formed on the leftmost lane.

4.4. Hardware and Costs for Smart Intersection and Micro-Tolling Applications

In order to accomplish an implementation of AIM, H-AIM, or D-tolling, the most critical piece of hardware necessary is a roadside unit (RSU) capable of supporting vehicle-to-infrastructure (V2I) communication. One solution to this requirement, provided by Applied Information and which leverages both cellular and DSRC communication technologies, costs approximately $6,000 per intersection. In addition to providing sufficient communication capabilities for implementing AIM or HAIM, this solution may also provide data on travel times over a road segment. Such a solution has been tested successfully for less involved projects by the Alabama Department of Transportation as well as Purdue University.

In implementing H-AIM, the need to continue operating traditional signal controllers requires a traditional traffic signal controller capable of network communication, and it must also provide a suitable advanced traffic controller (ATC) interface. An example of such a controller would be the Siemens M60 ATC signal controller, which is capable of running Linux and features a minimum of two network interfaces. The estimated price for this model controller is $5,000, which in most use cases is sufficient to control a single intersection. This model can be found commonly in the field where newer traffic controllers have been installed.

4.5. Conclusion and Future Work

Though previous intersection management protocols were shown to be extremely efficient in coordinating CAVs through an intersection, they were shown to provide no or little improvement until 90% of the processed vehicles are CAVs. This chapter presents Hybrid-AIM (H-AIM), an efficient intersection management protocol for early CAV penetration stages. H-AIM builds on the values representing maximal queue length and throughput.

Table 4.4 provides results for a three-way intersection scenario using different turning policy combinations and different CAV penetration levels (CAV ratio). The AIM protocol (Dresner and Stone, 2008) is applicable under the assumption that vehicles approaching the intersection can be sensed (on top of the assumptions required by AIM).

Table 4.4 Results for a three-way intersection scenario using different turning policy combinations and different CAV penetration levels (CAV ratio)

CAV ratio	Maximal Queue				Throughput		
	HV-0, CAV-6	**HV-0, CAV-0**	**HV-3, CAV-3**		**HV-0, CAV-6**	**HV-0, CAV-0**	**HV-3, CAV-3**
			300 vehicles/road/hour				
0	6.00	6.10	* **5.28**		583	587	**587**
0.01	5.90	5.85	* **5.00**		575	571	**589**
0.05	5.70	6.10	* **4.95**		575	**582**	576
0.1	5.88	6.25	* **5.10**		580	**590**	587
0.3	5.15	5.65	* **4.70**		587	579	**589**
0.5	* **4.00**	5.30	4.90		590	582	**592**
0.7	**3.60**	5.65	3.95		**587**	578	585
1	* **2.15**	3.20	2.45		595	589	**600**
			900 vehicles/road/hour				
0	18.93	18.83	* **14.10**		1,679	1,670	* **1,716**
0.01	15.95	19.50	* **13.20**		1,689	1,687	**1,700**
0.05	17.45	19.30	* **13.15**		1,682	1,680	**1,696**
0.1	14.67	15.28	* **13.08**		1,695	1,683	**1,707**
0.3	* **11.40**	13.75	12.30		**1,716**	1,691	1,713
0.5	* **10.00**	12.85	11.40		* **1,759**	1,694	1,723
0.7	* **8.25**	10.90	10.30		* **1,756**	1,680	1,723
1	* **4.65**	6.35	5.15		* **1,775**	1,699	1,727
			1500 vehicles/road/hour				
0	28.15	**28.05**	28.18		2,356	**2,360**	2,313
0.01	28.10	28.10	28.10		**2,381**	2,373	2,319
0.05	28.05	28.05	28.20		* **2,405**	2,389	2,348
0.1	**28.05**	28.08	28.10		* **2,467**	2,412	2,373
0.3	**28.00**	28.05	28.05		* **2,523**	2,476	2,496
0.5	28.15	28.00	**27.95**		2,534	2,562	* **2,632**
0.7	28.05	27.90	* **26.95**		2,629	2,637	* **2,781**
1	13.00	18.30	* **9.85**		* **2,905**	2,726	2,791

*An asterisk represents a significant advantage for one policy over the other two, as determined using a single tale unpaired t-test with 95% confidence.

When an approaching HV is sensed by H-AIM, the protocol examines whether the current traffic signal assignment allows the HV the right to pass. If this is the case, H-AIM reserves the relevant trajectory through the intersection and denies any conflicting reservation requests.

Results obtained from a comprehensive empirical study support the following general conclusions:

- At non-extreme CAV penetration levels (between 0 and 0.9), H-AIM is superior to previous approaches (AIM, traffic signals).
- At low and medium traffic demands, a turning policy that restricts HVs while allowing maximal flexibility to CAVs is recommended for reduced average delay, reduced congestion, and encouraging CAV adoption (since CAVs suffer from lower delays compared to HVs).
- At high traffic demand, restricting HVs while allowing CAVs maximal flexibility is beneficial only at early CAV adoption stages (≤ 0.1). Beyond early adoption stages, such a policy combination is counterproductive and other policies should be considered.

Future work will study the effects of H-AIM when semi-autonomous vehicles are considered (Au et al., 2015) and are assigned different turning policies. Future work will also examine how different traffic conditions affect the performance of H-AIM, where traffic conditions relate to the number of lanes on different incoming roads, turning ratios, traffic signal timing, imbalanced traffic (different volume of vehicles arrive on different incoming roads), safety buffer size, and speed limit. Ultimately, our goal is to test H-AIM on real intersections with real vehicles.

Chapter 5. Methods for Implementing Smart Intersections

5.1. Background

This project's Phase 1 report (0-6838-2) demonstrated via simulations of various Austin, TX networks that when autonomous vehicle (AV) penetration is high, changing traffic signals to smart intersections that adopt a "first come, first served" (FCFS) tile-based reservation (TBR) system control can often improve local throughput and system-wide conditions. For example, in the arterial Congress Avenue network, with all AVs in simulation, TBR slightly outperformed traffic signal control for all demand scenarios, except for the highest demand. Both scenarios greatly improved travel times when compared to the network with current conditions [traffic signals and only human-driven vehicles (HVs)].

However, this improvement trend was not evident for all networks tested, and a paradoxical effect was seen in some scenarios. In the Lamar & 38th Street arterial network, TBR control actually hindered system conditions at higher demands with a higher total system travel time (TSTT) than even the current conditions network. This system-wide degradation was most likely due to queue spillback onto links surrounding the major intersection and adjacent minor intersections of Lamar & 38th Street. The large spillback may have been due to the close proximity of major and minor intersections in the network along with the FCFS priority function of the reservation intersections. Because of FCFS, a small demand coming from a minor street approach to a TBR intersection could interrupt the progression of a much larger demand of vehicles approaching from a major or arterial street. It was also seen in the Congress Avenue network that at the highest demand, signals outperformed TBR with AVs. Although this paradoxical effect of smart intersection control appeared in the smaller arterial networks, the larger downtown Austin network covering both arterial networks saw only great improvements in system conditions, with a 55% lower TSTT using TBR compared to signal control at the highest demand. The paradoxical effect in the larger network was prevented due to the dynamic route choice of vehicles, which could allow avoidance of such intersection clusters.

This chapter presents methods to identify subsets of intersections where smart TBR-controlled intersections would provide not just local benefits of increased throughput in the intersection, but system-wide benefits specifically in terms of congestion and travel time. As previously found, reservation control can possess retrogressive impacts on system-wide congestion and some combination of TBR-controlled and signal-controlled intersections would offer better results than a network with only TBR control or only signalized control. To assess this proposition and identify favorable smart intersections, two main methods were used. First, a multilinear regression model was estimated to predict the differential impacts of individual intersection characteristics on travel times as a smart and as a signalized intersection. Second, a genetic algorithm (GA) was used in coordination with a dynamic traffic assignment (DTA) model solving for dynamic user equilibrium (DUE) to find system-optimal allocations of smart intersection subsets. Next, this

chapter presents results of the two methodologies when applied to a large-scale Austin city network. Finally, this chapter develops easy-to-use guidelines and prioritization rules for deploying smart intersections in a way likely to maximize benefits, and for the selection of real-world implementation testbeds.

5.2. Methodology

This section presents the formulation and methodology of two methods used to develop quantitative measures of the benefits and costs associated with replacing a traditional signal with a reservation-based smart intersection. To quantitatively identify a deployment strategy, the calculation of costs and benefits should be fairly simple and rely on a parsimonious set of independent variables, using regression from simulation results. Finding a system-optimal grouping of smart and signalized intersections in a network giving the minimum TSTT is a more difficult task. Due to the bi-level nature of the optimization problem, with the second layer defined by the solving of DUE on a large-scale network, the overall problem is NP-hard to solve exactly. Thus, a meta-heuristic is used to find a solution.

In an effort to generalize the effect of individual intersection characteristics on the differential impact of travel times of smart versus signalized control, a multilinear regression model is estimated using a mixture of DTA simulation data and readily available intersection data such as signal properties and turning demand. The model input data is collected from subsets of signalized intersections in different large-scale city networks. The aim of this model is to apply the regression to a subset of a network's intersections and output a ranking of the best smart intersection candidates. This ranking can then be tested in simulation to evaluate model accuracy, and the model can be used to easily develop smart intersection selection and prioritization rules.

In order to find a system-optimal grouping of smart and signalized intersections in a network, a GA is used. This section details the specific algorithm used to find a feasible solution and its steps.

5.2.1. A Mesoscopic Dynamic Traffic Assignment Model

This subsection serves to define the DTA model used to simulate all networks when solving for DUE, including the model specifications and assumptions/submodels. The mentioned model is a custom implementation of a mesoscopic DTA model in Java that can capture HV and AV behavior either separately or together on any network. A multiclass cell transmission model (CTM) is used to propagate flow through network links based on hydrodynamic flow theory. To model reservation-based intersection control, a conflict region model is used that divides an intersection region into larger and simplified conflict regions, each with a capacity, either accepting or rejecting vehicle requests based on some priority function. The priority function assumed in this study is a FCFS function in which the first vehicle to make a request with the intersection manager is processed by the manager first. The DTA model solves for DUE using the method of successive averages to a convergence defined by a 1% relative gap. Primarily, to simulate AV behavior, AVs are assumed a 0.5-second reaction time compared to HVs, which have a 1-second reaction time. This difference in reaction time leads to increased roadway capacity caused by increased backward

wave speed and reduced following headways as the proportion of AVs in a network increases, with the greatest increased capacity associated with 100% AVs. It is also assumed that only AVs can use the TBR intersections; thus, for the sake of this study, all DTA simulation runs are evaluated with only AVs in the networks.

For reference, the two real city networks used in simulation include the downtown Austin and downtown Dallas networks with a total demand of 62,783 and 167,592 vehicle trips over a 4-hour observation period, respectively. This total demand is considered 100% demand as the next two sections may refer to a proportion of this demand.

5.3. A Multilinear Regression Model

5.3.1. Motivation

The goal of the multilinear regression formulated in this section is to quantitatively predict any intersection's relative utility under smart control compared to traditional signal control, given some basic intersection characteristics. If this can be accomplished, then a set of intersections can effectively be ranked according to this utility and smart intersection deployment strategies can be easily developed based on quantitative metrics.

5.3.2. Formulation

To summarize the utility or performance of any intersection in a network, a difference in effective TSTT of an intersection under signal control and TBR control is used as the primary response variable. To obtain the regression input dataset, a set of N intersections is selected from a parent network. In order to effectively measure the effect on system-wide TSTT of a single intersection in a large network, a small subnetwork is created involving the observed intersection and the immediately adjacent links and nodes, with the nodes only acting as origins and destinations. The intersection is then assumed a control (TBR or signal) and is solved for DUE using a DTA simulator and a user specified origin-destination (OD) demand matrix. The response variable is then found by subtracting the TSTT under TBR control from the TSTT with signalized control. Multiple predictor variables described in Section 5.3.3 are obtained from given network input data and from simulation. Specific methods and sources of data collection are presented in Section 5.3.4. The general regression formula is as follows:

$$TSTT_{signal} - TSTT_{TBR} = \Delta TSTT = FFTT + \vec{\beta} * \vec{X}$$

where β is the vector of variable coefficients and $X$is the vector of predictor variables.

To encapsulate effects of different levels of demand on an intersection, the single intersection network is solved for DUE with the DTA simulator under each of six different demand levels: $d = \{10\%, 30\%, 50\%, 75\%, 85\%, 100\%\}$. The OD demand matrices for the listed demands are obtained through simulation, as detailed earlier. With N intersections selected to observe from the parent network and six demand levels considered, a total of $N * 6$ data points are found to estimate

the desired multilinear regression model in which a $\Delta TSTT$ is predicted given some intersection characteristics, $\vec{X}$.

The lower the $\Delta TSTT$, the better the intersection performs under TBR control compared to signal control. By obtaining this response variable, intersections can effectively be ranked according to the differential utility between the two controls.

5.3.3. Variables

Table 5.1 lists the set of possible predictor variables used in the regression model to predict the difference in TSTT between an intersection under signal control and TBR control.

Table 5.1 Possible predictor variables

	Predictor Variable	Description of Variable	Variable type
1.	Number of phases	The total number of signal phases across a cycle	Number of phases
2.	Cycle length	The signal's time of one complete phasing cycle	Time (seconds)
3.	Number of moves	The total number of non-restrictive turning movements for the intersection. Turning movements are defined by an approach link and an exit link.	Number of turning movements
4.	Average lane vehicle count	The average number of total vehicles using the intersections incoming and outgoing lanes	Number of vehicles/hour/lane
5.	Number of through turns	The total cumulative through demand of the intersection across all approaches	Number of vehicles
6.	Number of left turns	The total cumulative left-turn demand of the intersection across all approaches	Number of vehicles
7.	Number of right turns	The total cumulative right-turn demand of the intersection across all approaches	Number of vehicles
8.	Minimum length	The minimum length of a link entering or exiting the intersection	Length in feet
9.	Maximum length	The maximum length of a link entering or exiting the intersection	Length in feet
10.	Average length	The average length of a link entering or exiting the intersection	Length in meters
11.	Minimum link capacity	The minimum capacity of a link entering or exiting the intersection	Number of vehicles/hour
12.	Total link capacity	The total cumulative capacity of all links entering or exiting the intersection	Number of vehicles/hour

5.3.4. Data Collection

The primary response variable for this regression model is a difference in TSTT between an intersection under signal control and TBR control. To find the TSTT of an individual intersection, a new subnetwork of only the specified intersection is created. The new single-intersection network contains only one intersection, which acts as a real intersection to move vehicles across links, and all directly adjacent nodes to the intersection, which merely act as the network's origins

and destinations in which vehicles enter and exit the network. The new network also contains centroid nodes and connectors that act as the network's loading points in which vehicles exit and enter. To define the subnetwork's OD demand for a specific control at any given demand level, a DTA simulation of the larger parent network is run at the desired demand level with all eligible intersections assuming the specified control. (The demand level describes the proportion of total demand seen in a network.) The "eligible" intersection set is defined in Section 5.3.5. A cumulative count of vehicles is kept during simulation for each possible turning movement in each intersection, and because the new subnetworks contain only one intersection each, the counts for each turning movement in the parent run can be combined to define OD demand for each intersection. For each intersection, this demand is then set as the new subnetwork OD-demand matrix, a DTA simulation run of the network is completed under TBR and signal control with a TSTT output for each, and a $\Delta TSTT$ is found through the difference.

Average lane vehicle count, through turns, right turns, and left turns are all predictor variables found through simulation as well. While running the parent network DTA simulation to find OD-demand matrices for each intersection as described above, a cumulative count of vehicles using each possible turning movement was recorded, which was simply translated into total through, left, and right turns for the intersection. The average lane vehicle count was found by averaging the total experienced vehicle counts for all incoming and outgoing lanes in the intersection.

All other potential variables defined in Section 5.3.3 were obtained through city network datasets.

In this report, two parent networks—downtown Dallas and downtown Austin—were used to obtain regression input datasets and estimate the respective regression models. A total of 174 and 152 intersections were observed from the downtown Austin and downtown Dallas networks respectively.

5.3.5. Assumptions

The following assumptions were made while developing the regression model:

- The set of intersections that can be switched to autonomous intersections are the set of traffic signals in the network. The model does not consider the set of merges, diverges, or stop-sign-controlled intersections. This is assumed as signalized intersections are likely to be the priority intersections to change because TBR provides little system-wide benefit when applied to non-signalized intersections, as shown in previous studies (Patel and Levin, 2016).

- All DTA simulations are run using a demand composed only of AVs. Because HVs are assumed to not use TBR intersection control, a demand of only AVs is required for stable results and analysis between networks with mixtures of TBR and signal intersections.

- Because an independent subnetwork is created for each intersection containing only the intersection, it is assumed for the DTA simulation that the intersection is independent of

all others in the network. Other intersections do not impact the observed intersection over time; however, as described in Section 5.3.4, OD-demand matrices are found through observed vehicle counts in the parent DTA simulation run. The parent DTA simulation does include all intersections (which are eventually isolated as subnetworks), and so demand in the subnetwork OD matrices is from a complete network with intersection interactions.

5.3.6. Model Metrics

The following metrics are used to evaluate the performance and accuracy of the estimated linear regression models, including their variables and other specifications.

(1) R^2, or the regression's coefficient of determination is a statistical measure of how close the actual data are to the fitted regression model. It is measured as:

$$R^2 = SS_{reg}/SS_{total}$$

where $SS_{reg}$is the total variance in the data explained by the model and SS_{total} is the total variance in the data.

$t - tests$

(2) A standard t-test is conducted for each predictor variable with a confidence interval threshold of 95% ($t_{.95} = 1.645$), translating to a variable being significant if: $t_{variable} \geq 1.645$with 95% confidence. Although this defines a threshold for significance, some variables may be chosen to remain in the model if some other significance is seen.

5.4. A Genetic Algorithm for System-optimal Placement of Reservation-based Intersections

The goal of the GA approach is to identify the spatial orientation of smart intersections and traditional intersections in the network that provides the best benefit in terms of TSTT. However, it is impractical to assume that a smart intersection can be installed at every intersection in the network. Therefore, this section explores two sub-problems. First, it identifies the system-optimal allocation of reservations in the network that produces the best system-wide benefits. The second experiment finds the optimal allocation of TBRs that produces the best benefit when there is a limit to the number of TBRs that can be installed. In both cases, the GA is used in coordination with a DTA model to find the DUE solution.

5.4.1. Assumptions

The following were assumed while developing the model:

- The set of intersections that can be switched to autonomous intersections are the set of traffic signals in the network. The model does not consider the set of merges, diverges, and left and right turns.

- In the DTA model, the only demand in the network is the demand due to AVs.

5.4.2. A Background on Genetic Algorithms

A GA is a class of computational methods inspired by genetic evolution used to solve constrained and unconstrained optimization problems. In a GA, an initial random population of candidate solutions is created. An evaluation metric is used to determine the fitness value of each of these parents. At each step of the algorithm, a pair of parents is selected at random from the population to reproduce to create child individuals. A GA follows a set of steps at each iteration to create children, which ensures that the best characteristics of the parent generation are preserved. In our target problem, each individual in the population is a specific orientation of the network where each intersection is modeled either as a traffic signal or as a TBR. The characteristic that evolves from one generation to the next is the orientation of each of these intersections in the network. The algorithm is designed such that the orientation of each intersection that results in the best observed TSTT is retained in future generations. These steps are outlined below. Section 5.4.4 describes the specific design of the GA used to solve the problem at hand. A pseudocode of the algorithm is presented below.

5.4.3. Pseudocode

In this section, we define the steps involved in a GA followed by a detailed discussion of each of the steps. A pseudocode of the algorithm is presented below:

1. Create an initial random population of *n* individuals with random intersection controls.
2. Evaluate the goodness of each individual in the population to the problem.
3. Sort the population in decreasing order of the fitness measure.
4. While iteration < maxIterations repeat
    a. Select a pair of the individuals in the population at random as candidate parents to cross over to form children.
    b. Generate child individuals by crossing over candidate parents.
    c. Introduce mutation in the children. Evaluate the goodness of the child individuals created.
    d. Sort the population in decreasing order of the fitness measure.
    e. Remove the lowest-performing 50 individuals from the population.

5.4.4. Genetic Algorithm Steps

5.4.4.1. Candidate Selection

A pair of individuals is selected as parents at random from the best-performing individuals of the current generation. This depends on the proportion of the current population we choose to keep at

each step of the algorithm. For example, if we wish to create 20% children at each step of the algorithm, the GA chooses parents from the best-performing 80% of the current generation.

5.4.4.2. Crossover Probability

The crossover probability quantifies how often crossover will occur between parents. This helps decide which characteristics of each parent enters a child in the next iteration. We have developed a heuristic to determine this probability, described below. The probability is as follows:

$$Crossover\ Probabiltiy = 0.5 + 0.5 * \frac{|TSTT(Parent1) - TSTT(Parent2)|}{TSTT(100\%\ Signals) - TSTT(100\%\ TBRs)}$$

This probability equation's denominator is the upper bound on observable difference in TSTT between two different orientations of the network—one where every intersection is modeled as a traffic signal and one where every intersection is modeled as a TBR. The crossover probability is designed to ensure that the parent with the better TSTT has a higher chance of having its intersection orientation passed on to the child.

5.4.4.3. Mutation

Mutation is a genetic operator used to maintain genetic diversity from one generation to the next. It is usually applied with a low probability. There are many types of mutations used in GAs. In this model, we have chosen an inversion mutation. An inversion mutation is one in which the orientation of each intersection in the individual is inverted if the inversion probability is met. Mutation is done to introduce a degree of diversity to the genetic structure of each generation. In the experiments, the probability of mutation has been set to 0.7%

5.4.5. Model Inputs

5.4.5.1. Population Size

The population size is the number of parents created at the beginning of the GA. The population size often determines the quality of the solutions obtained from a GA. A small population size will result in quicker convergence but may result in the algorithm getting trapped in a local optimum. However, a large population will slow the model down significantly. It is prudent to choose a population size that avoids both of these pitfalls. The experiments use a population size of 100 parents to start our GA.

5.4.5.2. Proportion of Children

This parameter determines the proportion of children to generate at each iteration of the GA. Consequently, this affects the quality of the population in each generation.

5.4.5.3. Maximum TBRs

This parameter limits the number of TBRs in each individual in the population. This has been modeled as a parameter in the model, since it was observed that an all-TBR network performed significantly worse in some scenarios.

5.4.5.4. Convergence (Max Iterations)

The convergence criterion is the maximum number of iterations to be completed before the algorithm terminates. The max iterations criterion suffers from extremes. A small value for maximum iterations will result in the algorithm converging prematurely, sometimes before it hits a local optima. Similarly, a large value of maximum iterations will slow the algorithm from converging and may result in cycling after it reaches a local optima. It is thus prudent to choose an optimal convergence criterion that avoids both these problems.

5.4.6. Model Output: DTA DUE Solution

When a new individual is generated, a goodness calculation is performed to evaluate how good the solution is. In the problem at hand, the goodness calculation is the TSTT observed from solving the specific orientation of the network to equilibrium using a simulation-based DTA model.

5.5. Experimental Results

This section presents experimental results of applying the linear regression and the GA models, as described in Section 5.2, to a large-scale city network. The presented regression model is used to find a ranking of the "best" candidate intersections in the network to assume reservation control and the GA is used to find an optimal subset of intersections in various set sizes that minimize the TSTT of the network. The network used for all experimental simulation results is the downtown network of Austin, which contains 1,247 links, 546 intersections (174 signalized), 171 zones, and a total demand of 62,783 vehicle trips over a 4-hour observation period. This network includes the smaller arterial networks discussed in Section 5.1, including Congress Avenue and Lamar & 38th Street, where paradoxes in the benefit of TBR were seen. The DTA model used in this section is described in Section 5.4.6.

5.5.1. Linear Regression Results

With the input data for the Dallas network, a linear regression model was estimated with the goal of identifying significant intersection characteristics that affect the relative system-wide utility of a single intersection under reservation control compared to signal control. It is the primary goal that with an accurate regression model, a set of eligible (signalized) intersections in a network can be ranked in terms of their relative utilities. A subset of these intersections, starting with the intersection holding the largest utility value and moving in decreasing order, can be chosen to assume reservation control and provide the most benefit to the system's TSTT compared to any other same-size subset of the eligible set. Recall that the relative utility is $\Delta TSTT = TSTT_{signal} -$

$TSTT_{TBR}$, so a larger $\Delta TSTT$ means a larger benefit to the system with the intersection under TBR control. This section presents the estimated regression model and results of applying this ranking system to a network in simulation.

5.5.2. Model Outputs

This section presents the regression model used to predict $\Delta TSTT$ and the variables found to be significant in identifying an intersection suitable for reservation control compared to signal control.

Table 5.2 presents the results of the regression model. In using the intersections from Dallas as training data for the regression model, the model is able to predict 75% of the variation in the data for the Austin network. This is evident from the R^2 value of 0.754 in Table 5.2.

Table 5.2 Summary of the accuracy of the linear regression model

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.868	.754	.752	360.81758

5.5.3. Significant Variables

From the pool of variables presented in Section 5.3.3, the variables presented in Table 5.3 were found to be significant predictors for the regression model. The relative importance of each variable can be gauged by the magnitude of the coefficient of the variable in the model. This is also evidenced from the value of t-value for each variable at 95% confidence level. Recall that a variable is considered significant if $t_{variable} \geq 1.645$. Table 5.3 presents the relative importance of each variable in predicting the difference in TSTT for each intersection. From the model, it is evident that cycle length, number of moves, number of through turns, number of left turns, number of right turns, and minimum length have high t-values and are thus significant predictors in the model. However, the minimum capacity of link has t-value lower than the set threshold value. It has been retained in the model because it is indicative of queue spillback in the network (a link with a low capacity entering an intersection could result in spillback if there is significant network congestion).

Table 5.3 Summary of the significant variables in the regression model

Coefficients				
Model		Unstandardized Coefficients		t
		B	Std. Error	
1	(Constant)	-717.296	146.626	-4.892
	Cycle length	3.286	.295	11.126
	Number of moves	9.495	4.411	2.153
	Number of through turns	.261	.011	23.621
	Number of left turns	.430	.028	15.412
	Number of right turns	.414	.031	13.191
	Minimum length	.409	.231	1.767
	Minimum link capacity	-.025	.022	-1.140

5.5.4. Regression Results in Simulation

Before testing the regression model in simulation by applying it to the Austin intersections, the input data used to train the regression is tested to ensure this base data offers desirable results in simulation. To test the base data, the $\Delta TSTT$ objective function values for each eligible intersection in the Austin network are used to rank the intersections in descending order of their objectives, leaving the intersections which perform better as reservations to improve system-wide congestion, higher in the ranking list. To evaluate the ranking, an experiment with each of a different proportion (including 20%, 40%, 60%, and 80%) of the highest ranked intersections is conducted by forcing this subset of intersections to assume TBR control and the rest to assume signal control. For example, in the 40% experiment, intersections ranked 1 through 70 (40% of 174 eligible intersections) are chosen to take on TBR control, and the rest signal control. This subset of intersections is predicted to provide more system-wide benefit to TSTT than any other same sized subset of intersections. A DTA simulation was then run for each before mentioned proportion of the top TBR candidate intersections. To analyze the margin of effect seen in TSTT for all presented results, two control or base case scenarios are run to provide a bound on currently known TSTT's. The first is the case of the network with only signal control and the second is the case of the network with only TBR control for all eligible intersections, yielding a TSTT of 6443.22 hours and 4560.14 hours respectively. Keep in mind that all experiments in this study are run with a demand of only AVs.

As shown in the graph in Figure 5.1, results for the base input data at 20% TBR-controlled intersections show a 20% drop in TSTT from the network with only signals which is to be

somewhat expected. In general, TSTT decreased at a slightly decreasing rate as the proportion of TBR control in the network increased which is also to be expected. However, the base data results actually show a 4% and 8% lower TSTT at 60% and 80% TBR control, respectively, compared to the base case with 100% TBR control. The graph in Figure 5.1 shows a red and green horizontal line showing the 100% signal and 100% TBR base cases respectively, and the Raw TSTT_DELTA data points show the decreased TSTT of the 60% and 80% experiments, and even show that at 40% TBR control, the difference in TSTT with the 100% TBR control base case is quite small. These decreased TSTT's, although not large in size, show that more system-wide benefit can be seen using only a fraction of the reservation control. This also backs the notion that some intersections, typically lower in the ranked list of TBR candidates, tend to help the traffic system as signals. In the 100% TBR base case, great improvement is seen over the 100% signals base case and there is no real sign of TBR causing any increased congestion compared to a better performing and smaller subset of possible TBR intersections, partially due to dynamic route choice. As vehicles can dynamically alter paths based on their observation of the network's state, they can avoid such TBR intersections which become very congested, however this can lead to congestion in other portions of the network. Such rerouting due to paradoxical queue spillback seen near TBR intersections can be avoided with the right intersections.

Next, the Dallas-based regression is applied to the 174 eligible Austin intersections to predict their relative utilities, and the intersections are ranked in descending order as was done with the base input data. The same experiments for the predicted ranked list of intersections were then conducted and shown on the graph in Figure 5.1. Results show the same general trend of decreasing TSTT as the proportion of TBR control increased held, however all simulation outputs of TSTT were approximately 17% higher than those found in the pre-regression input data experiments. This large difference is not expected as the predicted regression model showed a somewhat close fit to the data with an R^2 of 0.754. Although the regression results outperformed the 100% signals base case, this was to be expected of almost any network with a subset of intersections using reservation control. At a TBR proportion of 80%, there does seem to be a TSTT falling very close to the base case of 100% TBR control showing that a similar result can be seen with 35 less reservation intersections, however it does not outperform the 100% case like the data the regression was modeled after.

To confirm the validity of a set of downtown Dallas intersections being able to accurately predict a set of Austin intersections, the Austin network base characteristic data was used to estimate a new regression model so that the Austin network could predict the relative utilities of its own intersections. DTA simulation results show a set of points nearly identical to that seen in the Dallas-based regression, shown in Figure 5.1. This confirmed that intersections from the two different networks seem to have no significant difference in their respective characteristics' relation to their differential impact on system-wide congestion under the two intersection controls. Finally, it is not clear the magnitude of the regression ranked list's inaccuracy compared to the original input data only through simulation. As a comparison and means of testing the accuracy of the regression's predicted results, for each analyzed proportion of TBR control in a network, 20 networks are randomly generated subject to the number of TBRs remaining equal to the respective

proportion and their TSTT values are averaged for a new data point. As shown in Figure 5.1 with the dashed plotted line, on average, the randomly generated networks actually outperformed both regressions for each TBR proportion, suggesting that the ranked list predicted by the regression was inaccurate.

The shortcomings of the regression model are most likely due to uncaptured interdependencies between intersections. Results show that the base data obtained through the simulation of single-intersection networks yielded a high performing ranking of intersections in terms of their differential utility under reservation and signal control. This suggests that some intersections are better for system-wide congestion under TBR control compared to others, and that this trend cannot be captured linearly. Most likely, because demand for the single-intersection networks for each intersection was directly obtained through simulation of the parent network, some intersection dependency was captured even though intersections were tested as independent. The regression also only identifies single intersections as TBR-controlled and does not identify the clustering of intersections where reservations could provide the most benefit. This interdependency between intersections may heavily impact the system-wide effect that a single intersection carries, most likely through demand allocation. A group of intersections may not necessarily improve the TSTT unless they are placed optimally in the network. Regression results showed the most significant predictor variables to be different turning demands which may be true, however these most likely cannot captured linearly, as previously stated.

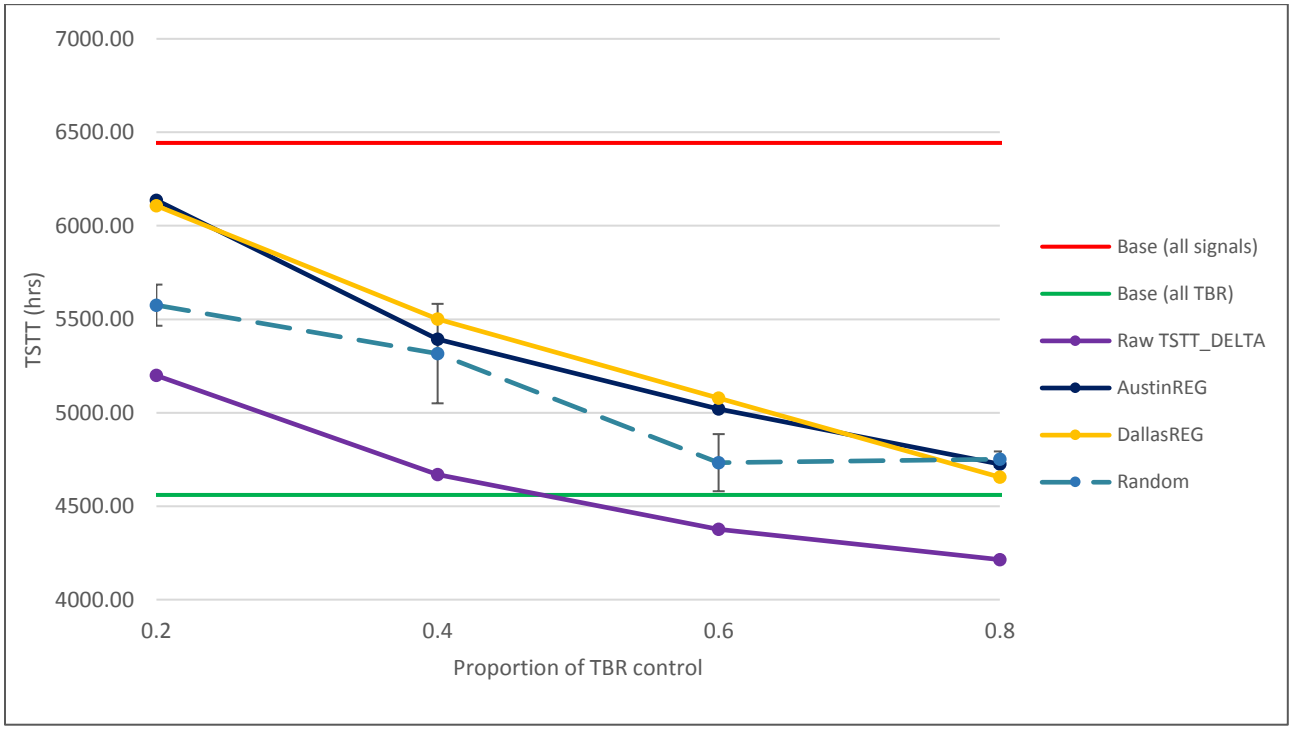


Figure 5.1 A summary of regression and raw data results in simulation across TBR control proportions

5.6. Genetic Algorithm Results

The regression identifies the intersections with characteristics desirable for a reservation. However, the regression model is incapable of identifying the optimal subset of intersections in a network that would produce the best benefit in terms of TSTT. As discussed earlier, previous studies found that a system with 100% reservations did not necessarily improve TSTT. The methodology explained in Section 5.2 was used to identify the subset of intersections in Austin that produce the best benefit when modeled as a reservation. Two sets of problems were solved using the GA: 1) identify the system-optimal allocation of reservations in the network that produce the best system-wide benefits, and 2) find the optimal allocation of TBRs that produces the best benefit when the number of TBRs is limited to 35 and 70 intersections. In all these cases, the GA was used in coordination with a DTA model to find the DUE solution. In the model used, an initial population of 200 individuals was used with a proportion of 0.75 of the population kept at each iteration. The probability of mutation was set at 0.7% per intersection and the convergence criterion was set at a maximum number of iterations of 100.

5.6.1. System-Optimal GA

For the system-optimal experiment, the initial population had 200 individuals with 50-50 split between traffic signal and reservations. The initial solution for the system-optimal genetic algorithm (SOGA) is much lower than that produced by the solution from the regression. In fact, the SOGA starts with a TSTT that is comparable and quickly reaches the levels produced when 100% of intersections are modeled as reservations. The algorithm terminates abruptly just as it begins to converge, as evidenced from Figure 5.2. This indicates that for a larger proportion of reservations, a larger convergence criterion must be used. Figure 5.3 shows that the proportion of reservations in the solution increases gradually with the number of iterations. The proportion of reservations in the solution when the GA terminated is at 0.9. This validates the paradox observed in earlier studies.

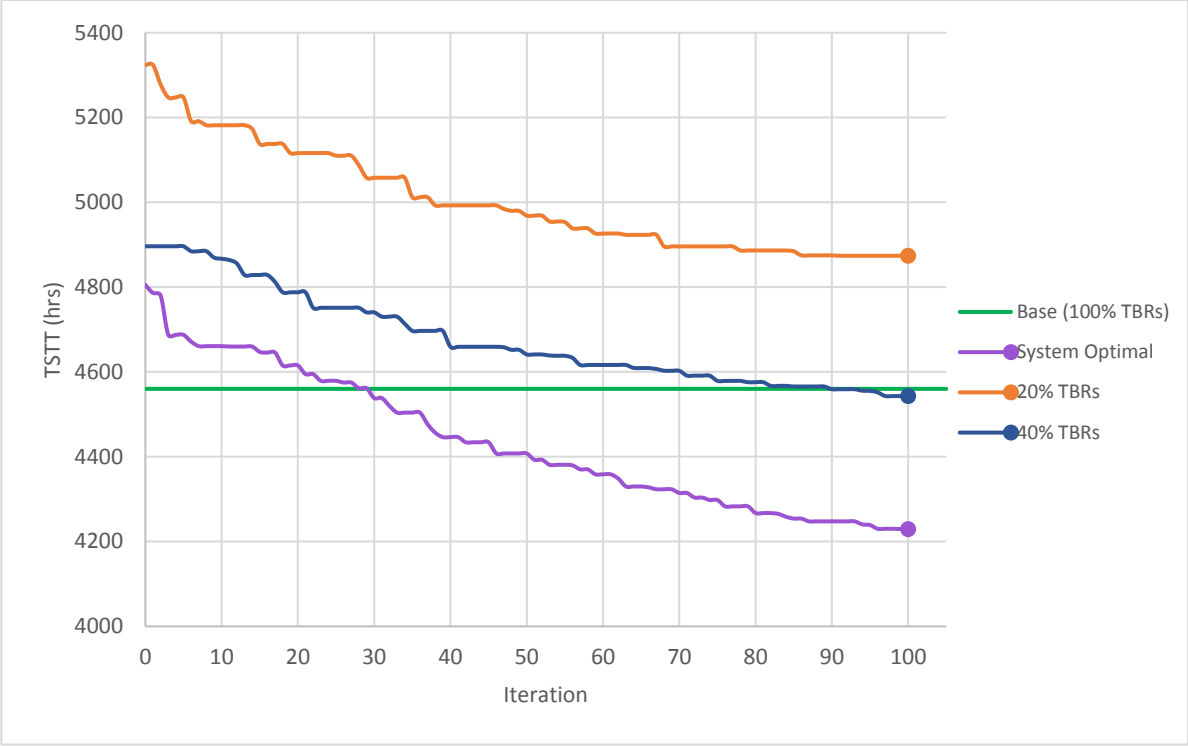


Figure 5.2 Variation of TSTT with number of iterations

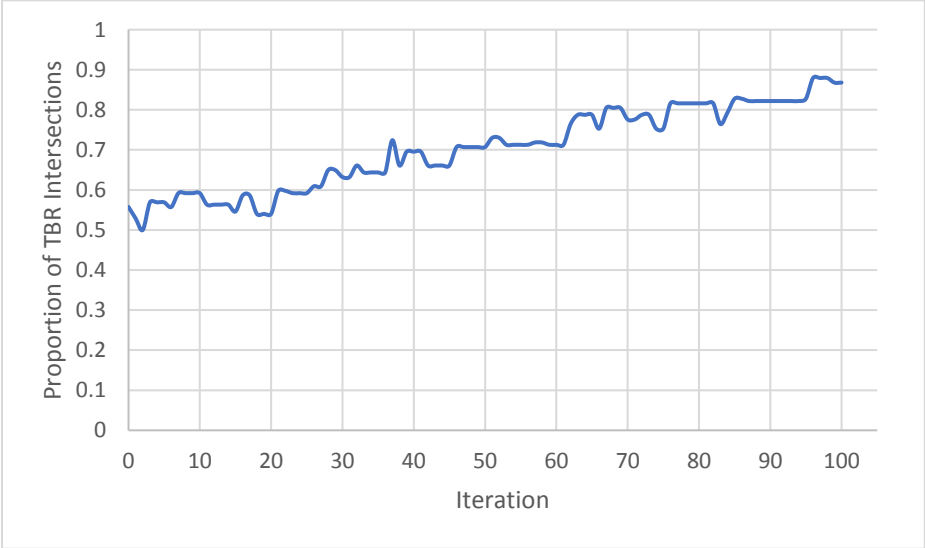


Figure 5.3 Variation of proportion of TBR with number of iterations

5.6.2. Limited TBRs GA

The limited TBR experiment is similar to the SOGA with the exception that there is now a limit on the number of reservations in the network. An initial population of 200 individuals was produced, containing 20% and 40% TBRs in each individual. The population was then evolved by crossing the parent generation to produce children. At each iteration, 0.75 of the parent generation were kept in the population. The probability of mutation was kept at 0.7% and the termination

criterion was kept at 100 iterations again. It is evident from Figure 5.2 that the TSTT drops with increase in the proportion of TBRs in the network. For the 40% TBR experiment, the GA evolves to produce results comparable with the default 100% TBR case. This indicates that a 40% TBR network can provide the benefit of a 100% TBR network if the location of TBRs is chosen wisely. This is also indicative of the paradoxes observed in earlier studies. Although the system-wide benefit of implementing TBRs increases with the proportion of TBRs in the system, the improvement is not uniform. From Figure 5.2, the SOGA reaches the 100% TBR solution at iteration 30. At this point, the proportion of TBRs in the model is 65% (approximated from Figure 5.3). However, the limited 40% TBR solution reaches achieves the 100% TBR solution by iteration 90. This implies that an optimal placement of a limited number of TBRs can achieve the system-wide benefit of a network with a higher number of TBRs.

5.6.3. Network Trends

In order to observe trends and find commonalities between methods of finding optimal TBR placement, experimental results are plotted onto maps using the ArcGIS software. In Figure 5.4, both the pre-regression base data and the Dallas-based regression solutions are mapped in the Austin network. Interestingly, both maps show most of the TBR intersections in series along corridors. Although both maps show similar clusters of TBR control, the regression map shows little to no reservations in the center of downtown, whereas the base data map shows short corridors of reservations along highly congested streets. Both show TBR corridors along Lamar Blvd, 1st Street, and 15th Street, all of which are known to be major arterial roads and/or become very congested in peak hour traffic.

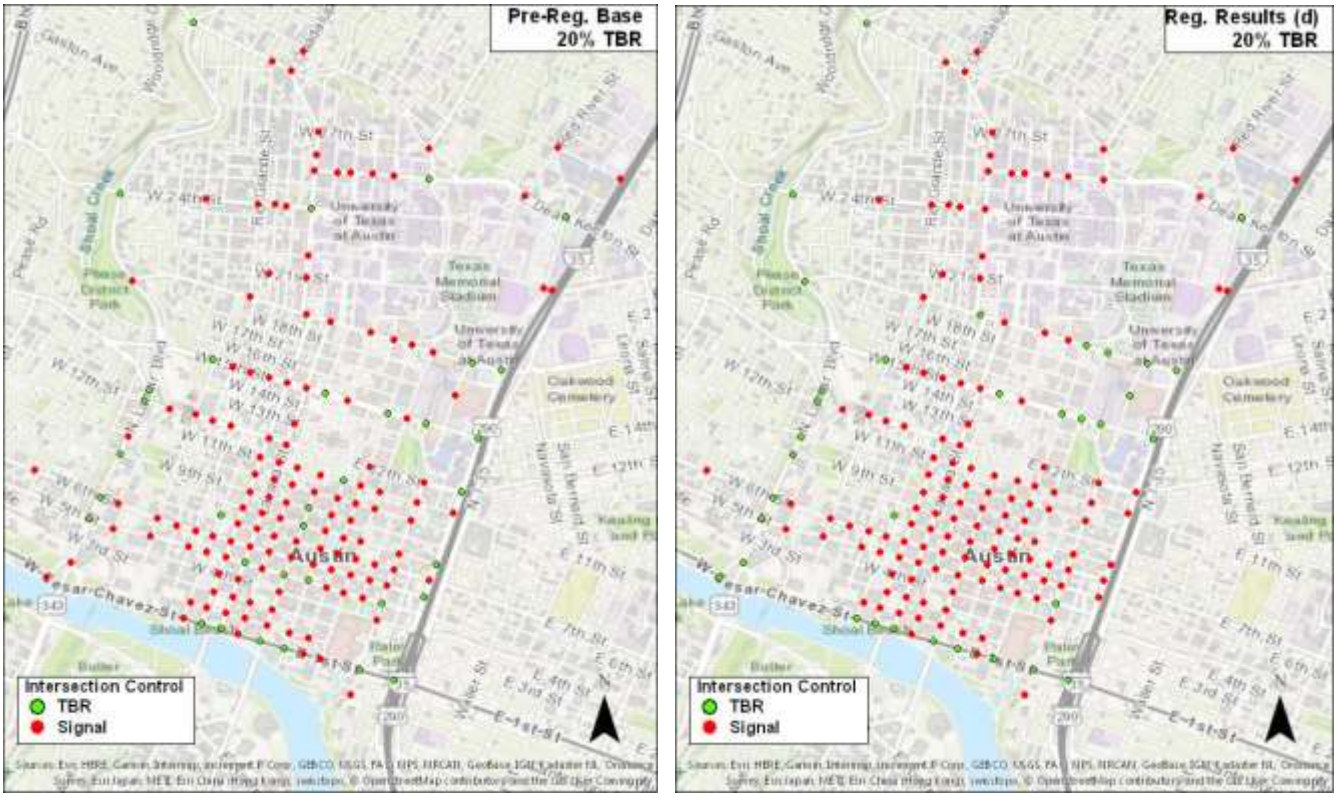


Figure 5.4 Intersection placements found by pre-regression raw input data (left) and the Dallas-based regression (right)

As Figure 5.5 demonstrates, the GA mapping at a 20% proportion of TBRs shows a similar trend in TBR corridors along major arterials and roadways such as Lamar Blvd and 1st Street. However, these corridors are not as continuous as those seen in the regression and base data results, with the exception of 15th Street. On 15th Street, the GA placed seven consecutive TBR intersections of which there are almost no directly adjacent signalized/TBR intersections. This aligns with the prediction that those intersections that are not within very close proximity to other minor intersections would do better as reservations, as there is no significant interruption of major flow through the corridor. The 20% TBR GA solution had the lowest TSTT of all 20% TBR proportion experiments. Similar trends are seen in the map of the 40% TBRs GA solution in Figure 5.5. The GA places TBR intersections in longer chains of multiple intersecting corridors, especially in the central downtown areas. Aside from some of the intersections in the middle of central downtown, most TBR intersections tend to have few signals within close proximity and are clustered mainly in straight corridors.

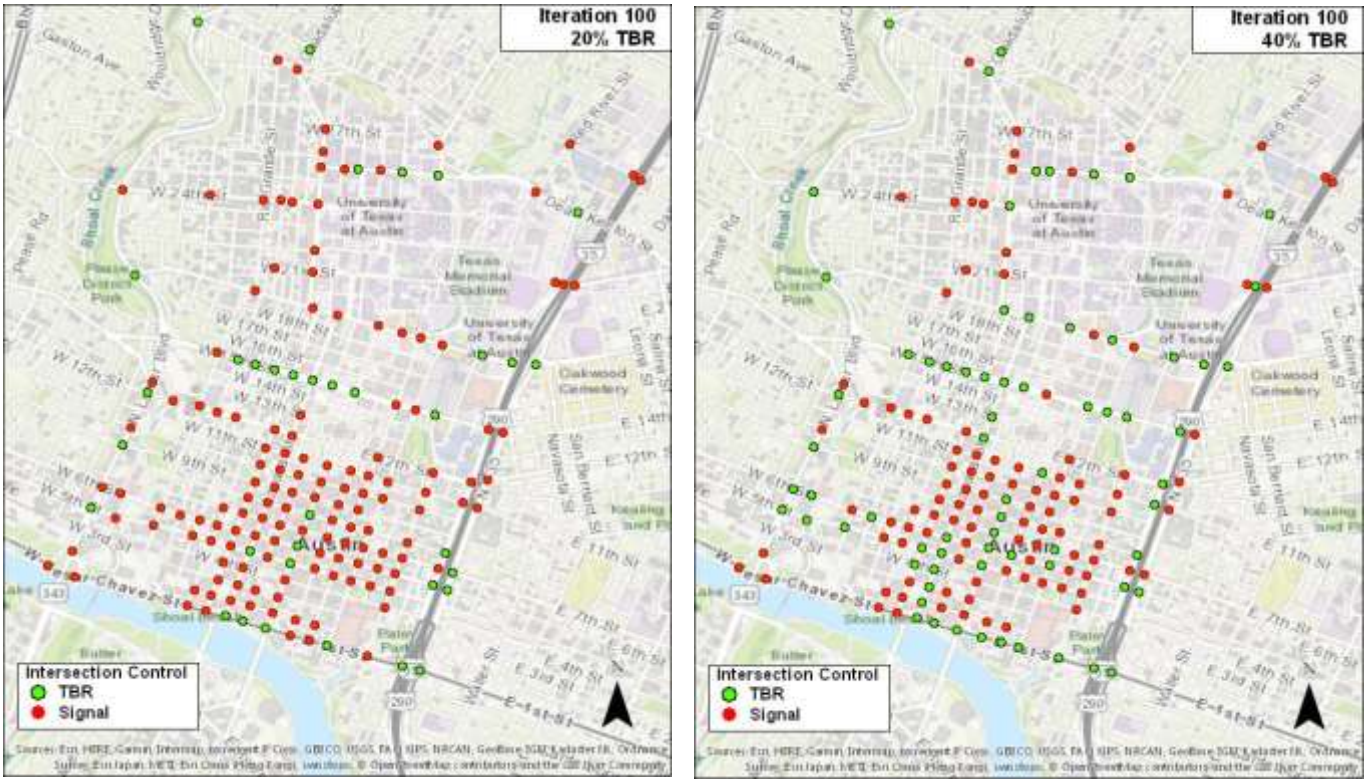


Figure 5.5 Intersection placements found by the limited reservation GA with 20% TBRs (left) and 40% TBRs (right)

As seen previously and in Table 5.4, the pre-regression raw data at 80% TBRs, when simulated, actually slightly outperformed the GA's system-optimal solution, which came to an 86% proportion of TBR control in the network. This could be because the GA was allowed too much freedom when adding to the proportion of TBRs, and skipped over a more optimal solution, as was found in the pre-regression base results. Although TSTT was approximately the same between the two, as seen in Figure 5.6, the orientation of signals across the network was different. The base data ranking of intersections left more clusters of signals crowded together compared to the GA, which had a wider spread of signal locations. Still, similarities were seen in signal placement at certain nodes, such as around the UT campus and in parts of central downtown. The same trend of TBR corridors is seen in both map visualizations of the two results, which gave the two lowest TSTT values seen across all experiments in the study.

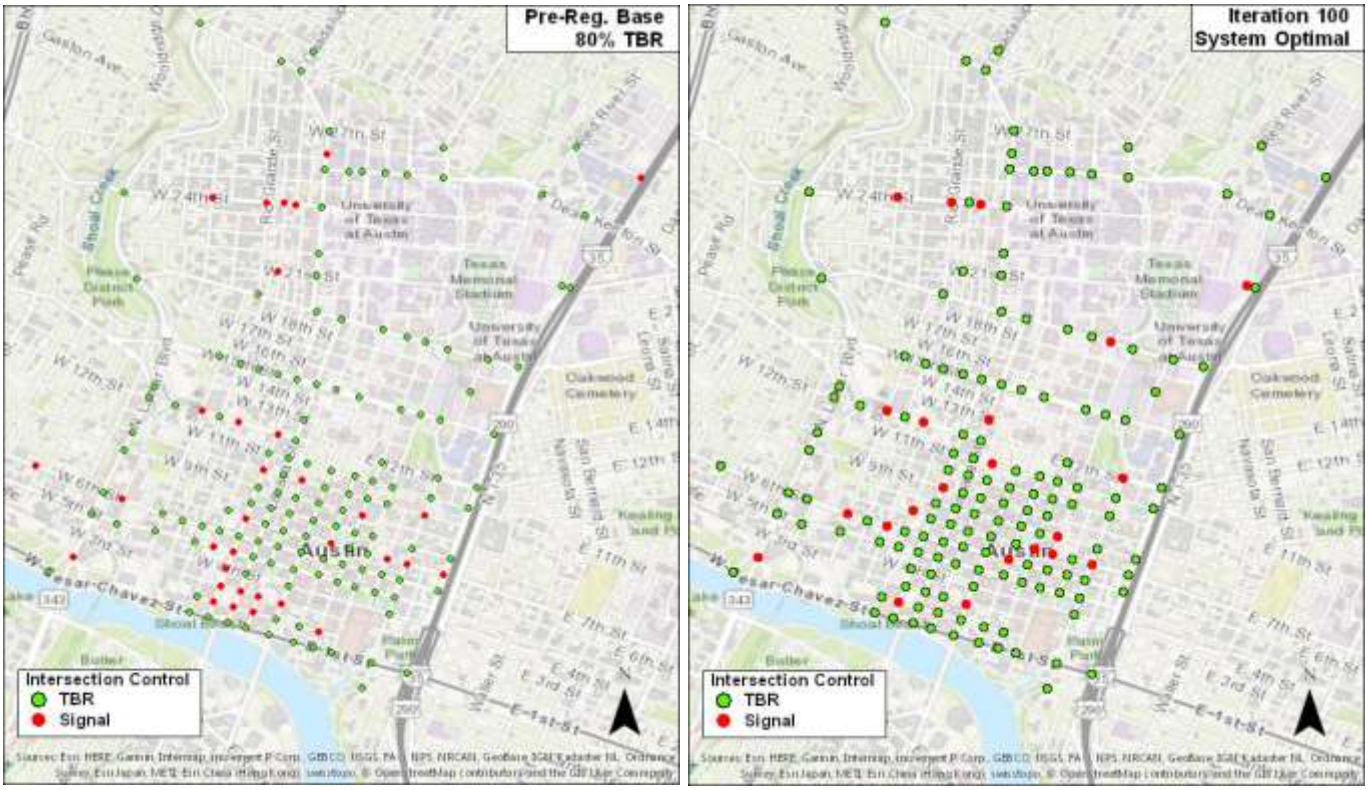


Figure 5.6 Intersection placements by pre-regression raw input data at 20% TBRs (left) and the SOGA solution

Table 5.4 A summary of used methods and TSTT

	TSTT						
Proportions	**Base Cases**	**Dallas-based Reg.**	**Austin-based Reg.**	**Raw TSTT_DELTA**	**20% TBR GA**	**40% TBR GA**	**SO GA**
0% TBR	6443.2						
20% TBR		6107.3	6136.8	5199.0	4873.9		
40% TBR		5501.0	5393.9	4669.4		4543.1	
60% TBR		5077.8	5020.7	4375.6			
80% TBR		4655.2	4725.1	4214.0			
100% TBR	4560.1						
86% TBR							4229.2

5.7. Conclusion

This chapter discussed three methods to identify the intersections where reservations would produce the best system-wide benefits in terms of TSTT in Austin. In this section, the three methods are discussed in detail with regard to their relative utility in addressing the problem at hand. The raw data approach and the regression model approach look to identify the best-performing intersections when modeled as reservations while the GA targets the best spatial orientation of a limited number of TBRs that produce the best system-wide benefits.

The raw data method ranks intersections in the decreasing order of $\Delta TSTT$ when the intersection is modeled as a reservation and as a traffic signal. Although this approach may be naïve in identifying a solution, it produces encouraging results in terms of TSTT. In fact, a 50% TBR raw data solution produces the TSTT observed by the base 100% TBR solution. This indicates that each intersection has an effect on TSTT and optimizing for each intersection can produce significant benefits in terms of TSTT. However, the raw data approach cannot identify the interdependencies within the network and, hence, cannot identify congestion due to queue spillback at intersections with low capacity links.

The regression approach alleviates some of the drawbacks of the raw data approach. The regression model helps quantify the characteristics that contribute to the performance of an intersection modeled as a traffic signal and as a reservation. This method can be used to characterize intersections as TBR or signal simply based on the properties of the intersection. However, the regression model did not produce satisfactory results in doing so. This is partly because the data may contain several interdependencies that the regression is incapable of identifying. It is entirely possible that data from the Dallas network was insufficient for the model to learn enough about the subtle differences between different intersections and, hence, the model did not generalize well enough when used to predict on the Austin network. This was evident from the different $\Delta TSTT$ observed between the raw data approach and the regression. Although the regression was able to generalize sufficiently well in most cases, for some intersections the regression produced a significantly different $\Delta TSTT$ than the raw approach. The regression and the raw data approach look to address a fundamental aspect of the problem—they try to classify an intersection as a TBR or a signal based on its characteristics. However, neither model does well enough to capture subtle interactions in the network.

The GA approach is different from the previous two approaches in that it captures the interactions in the network, such as congestion due to spillback from low-capacity intersections. However, the GA cannot identify classify intersections as TBR or traffic signal from the characteristics of each intersection and neither is it designed to do so. The GA approach is an optimization problem that identifies the spatial orientation of a given number of TBRs that produce the best system-wide benefits. This is evident from the discussion in Section 5.4.4. For example, among the 40% TBR solutions for the raw approach, regression approach and the GA, the GA produces the best TSTT. Thus, it is imperative to find a middle ground between the regression approach and the GA approach in solving the problem of identifying intersections to model as TBRs.

Trends in the networks for the optimal placement of smart intersections were primarily seen in the form of corridors of consecutive smart intersections. These corridors are highly congested roadways within the Austin network, including streets such as Lamar Boulevard and 1st Street, and seemed to be primarily targeted by all methods of intersection placement. Many networks, such as the regression resulting network and the raw data resulting network, appeared to have many similar features, including similar corridors in areas of the network in which few adjacent intersections were present. However, some slight differences in smart intersection placement caused a rather large difference in TSTT between the two networks. The placement of smart intersections in corridors with few adjacent intersections seems to be a trend seen enough to be considered a general rule. However, further experimentation would need to be done to confirm the relationship between smart intersection placement and clustering.

It would be wise to combine the relative benefits of the GA and regression approaches to identify a solution that produces the best system-wide benefits for a given budget of TBRs. One of the drawbacks of the GA model is that the initial population consists of individuals where TBRs are randomly distributed in the network. Due to this randomization, it is necessary to produce a large number of individuals in the initial population so as to ensure every possible orientation of the network is captured. It would be desirable to reduce the number of individuals in the initial population because such a change would drastically reduce the running time of the GA and would also ensure that the GA converges to the optimal solution quickly or at least converges to a local minimum quickly. It is worth exploring the possibility of using the results of the regression to identify the initial population of the GA. This would enable the GA to solve the problem quickly and also combine the benefits of both approaches.

Future work includes obtaining more data for a possibly different type of nonlinear regression as well as developing different optimal solution heuristics and algorithms tailored to the DTA model used. Including clustering into an algorithm or conducting isolated intersection experiments would be useful in deciding the relation between smart intersection proximity and system-wide congestion and would help to uncover possible network interdependencies between intersections and generalize them.

Chapter 6. Road Pricing

This chapter proposes Δ-tolling, a simple adaptive pricing scheme that requires only travel time observations and two tuning parameters. These tolls are applied throughout a road network and can be updated as frequently as travel time observations are made. Notably, Δ-tolling does not require any details of the traffic flow or travel demand models other than travel time observations, rendering it easy to apply in real time. The flexibility of this tolling scheme is demonstrated in three specific traffic modeling contexts with varying traffic flow and user behavior assumptions: a day-to-day pricing model using static network equilibrium with travel time functions; a within-day adaptive pricing model using the cell transmission model (CTM) and dynamic routing of vehicles; and a microsimulation of reservation-based intersection control for connected and autonomous vehicles with myopic routing. In all cases, Δ-tolling produces significant benefits over the no-toll case, measured in terms of average travel time and total travel time cost, while requiring only two parameters to be tuned. The team further examined the use of reinforcement learning for tuning the parameters used by Δ-tolling. Some optimality results are also given for the special case of the static network equilibrium model with BPR-style travel time functions.

6.1. Related Work

Road pricing has received considerable attention due to its potential to reduce congestion, and the economic fairness of charging users for the delays they cause to other travelers. It has long been established that in a static equilibrium setting, marginal tolls can eliminate the inefficiency associated with selfish routing (Pigou, 1920a; Beckmann et al., 1956). A detailed history along with practical aspects of congestion pricing can be found in de Palma and Lindsey (2011). However, such steady-state conditions rarely exist in practice. Changes in supply, demand, and other driver characteristics such as bounded rationality and value of time result in traffic that is dynamic both day-to-day and within-day. To control congestion in the presence of these factors, researchers have proposed a wide range of tolling models, based on different representations of traffic flow and different assumptions on the source of variability. In this section, we review these studies and highlight some of the gaps that will be addressed in this article.

6.2. Pricing Models Using Travel Time Functions

The simplest way to model congestion is by using travel time functions that predict the travel time on links as a function of its traffic volume. In this subsection, we review adaptive pricing articles built on this assumption. These studies can be broadly grouped into the following three categories: (1) pricing models with route switching behavior, (2) congestion pricing under supply and demand side uncertainty, and (3) trial-and-error methods for congestion pricing.

Friesz et al. (2004) proposed an optimal control formulation for finding tolls that minimize total travel time cost, while also achieving a minimum revenue target. Traveler choices were represented by an ordinary differential equation that corresponds to a *tˆatonnement* route switching

process. In a similar vein, Yang et al. (2007) suggested an adaptive tolling framework that reaches the system optimum state, assuming that travelers follow the proportional switch adjustment process (Smith, 1984). Tan et al. (2015) incorporated user heterogeneity in such day-to-day pricing models by proposing a multi-class flow evolution dynamic in which users with different values of time respond differently to current congestion levels. Adaptive tolls that minimize a weighted sum of system cost and time were sought. Guo et al. (2016) and Rambha and Boyles (2016) studied similar problems with slightly different objectives in discrete time settings where travelers' choices depended on the previous day's flow. While the former focused on the asymptotic behavior of the system under the rational behavior adjustment process (Guo et al., 2015), the latter supposed that users select routes according to the logit choice model and the tolls were used to minimize the average system travel time over an infinite horizon.

Marginal tolls are usually computed assuming fixed trip tables and fixed network parameters such as capacity and free flow travel time. However, when these supply and demand inputs are uncertain, the marginal prices can be non-optimal and in some cases may worsen the network performance (Gardner et al., 2008, 2010; Boyles et al., 2010). To address this issue, Gardner et al. (2011) defined six scenarios that take into account the information states of the system manager and travelers and suggested different optimization models in which responsive tolls are designed based on the realizations of the supply and demand. Recently, Rambha et al. (2017) extended the problem of finding the optimal tolls under supply side uncertainty to cases in which travelers respond to online information by changing their decisions en route.

A third class of adaptive pricing models is called *trial-and-error methods*. Tolls in these models vary across different days but are not set to address the variability in network congestion. Instead, the tolls are adaptive because the system manager may in reality not know the demand and delay functions. By levying certain "trial-and-error tolls" and updating them over different days using observed link volumes, provable convergence to system-optimal tolls can be guaranteed (Yang et al., 2004; Han and Yang, 2009; Yang et al., 2010). While travelers in these models are assumed to be aware of current day's tolls and react rationally, extensions in which travelers respond using day-to-day route dynamics also have been proposed (Ye et al., 2015). The Δ-tolling framework with travel time functions that will be described in Section 6.10 can be seen as a variant of the trial-and-error method.

6.2.1. Pricing Models Using Macroscopic Traffic Simulators

The earliest work studying dynamic congestion used Vickrey's (1969) bottleneck model. Road pricing also affects the number of trips and their departure times, and this elasticity was added by Arnott et al. (1993). Demand elasticity could be affected by alternative modes, so Danielis and Marcucci (2002) combined the bottleneck model with a railroad mode. Similarly, Huang (2002) studied a bottleneck model with a parallel mass transit alternative mode and compared the effects of several pricing schemes on congestion and overall system efficiency. Verhoef (2003) proposed a heuristic for adaptive tolling for dynamic traffic congestion in continuous time. In their study, van den Berg and Verhoef (2011) extended the results to continuous distributions of values of time

and found that congestion pricing could improve the total travel time cost of the majority of travelers even without returning toll revenues.

Adaptive tolling has also been widely studied in the context of managed lanes. Yin and Lou (2009) suggested a feedback control approach in which tolls are raised or lowered proportionally to the difference between the current and desired occupancy. They also proposed a self-learning approach in which the willingness to pay is estimated in an online manner and the lane choice is captured using a logit choice model, which was later extended to a multi-lane hybrid traffic flow model (Lou et al., 2011). Gardner et al. (2013) analyzed managed lane pricing using an additive logit model and an all-or-nothing assignment. Extensions that incorporate demand uncertainty (Gardner et al., 2015) and departure time choices (Boyles et al., 2015) were also studied.

For pricing at a network level, Carey and Srinivasan (1993) define dynamic externalities and tolls using exit flow functions as defined by Merchant and Nemhauser (1978). Wie and Tobin (1998) formulated optimal control programs assuming point-queue models. Wie (2007) suggested a bi-level model in which the lower level involves a simplified dynamic traffic loading mechanism. It was assumed that a subset of arcs can be tolled and the objective was to maximize net consumer surplus. However, the traffic flow models in these approaches do not capture queue spill-backs. Tsekeris and Voß (2009) review several studies that used bottleneck or point queue models, and therefore did not capture the effects of queue spillback. Waller et al. (2006) and Lo and Szeto (2005) showed that these traffic flow dynamics were important to the effectiveness of congestion pricing, and that ignoring them could result in tolls that *increased* congestion. Therefore, it is important to study tolling schemes such as Δ-tolling on mesoscopic or microscopic models that properly capture spatial propagation of congestion.

For more realistic flow models applied to large networks, such as the hydrodynamic model (Lighthill and Whitham, 1955; Richards, 1956), marginal costs are much more difficult to compute because of discontinuities in the flow model and congestion effects that transcend link boundaries. For such models it is not known how to reduce the problem beyond its fundamental bi-level form. The upper-level problem chooses the optimal tolls subject to route choice constraints, which form the lower-level problem. These route choice constraints are often in the form of dynamic traffic assignment (DTA) (Chiu et al., 2011), which itself is a difficult optimization problem. Such bi-level problems have been studied extensively for both static and dynamic flow models as *network design problems* (Farahani et al., 2013), and are known to be NP-hard even when both the upper-level and lower-level problems are convex. Consequently, they are typically solved using heuristics or meta-heuristics. Lin et al. (2011) formulated such a bi-level program in which route and departure time choices for a single destination network were captured with the CTM. A dual variable-based heuristic was used to solve the proposed MPEC (mathematical program with equilibrium constraints). Joksimovic et al. (2005a) and Joksimovic et al. (2005b) included both departure time and route choice in the lower level through a discrete choice model (stochastic user equilibrium). Although results used a small test network, tolls were observed to encourage travel on less congested routes or departure times. Ekstrm et al. (2016) devise a surrogate-based

optimization method in which a small number of dynamic cordon tolling schemes are tested on the Stockholm network using VisumDUE, a DTA tool.

When applied to real-world traffic networks, Δ-tolling can be classified as a within-day pricing scheme, as described in Sections 6.3 and 6.4. *Within-day pricing* varies the tolls at different times of day in response to (expected) congestion. Within-day strategies can affect both route choice and departure times because travelers may respond to congestion pricing by delaying their trip until a less congested time. While many within-day strategies set a predictable schedule of tolls that human drivers can react to, Δ-tolling assumes the use of route guidance software that can more quickly react to rapidly changing tolls. Furthermore, Δ-tolling is responsive to fluctuations that may be caused due to uncertain demand.

6.2.2. Pricing Models Using Microscopic Traffic Simulators

Very few researchers have used microscopic traffic simulators to study congestion pricing. The outputs of microscopic simulators cannot be analytically expressed, a challenge in optimizing control strategies. Thus, existing studies have used feedback mechanisms for computational tractability. The Δ-tolling approach proposed in this article can also be seen as a feedback mechanism much like ramp metering strategies such as ALINEA (Papageorgiou et al., 1991). Zhang et al. (2008) developed a feedback control theory-based tolling for high-occupancy/toll lanes in VISSIM to avoid the potential hysteresis problem and was later extended by Cheng et al. (2014) to include the effects of travel time reliability and income levels of users. However, their pricing model is fairly complicated and may be difficult to apply to other traffic flow models, and testing was limited to several connected freeway segments. Zheng et al. (2012) and Simoni et al. (2015) used a hybrid approach involving a microscopic simulator MATSim and a macroscopic fundamental diagram flow model to set cordon tolls in the city of Zurich. This model was used to calculate the aggregate density, from which the cordon tolls were inflated or deflated based on a linear feedback control strategy. Grether et al. (2008) also used MATSim along with an activity-based model that simulated users' plans, modes of travel, and values of time, but only evaluated fixed time-of-day dynamic tolls.

6.3. Framework

The Δ-tolling framework is designed to be widely applicable across a broad range of traffic and user behavior assumptions. The modeling framework has three major components:

- The traffic model.
- The travel time calculation model.
- The tolling model.

These three models make use of four variables:

- τ - the vector of tolls applied to each link.

- d - the travel demand, expressed in number of vehicles departing each origin towards each destination.
- X - the system state, a tuple of sets or vectors reflecting current traffic conditions corresponding to a particular traffic model.
- L - the vector of measured link travel times (travel time).

If the underlying network is represented by $G = (V,E,Z)$ where V and E are the sets of nodes and links, and $Z \subseteq N$ is the set of origins and destinations where trips start and end, then we have $\tau \in \mathbb{R}^{|E|}$, $d \in \mathbb{R}^{|Z|\times|Z|}$, and $T \in \mathbb{R}^{|E|}$. We use $l_{e\in E}$ to represent the travel time on link e; the same goes for x_e and τ_e (a link (e) might also be expressed as a pair i,j representing the link connecting node i with j). Similarly, $d_{s,t\in Z}$ is the demand originating at node s towards node t.

Each of these variables evolves over time according to the traffic flow, travel time, and tolling models, which are described next. The reader may find it useful to refer to Figure 6.1 during this discussion. Arrows in this figure reflect direct dependencies between the variables, as described below.

The **traffic flow** model M encompasses the routing decisions made by drivers, as well as the congestion effects caused by interactions amongst drivers. We express this relationship as

$$X^i = \mathrm{M}(X^{i-11}, L^{i-1}, d^i, \tau^i). \tag{6.1}$$

This equation represents the following potential dependencies: the system state at time step k may depend on the system state at the previous time interval (X^{k-1}); the measured link travel times at the previous time interval (L^{k-1}); the vehicles departing during time k (d^k), and the current tolls (τ^k).[1] Particular traffic flow models may not make use of all of these dependencies (for instance, the traffic state for Model A in Section 6.4 does not explicitly depend on previous time intervals) but they are included for generality. All of the models in this article assume that the last measured travel times L^{i-1} and current tolls τ^k are communicated to all vehicles (based on the assumption of CAV technology), but in principle the framework could allow for route choice decisions made without perfect knowledge of these. Specific examples of traffic models M and the corresponding system states X are given in Sections 6.4–6.6.

The **travel time calculation** model L maps the system state to link travel times used for tolling:

$$L^i = \mathrm{L}(X^i). \tag{6.2}$$

Although the travel times L^{k-1} may be part of the system state X^{k-1}, we include a separate dependence on L^{k-1} to allow for measurement errors, as might occur in practice if travel times are measured from sensors or probe vehicles in the field, or to allow for approximate travel time

[11] The time interval between time steps $(k, k + 1)$ may differ between models and between instances of the same model. Examples are given in the following sections.

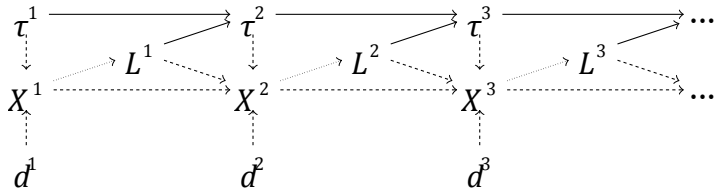
calculations, as are often obtained from simulation-based traffic models (such as Models B and C below). This separation also emphasizes that the toll calculation only relies on travel times and does not require additional information about the system state or demand. Examples of travel time models are also found in the sections that follow.

The **toll calculation** model for Δ-tolling is found in Equation 6.3:

$$\tau^i = (1 - R)\tau^{i-1} + R\beta(L^{i-1} - T^0) \tag{6.3}$$

where T^0 is the vector of link free-flow times,[12] and R and β are tuning parameters. Parameter β is the proportionality constant relating travel time and the toll value, while R is a weighting parameter that results in an exponential decay effect for tolls assigned in previous time steps. Both the R and β parameters must be tuned for a given network.

To summarize, the dependencies between the variables are intended to fit the following story: at time step i all users are informed with the tolls (τ^k) imposed during time interval k to $k+1$, as well as the travel times (L^{k-1}) measured at the end of the previous time interval. Given this data, users choose and follow a route leading from their current location to their destination that optimizes their utility. Based on the routes they choose, the system evolves to state X^k, and at the end of this time interval, the updated travel times (L^k) are measured. These updated travel times, along with the updated tolls, are fed-back to all users which, once again, re-optimize their route. Figure 6.1 presents a schematic illustrating these dependencies.



Dashed lines reflect the traffic model, dotted lines the travel time calculation model, and solid lines the tolling model. Note that the toll calculation only requires travel times, and no other details of the demand or traffic model.

Figure 6.1 Schematic for Δ-tolling framework

Note that allowing users to predict and react to future congestion and tolls can have a positive effect on the system as the convergence towards a user equilibrium would be faster or even instant. On the one hand, assuming such capabilities is reasonable from a practical standpoint (recurring congestion can be predicted). On the other hand, considering such capabilities significantly complicates our theoretical and empirical models. As a result they are not assumed in this study. Nonetheless, due to its relevance to the application of Δ-tolling, we intend to explore this topic in future work.

The following three sections show specific instantiations of this framework for varied traffic models: one inspired by day-to-day pricing in static traffic assignment, and two meant to represent

[12] Free flow travel time T^0 may change over time, e.g., due to weather conditions, and should be appropriately updated.

within-day pricing in dynamic models (the CTM and a microsimulator). The intent of these demonstrations is to study the performance of Δ-tolling across different modeling contexts, in contrast to many prior studies that demonstrate effectiveness only in a single traffic model (often the same model used to derive the tolling scheme). The focus of the presented experiments is on robustness of performance across widely varying models, rather than claiming that any of the specific models is the "right" one for any particular application.

6.4. Model A: Link Performance Functions

In the first model, the traffic model is based on link performance functions that map the flow on each link to its travel time (travel time). The presumption is that the intervals between successive time steps (and toll updates) are large enough that most trips can be completed during a single time interval, and that delays and flows can be well-approximated by deterministic, steady-state conditions. In this model, we assume that drivers choose routes to minimize their travel cost (time+tolls). Because delays are deterministic and the interval between time steps is assumed long, we further assume that CAV technology can choose routes rationally, leading to a Nash equilibrium in which each vehicle chooses a route with minimum cost (Wardrop, 1952).

6.4.1. Model Specification

For a source-target pair $(s,t) \in Z^2$, let Π_{st} denote the set of simple paths in G connecting source s to destination t, $\Pi_t = \cup_{s \in Z} \Pi_{st}$ the set of paths ending at t, and $\Pi = \cup_{t \in Z} \Pi_t$ the set of all network paths. Let d^k be the demand during time period k. The vector of path flows $h^k \in \mathsf{R}^{|\Pi|}$ is feasible if each element is nonnegative, and if $\sum_{\pi \in \Pi_{st}} h^k_\pi = d^k_{st}$ (each vehicle is assigned a path leading to its target). Each vector of path flows generates a vector of link flows $f^k \in \mathsf{R}^{|A|}$ given by $f^k_{ij} = \sum_{\pi \in \Pi : (i,j) \in \pi} h^k_\pi$. Furthermore, we assume that each link $(i,j) \in E$ is equipped with an increasing and differentiable travel time function $l_{ij}(f_{ij})$, giving the travel time on link (i,j) as a function of its flow f_{ij} alone. Let $X(d^k)$ denote the set of feasible link flows when the demand is d^k, that is, the set of link flow vectors corresponding to a feasible path flow vector.

For this model, the state vector is simply the vector of link flows:

$$X_k = f_k \tag{6.4}$$

and the traffic model M in Equation (6.1) is specified with the following formula:

$$f^k = \arg \min_{f \in \mathcal{X}(d^k)} \sum_{(i,j) \in E} \int_0^{f_{ij}} \left(l_{ij}(f) + \tau^k_{ij} \right) dx \,. \tag{6.5}$$

In this formula we assume that the value of time is homogeneous among all travelers, allowing us to choose units so that the tolls τ and travel time l can be directly added. The minimizer of the function on the right-hand side is known to satisfy the Nash equilibrium principle, and to be unique under the assumption that the l_{ij} are increasing. Note that there is no explicit dependence on X^{k-1} in Model A.

One special case of Model A occurs when the demand is stationary with time, $d^k \equiv d$. In this case, Model A can be thought of as a day-to-day tolling model, where the tolls are updated on sequential days as drivers make the same trips.

6.4.2. Scenario Specification

Every scenario simulated through Model A follows the following principles:

- Demand model - demand is modeled as a fixed amount of flow that needs to be routed between any two given nodes in the network at any time step.
- Vehicle model - in model A there is no notion of atomic vehicles; traffic is viewed as a set of infinitely divisible flows.
- Path assignment model - instead of assigning paths to vehicles, Model A assigns flows to paths. The vehicles comprising each flow are assumed to be self-interested and are assigned the minimal generalized cost path (travel time + tolls). Such a policy leads to the Nash equilibrium that is defined by Equation 6.5.

6.4.3. Theoretical Results

One advantage of Model A is that the analytical form of the traffic model (6.5) is amenable to mathematical analysis. In particular, we are able to show several optimality results if demand is stationary with time and the link performance functions are of the form specified by the Bureau of Public Roads (BPR),

$$l_{ij}(f_{ij}) = T_{ij}^0 \left(1 + A\left(\frac{f_{ij}}{u_{ij}}\right)^B \right. \qquad ! \tag{6.6}$$

where T_{ij}^0 is the free-flow travel time, u_{ij} the practical capacity, and A and B are shape parameters assumed uniform throughout the network. In this section, we mean "optimality" in the sense of minimizing the average travel time, which is proportional to $\sum_{(i,j)\in E} f_{ij} l_{ij}(f_{ij})$. This is a convex function of the link flows, so optimal link flows f_{ij}^* exist and are unique.

The next results make use of the following well-known facts. (Both can easily be shown by writing the optimality conditions of the associated convex minimization programs.)

> ***Fact 1.*** *Let $f^k \in X(d^k)$ and let h^k be any feasible path flow vector generating f^k. The flows f^k satisfy (6.5) if and only if every positive component of h^k corresponds to a path whose generalized cost (the sum of $l_{ij}^k + \tau_{ij}^k$ along its links) is minimal for its source-target pair.*

> ***Fact 2.*** *Let $f^k \in X(d^k)$ and let h^k be any feasible path flow vector generating f^k. The flows f^k are optimal if and only if every positive component of h^k corresponds to a path whose marginal cost (the sum of $l_{ij}^k + f_{ij}^k l_{ij}^0 (f_{ij}^k)$ along its links) is minimal for its source-target pair.*

The first result shows that the Δ-tolling rule is equivalent to marginal cost pricing if the parameter β is chosen correctly:

Proposition 1. *If the link performance functions are of the BPR form and* $\beta = B$, *then* $\beta(l_{ij} - T_{ij}^0)$ $= f_{ij} l_{ij0}(f_{ij}) = BAT_{ij0}(f_{ijk}/u_{ij})_B$.

Proof. Routine. □

Next, we show that if the tolls do not change from one time step to the next, then the resulting link flows must be optimal. Furthermore, the system has reached a stable state, and the optimal state will persist for future time iterations. In other words, if the tolls are stable, the flows are optimal and stable.

Proposition 2. *Let the link performance functions have the BPR form, and let demand be stationary with time. If* $\beta = B$ *and* $\tau^k = \tau^{k+1}$ *for any time interval k, then* f^k *is optimal, and furthermore* f^K *is optimal for any* $K > k$.

Proof. Stationary demand implies that the set of feasible *f* and *h* are stationary. By the Δ-tolling update rule (6.3), if $\tau^k = \tau^{k+1}$ then we must have $\tau^k = \tau^{k+1} = \beta(L^k - T^0) = BAT^0(f^k/u)^B$, where this vector equation holds component-wise. Let h^t be a feasible path flow vector generating f^k. By Fact 1, every positive component of h^k corresponds to a path whose sum of $l_{ij}^k + BAT_{ij}^0(f_{ij}^t/u_{ij})^B$ along its links is minimal. But by Proposition 1, this implies that every component of h^t corresponds to a path whose sum of $l_{ij}^k + f_{ij}^k l'_{ij}(f_{ij}^k)$ is minimal, and thus f^k is optimal by Fact 2. Furthermore, since the minimizer of (6.5) is unique, $\tau^k = \tau^{k+1}$ implies $f^k = f^{k+1}$, and thus $\tau^K = \tau^k$ and $f^K = f^k$ whenever $K > k$. □

The converse of Proposition 2 is not true, because the vector of tolls which produce a particular flow *f* under the mapping (6.5) is not unique, and only the marginal-cost tolls corresponding to Proposition 1 are fixed points of the toll update rule (6.3). If f^k is optimal but generated by a different toll vector, we will not have $\tau^k = \tau^{k+1}$. Nevertheless, we can show that if f^k is optimal, the flows are stable in subsequent iterations, even if the tolls still change.

Proposition 3. *Let the link performance functions have the BPR form, and let demand be stationary with time. If* $\beta = B$ *and* f^k *is optimal at any time interval k, then* f^K *is optimal whenever* $K > k$.

Proof. Let h^t be any feasible path flow vector generating f^k. By Fact 1, every positive component of h^k corresponds to a shortest path with link weights $l_{ij}^k - \tau_{ij}^k$. Since these flows are optimal, Fact 2 and Proposition 1 imply that these are also shortest paths with link weights $l_{ij}^k + \beta(l_{ij}^k - T_{ij}^0)$. That is, for each node *i* and source *s*, there exist node potentials π_i^s and $\rho^s{}_i$ such that

$$\pi_i^s + l_{ij}^k + \tau_{ij}^k \geq \pi_j^s \tag{6.7}$$

$$\rho_{si} + l_{ijk} + \beta(l_{ijk} - T_{ij0}) \geq \rho_{sj} \tag{6.8}$$

for each link $(i,j) \in E$ and source $s \in Z$, with equality holding along all of the paths with positive flow at h^k. Multiplying inequality (6.7) by $1 - R$, inequality (6.8) by R, and adding, we have

$$(1-R)\pi_i^s + R\rho_i^s + l_{ij}^k + \tau^{k+1} \geq (1-R)\pi_j^s + R\rho_j^s, \tag{6.9}$$

implying that $(1 - R)\pi + R\rho$ form valid node potentials for the new toll vector τ^{k+1}, with equality holding for exactly the same links and origins as before. Thus, the shortest paths with respect to $L^t + \tau^{k+1}$ are the same as those with respect to $L^k + \tau^k$, and f^k remains optimal for f^{k+1}. The argument can be repeated for any $K > k$.□

6.4.4. Experiments and Results

Model A was implemented in C using Algorithm B (Dial, 2006a) to solve the equilibrium subproblem. It was tested on two city networks, representing the cities of Sioux Falls, SD and Austin, TX. The Sioux Falls network is a standard test instance in the transportation network literature (Bar-Gera, 2014), with 76 links, 24 nodes, and 360,600 trips spanning 24 hours. The Austin network represents the central business district of the city, and contains 1247 links, 546 nodes, and 62,836 trips over a two-hour morning peak period. Additional details on the Austin network can be found in Levin et al. (2015a). Figure 6.2 shows schematics of both networks.

In both networks, the link performance functions are BPR functions, using the standard values of the shape parameters: $A = 0.15$ and $B = 4$. For the experiments in this article, the time intervals k were interpreted as subsequent days, so demand was assumed stationary and the experiment represents a day-to-day pricing scenario.

Because Model A assumes fixed demand and homogeneous travelers, we can use the average travel time $ATT^k = (f^k \cdot L^k)/(d \cdot 1)$ as a performance metric. In this demonstration, the weights $R^k = 1/(k + 1)$ were chosen, effectively setting the toll during time step k to the average of the daily "target" tolls $\beta(L^k - T^0)$. This choice of R^k value was inspired by the method of successive averages (Liu et al., 2009).

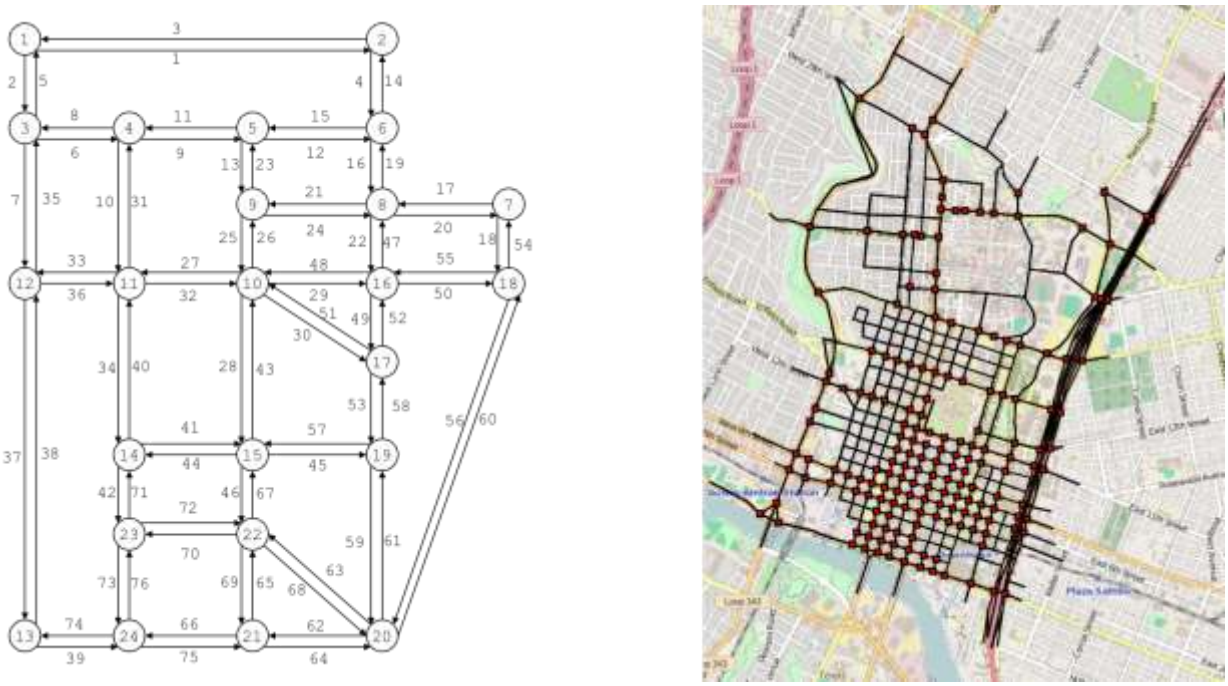


Figure 6.2 Sioux Falls (left) and Austin (right) networks

Table 6.1 shows the steady-state average travel time for both networks as the sensitivity parameter β varies. For both networks, with the above choice of R^k, fairly rapid convergence was obtained to a steady state.[13] For β values of 1, 2, 4, and 8, convergence on Sioux Falls required 95, 27, 11, and 94 iterations, respectively. For Austin, these values required 24, 16, 27, and 42 iterations to converge, respectively.

Table 6.1 Average travel time (minutes) at UE for different β values using Models A (Rt = 1/(t + 1)) and B (R = 10−4).

Note that for Model A with β = 4 the UE and SO align (this is the provable SO for model A). The SO for Model B is not applicable.

Scenario	No tolls	SO	$\beta = 1$	$\beta = 2$	$\beta = 4$	$\beta = 8$
Sioux Falls						
Model A	20.74	19.95	20.09	19.98	**19.95**	19.96
Model B	24.74	NA	20.28	20.08	**19.92**	20.26
Downtown Austin						
Model A	21.92	21.78	21.81	21.79	**21.78**	21.79
Model B	20.67	NA	16.06	**15.64**	15.82	17.39

Over the range of β values tested, the Δ-tolling strategy always reduced average travel time from the no-toll value. When β was set equal to the B exponent in the link performance functions, the travel times were the lowest observed, and in fact correspond to the system-optimal solution, as suggested by Proposition 1. Note that the performance of Δ-tolling seems to be insensitive to the

[13] The system is said to converge to a steady state if the change in average travel time between successive time steps was less than a tenth of a millisecond.

chosen β value as *beta* values twice as big (8) or small (2) from the optimal (4) result in a system performance, which is almost identical to the optimal one.

6.5. Model B: Cell Transmission Model

Model B implements the Δ-tolling framework in the CTM developed by Daganzo (1994a, 1995a) as a discrete, explicit solution method for the hydrodynamic theory of traffic flow proposed by Lighthill and Whitham (1955) and Richards (1956). CTM is frequently used in DTA. The time step used in this model is typically short, on the order of a few seconds. When used with Δ-tolling, this allows for a truly adaptive toll that can be updated based on traffic conditions. Because the tolls are updated frequently, we believe that the equilibrium behavior assumed in Model A is not appropriate. Rather, we assume that CAVs dynamically receive updates of travel times and tolls, and may alter their route to the destination in response to receiving such information. This update is based on the currently reported travel times and tolls when they reach the diverge point.[14]

6.5.1. Model Specification

The CTM divides each link into a set of cells, each of length equal to the distance a vehicle would travel in one time step at free-flow conditions. Cells result from a fixed division of a link into discrete segments. The length of each segment corresponds to the distance a vehicle would travel on that link at free-flow speed in one time step. This choice of cell length ensures stability of the CTM (it satisfies the Cournout-Friedrich-Lewy conditions for the underlying system of partial differential equations). Let C be the set of cells in the entire network, and for a given cell $c \in C$, let $C^+(c)$ denote the set of cells immediately downstream of c, and $C^-(c)$ the set of sells immediately upstream of c. For cells in the interior of a link $|C^+(c)| = |C^-(c)| = 1$, whereas if c is at the upstream end of a link we may have $|C^-(c)| > 1$ (if there are multiple links incident from upstream) and if c is at the downstream end we may have $|C^+(c)| > 1$ (if there are multiple links adjacent downstream). Let $x_{c,\pi}^k \in \mathbb{V}$ denote the number of vehicles in cell c at the start of time interval k which are currently following path $\pi \in \Pi$, and $x_c^k - \sum_{\pi \subset \Pi} x_{c,\pi}^k$ the total number of vehicles in the cell. Based on these values, the CTM respectively defines the *sending flow* S_c^k and *receiving flow* R_c^k as the maximum number of vehicles which could possibly exit cell c during time step k, and the maximum number of vehicles which could possibly enter cell c during time k. If we denote Q_c as the capacity of cell c, V_c the maximum number of vehicles that can physically occupy cell c, and δ the ratio between the maximum backward shockwave speed and free-flow speed, common expressions for the sending and receiving flow are

$$S_c^k = \min\{x_c^k, Q_c\} \tag{6.10}$$

and

$$R_c^k = \min\{\delta(V_c - x_c^k), Q_c\}. \tag{6.11}$$

[14] Particularly, vehicles do not anticipate future changes in travel conditions, nor the impact of receiving future information, cf. Waller and Ziliaskopoulos (2002); Boyles (2009). While such computations are more involved, they may be feasible with CAV technology, and would be an interesting topic for future study.

See Daganzo (1995a,c) for additional details.

The update rules for the CTM (the traffic model M) involve the auxiliary variables y_{cd}^k representing the number of vehicles moving from cell c to cell $d \in C^+(c)$ during time step k. If c and d belong to the same link, we have

$$y_{cd}^k = \min\{S_c^k, R_d^k\}. \tag{6.12}$$

If c and d belong to different links, y_{cd}^k is calculated using various *intersection models* representing traffic behavior at diverges, merges, traffic signals, roundabouts, or other intersection types; see Tampère et al. (2011) for discussion of intersection model desiderata and some examples. The simulations reported in this section use intersection models reflecting traffic signals.

Once the y^k values are calculated, cell occupancies update in the natural way:

$$x_c^{k+1} = x_c^k + \sum_{b \in C^-(c)} y_{bc}^k - \sum_{d \in C^+(c)} y_{cd}^k \tag{6.13}$$

with the path-disaggregated $x_{c,\pi}^k$ values updated according to the first-in, first-out principle.

The travel time model for l_{ij}^k calculates the average time spent on link (i,j) by the vehicles which most recently exited the link, based on the difference between their entry and exit times. This results in a slight lag in the travel times used for Δ-tolling and for the adaptive routing procedure, since link travel times for vehicles are not computed until they leave the link, even though the routing and tolls are based on decisions made as vehicles enter a link.

6.5.2. Scenario Specification

Every scenario simulated through Model B follows the following principles:

- Demand model - demand is given as the number of vehicles originating at node n_1 at time step k and are assigned a destination n_2 for any $n_1, n_2 \in V$ and $k > 0$ combination.
- Vehicle model - each vehicle is affiliated with a *value of time*[15] (VOT). Vehicles seek to minimize their generalized cost that is defined as travel time + tolls × value of time.
- Path assignment model - let l_π^{k-1} be the sum of travel times along path π during time step $k-1$ and let τ_π^k be the sum of tolls along π during time step k. When reaching a diverge node n at time step k all paths (Π_{nt}) leading from n to target t are considered. The vehicle in question is assigned the path $\text{argmin}_{\pi \in \Pi_{nt}} cost(\pi) = \{\tau_\pi^k + l_\pi^{k-1} \cdot VOT\}$. An additional rule was added to prevent gridlock problems, which can arise in dynamic traffic models when a cycle of links is at jam density: if a vehicle is unable to enter a link because its receiving

[15] Value of time represents the monetary value of a single unit of time. It is used to map time into monetary cost.

flow is zero for more than 96 seconds,[16] the vehicle is assigned the least cost path to its destination that avoids that link, if such a path exists.

6.5.3. Experiments and Results

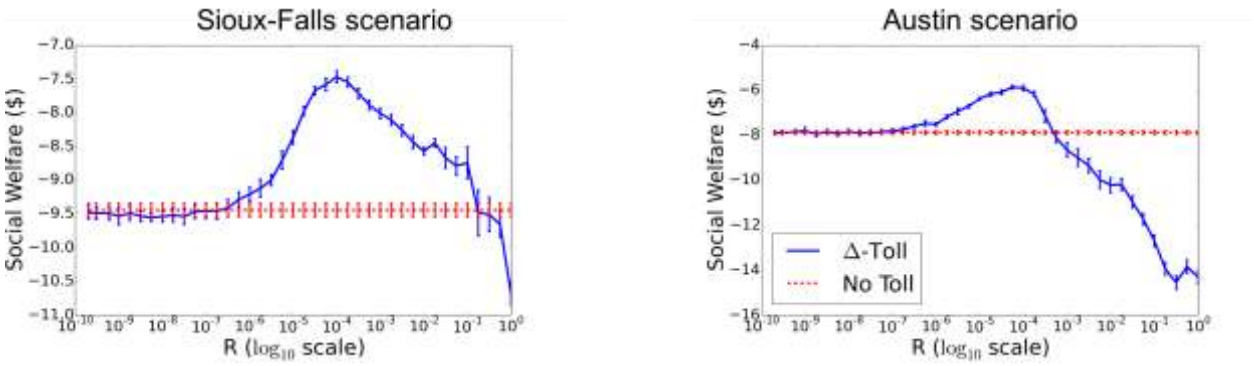
For running Model B we used the DTA simulator (Chiu et al., 2011) implemented in Java. Model B was tested on the Sioux Falls and Austin networks also used for the Model A results. A few changes were needed to accommodate the differences in time scale and modeling assumptions between Models A and B. Because the original Sioux Falls demand was specified over 24 hours, the demand was modified to fit a 3-hour experiment more suitable for DTA, including 28,835 trips over this time period. Both networks also required traffic signal timings. The Austin network data was originally used for DTA and contained real-world signal data. The Sioux Falls network does not include this data, so we generated an artificial timing based on Webster's formula (1958) for signal timing. The variations in departure rates over time for these scenarios can be seen as the solid black line in Figure 6.4. Whenever a vehicle is loaded onto the network, it is assigned a value of time randomly drawn from a Dagum distribution with parameters $\hat{a} = 22020.6$, $\hat{b} = 2.7926$, and $\hat{c} = 0.2977$, reflecting the distribution of personal income in the United States; see Lukasiewicza et al. (2012).

The average travel times for Model B for particular values of R and β are shown in Table 6.1. Since Model B allows heterogeneity in travelers' values of time, in addition to evaluating average travel times, we also evaluate the *total travel time cost*, defined as the weighted sum of each traveler's travel time according to his or her value of time.[17] If A is the set of vehicles, and vehicle a experiences a travel time of l_a and has a value of time of α_a, total travel time cost is defined as $\sum_a l_a \alpha_a$. The tolls are not included in the calculation of total travel time cost, because we assume that toll revenues are transfer payments which remain internal to society. This assumption was made implicitly in the use of average travel time as the metric for Model A.

Figure 6.3 shows the effects of Δ-tolling on total travel time cost (denoted social welfare) as the responsiveness parameter R varies. The red series indicates the no-toll scenario used as a benchmark, while the blue series shows total travel time cost under the Δ-tolling regime. Each data point represents the average of ten scenario runs, and the bands represent 95% confidence intervals around each point. Scenario runs differ from each other in the value of time that is (randomly) assigned to each vehicle. For the experiments in this plot, a fixed value of $\beta = 4$ was used. This value was optimal in Model A for its assumption of BPR-type delay functions, but this choice is not necessarily optimal for Model B, which uses a dynamic network loading procedure. (As shown in Table 6.1, in Austin, a slightly better performance was obtained with $\beta = 2$ compared to $\beta = 4$.) In this plot, note that extreme R values are less effective, whether high or low: when R is near 1 tolls oscillate rapidly (worsening performance over the no-toll baseline), and when R is near zero the tolls have little impact. The best performance occurred in a narrow band around $R \approx 10^{-4}$, yielding increases of 26% and 33% in total travel time cost for the Sioux Falls and Austin

[16] This value was chosen by trial-and-error and resulted in the best performance.
[17] For Model A, where value of time is homogeneous, total travel time cost is directly proportional to the total travel time.

scenarios, respectively. Indeed, for any given β value, near-optimal tolls cluster around a single order of magnitude of R values.

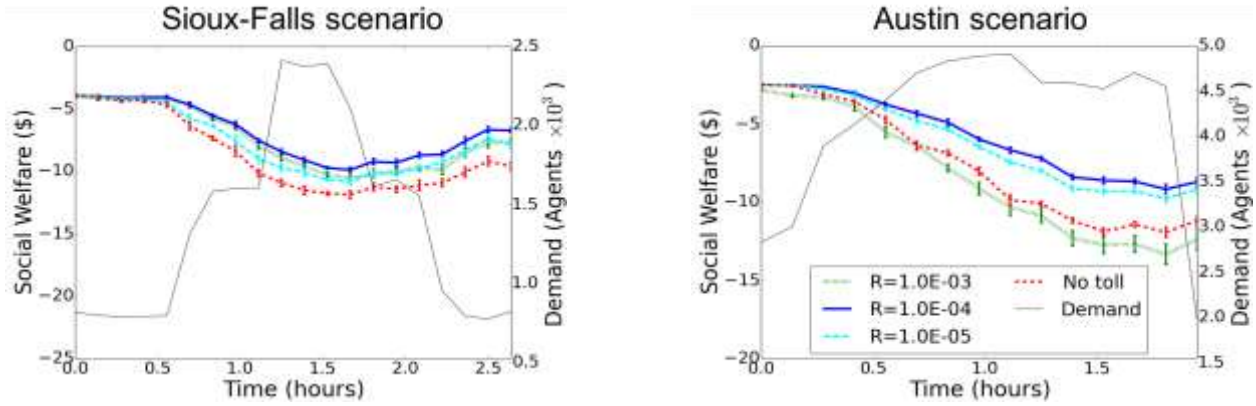


Note: The legend is identical in both plots.

Figure 6.3 Total travel time cost (social welfare) in Sioux Falls (left) and Austin (right) as responsiveness parameter R varies using Model B

Figure 6.4 shows the impacts on total travel time cost for travelers departing at different times during the simulation for R values in the range 10^{-5}–10^{-3}. All series in this plot correspond to β = 4. This plot shows the superior performance of $R = 10^{-4}$ in a different way and indicates that the benefits from tolling increase throughout the peak period, and that the onset of congestion is delayed. In the Sioux Falls scenario, which includes the end of the peak period, we see that the recovery from congestion occurs earlier as well.

Figure 6.5 shows performance of the tolling scheme as both R and β vary, depicting the difference in total travel time cost (in percentage) compared to a scenario where no tolls are applied (a value of 126, for instance, correspond to a 26% increase in total travel time cost). For any fixed value of one parameter, there is a near-optimal value for the other parameter. This observation leads us to suspect that there are dependencies between the two parameters. We leave exploring such dependencies for future work. Nonetheless, this observation justifies the use of a single β value in the results described earlier.



Note: The legend is identical in both plots.

Figure 6.4 Total travel time cost (social welfare) over time in Sioux Falls (left) and Austin (right), R varies using Model B

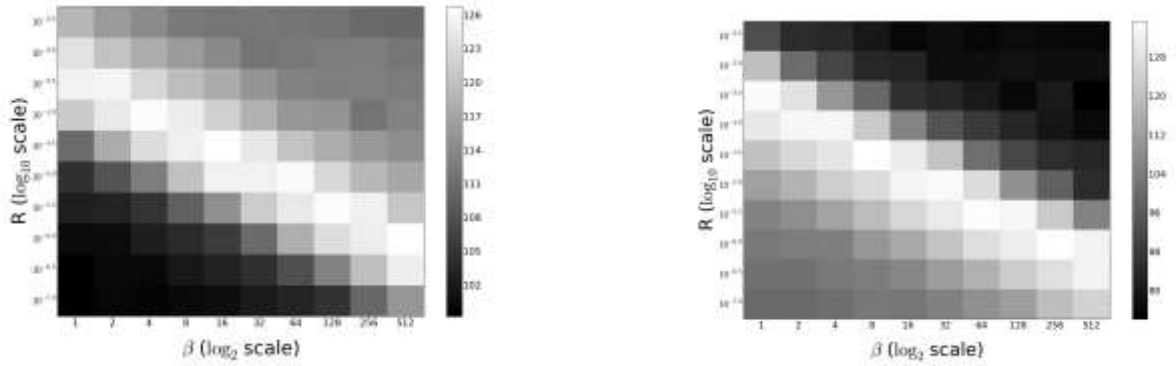


Figure 6.5 Heat maps showing the difference (in percentage) in total travel time cost (social welfare) compared to the no-tolls scenario for different R and β values in Sioux Falls (left) and Austin (right) (Model B)

6.6. Model C: Microsimulation

AIM (Autonomous Intersection Manager) is a traffic microsimulator first developed by Dresner and Stone (2004) to model the flows of CAVs at intersections where priority is granted by reservation, rather than with signals.

AIM provides a multiagent framework for simulating autonomous vehicles on a road network grid, and it presents a much more detailed traffic flow model than Models A or B. The AIM simulator uses two types of agents: intersection managers and driver agents. Intersection managers are responsible for directing the vehicles through the intersections, while the driver agents are responsible for controlling the vehicles to which they are assigned. To improve the throughput and efficiency of the system, the driver agents "call ahead" to the intersection manager and request a path reservation (space-time sequence) within the intersection. The intersection manager then determines whether or not this request can be met. If the intersection manager approves a driver agent's request, the driver agent must follow the assigned path through the intersection. On the other hand, if the intersection manager rejects a driver agent's request, the driver agent may not pass through the intersection but may attempt to request a new reservation. AIM has been used in various studies on reservation-based intersection control: Dresner and Stone (2006, 2007) studied variants of the reservation protocol that provided intersection access to human drivers through an occasionally activated traffic signal, and Fajardo et al. (2011) found that reservations had lower delays than optimized traffic signals for a symmetric intersection. Figure 6.6 shows a typical snapshot of simultaneous vehicle flow at a congested intersection.

Figure 6.6 The AIM simulator depicting a reservation-based intersection in operation

Link travel times L^{k-1} are estimated as an average of the time spent on each link by the vehicles most recently exiting. The scenario specifications for this model are identical to those specified for model B. As with Model B, due to the frequency of updates there is no presumption that an equilibrium is reached when vehicles choose routes.

Unlike Models A and B, the microsimulation environment is not well-suited to explicit mathematical description. The state vector can be defined as $X^k = (A^k, w^k)$, where A^k is the set of vehicles on the network at the start of time step k, including associated information such as their value of time and position, velocity, and acceleration in the network, and w^k is the set of intersections and associated information at time k, such as the trajectories of scheduled reservations. This information is updated according to the rules described above.

6.6.1. Experiments and Results

For running Model C we used the AIM4 microsimulator (http://www.cs.utexas.edu/aim/aim4sim/aim4release-1.0.3/aim4-root/docs/install.html). AIM4 is unable to model large networks of the type used for Models A and B, because of the level of detail in its representation of agent behavior (both vehicles and intersection reservations). Hence, the Sioux Falls and Austin scenarios are intractable within AIM. Figure 6.7 shows the 3×3 grid network used for these experiments. Vehicle agents are generated randomly according to a Poisson process, at a mean rate of 500 vehicles per hour per incoming lane. Each vehicle is assigned either to destination D1 or D2. The network also includes alternative destinations for vehicles headed to either of these destinations. Alternative destinations are used to simulate route choice effects on a network much smaller than the city networks used in Models A and B. These alternative destinations, marked as A1 and A2 in Figure 6.7, are associated with a time penalty if vehicles leave the network through them instead of their original destination. Vehicles may opt for a path ending at an alternative destination when performing the A^* search when arriving at each intersection.

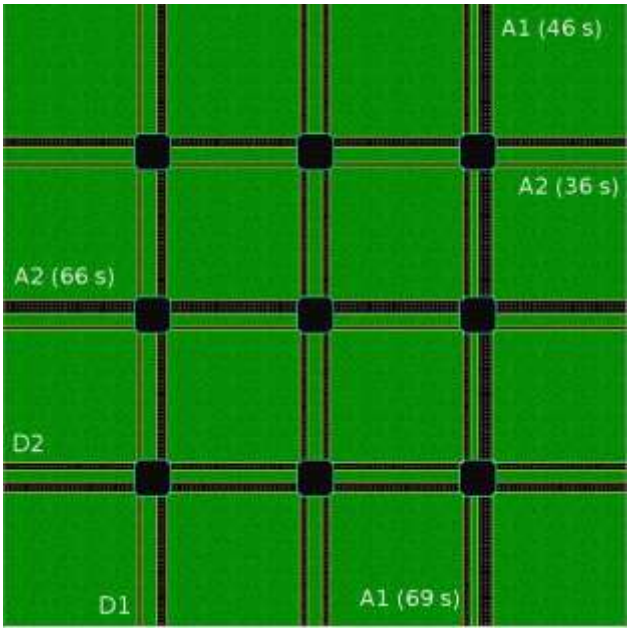


Figure 6.7 Grid network used for Model C results, with destinations and alternatives marked

Figure 6.8 present results that are similar in format to those presented for Model B (Figures 6.3 and 6.5). That is, in the left figure, total travel time cost as responsiveness parameter R varies and β is set (β equals 16). In the right figure, heat map showing total travel time cost with different R and β values.[18] Each data point in the right figure, and each bracket shade in the left figure, represents the average travel time cost over 30 simulation runs, where travel time cost for a single simulation is the average utility over all agents (vehicles). Each run simulates one hour of traffic. Error bars reflecting 95% confidence intervals are shown in the right figure (social welfare vs R).

The general trends that are observed in these results are very similar to those observed in the CTM model: that is, reducing R to approximately 10^{-4} improves system performance (due to mitigation of oscillation and spike effects), and that near-optimal performance can be achieved with most β values by properly tuning the R values. Nonetheless, there are two notable differences between these results and those presented for the CTM:

- Low β values (< 8) do not have a suitable R that yield optimized performance. We believe this discrepancy stems from differences in the congestion accumulation model. Recall that AIM manages intersections in a way that is conceptually different than traffic signals. When traffic signals are considered, the marginal impact of a single vehicle is negligible at low traffic levels (low demand) since vehicles must wait for a green signal regardless of the number of vehicles arriving from other directions. With AIM, however, the marginal impact of a vehicle is noticeable even at low traffic levels.

[18] Results in a format similar to that in Figure 6.4 are not presented for this model since, unlike the Model B results, traffic demand is not time varying in this experiment.

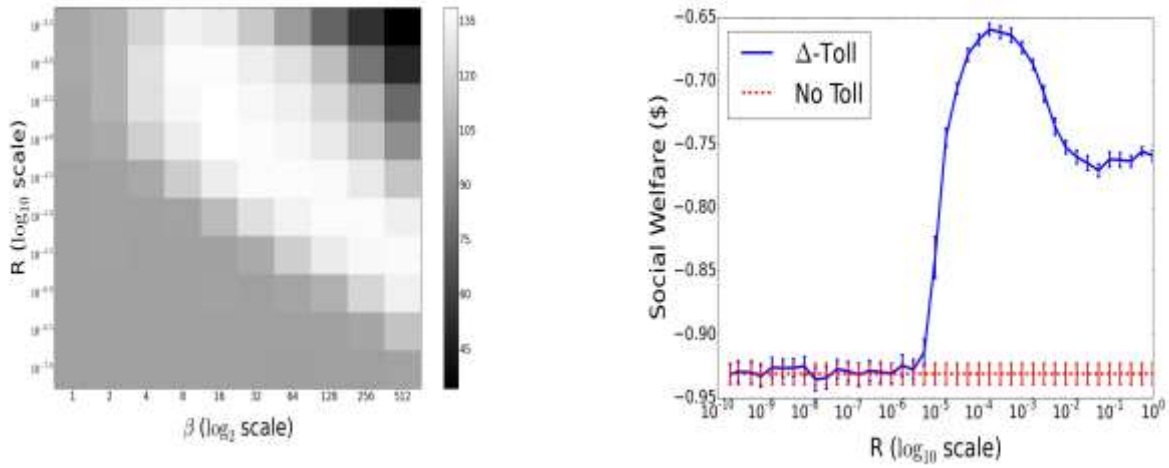


Figure 6.8 Results from running AIM in the 3 × 3 grid network. Left: Heat map showing the difference (in percentage) in total travel time cost (social welfare) compared to the no-tolls scenario for different R and β values. Right: Total travel time cost as responsiveness parameter R varies (Model C)

- R = 1 (rightmost data point in the right figure) presents performance that is better than applying no tolls. We believe this discrepancy also stems from the fact that AIM does not use traffic signals, which contributes to the negative effect of spikes and oscillation.

Though these results are not identical to those obtained by the CTM, their similarity still provides additional evidence of the robustness of Δ-tolling across different models and network topologies. On the other hand, the listed discrepancies suggest that the parameters used by Δ-tolling need to be re-tuned following changes in traffic flow modeling (such as changes to the intersection management policy).

6.7. Enhanced Delta-tolling

Δ-tolling, as presented above, includes two global parameters: β, which is a proportionality parameter, and R, which influences the rate of change of toll values across all links. We now turn to introduce a generalization of Δ-tolling that accounts for different β and R values on each link in the network. While this enhanced Δ-tolling algorithm requires setting significantly more parameters, we show that they can be tuned effectively via policy gradient reinforcement learning. Experimental results from several traffic scenarios indicate that enhanced Δ-tolling reduces total travel time by up to 33% compared to the original Δ-tolling algorithm, and by up to 52% compared to not tolling. Our detailed empirical study in Section 6.10 validates our claim that enhanced Δ-tolling has the potential to greatly improve upon the already impressive results of Δ-tolling when it comes to incentivizing self-interested agents to coordinate towards socially optimal traffic flows.

6.8. Background

Policy gradient RL is a general-purpose optimization method that can be used to learn a parameterized policy based on online experiential data. While there are several different methods for estimating the gradient of the policy performance with respect to the parameters (Peters and Schaal, 2006), one of the most straightforward, and the one we use in this report, is *Finite Difference Policy Gradient RL* (FD-PGRL) (Kohl and Stone, 2004), which is based on finite differences.

FD-PGRL is presented in Algorithm 1 (see Section 6.9). Under this framework, the policy is parameterized using the parameter vector $\pi = [\theta_1,\ldots,\theta_N]^\intercal$. The algorithm starts with the initial parameters $\pi^0 = [\theta_1{}^0,\ldots,\theta_N{}^0]^\intercal$ (line 1). At each step k, the policy gradient is estimated by running a set of randomly generated policies $\Pi^k = \{\pi_1^k, \ldots, \pi_M^k\}$ (lines 5–7) where each policy is defined as:

$$\pi_m^k = [\theta_1^{k-1} + \delta_{1,m}^k, \ldots, \theta_N^{k-1} + \delta_{N,m}^k]^\intercal, \tag{6.14}$$

where $\delta_{n,m}^k \in \{-\epsilon_n, 0, \epsilon_n\}$. The generated policies in (6.14) are obtained by randomly changing each parameter from the previous policy by a small ϵ_n, relative to θ_n. The cost of each newly created policy, $\pi_m{}^k$, is observed and denoted by $c^k{}_m$ (lines 8- 9).

To estimate the policy gradient, the policy set in (6.14) is partitioned to three subsets (lines 11–14) for each dimension depending on whether the change in the policy in that dimension is negative, positive or zero, that is the three subsets are:

$$\pi_m^k \in \begin{cases} \Pi_{-\epsilon,n}^k = \{\pi_m^k : \delta_{n,m}^k = -\epsilon\} \\ \Pi_{0,n}^k = \{\pi_m^k : \delta_{n,m}^k = 0\} \\ \Pi_{+\epsilon,n}^k = \{\pi_m^k : \delta_{n,m}^k = \epsilon\}. \end{cases} \tag{6.15}$$

The average costs of above policy subsets are denoted by $c_{-\epsilon,n}^k$, $c_{0,n}^k$ and $c_{+\epsilon,n}^k$ (lines 15–17). The adjustment vector $A^k = [a_1^k, \ldots, a_N^k]^\intercal$ can be constructed by the following equation for each dimension (lines 18–21):

$$a_n^k = \begin{cases} 0, & \text{if } \bar{c}_{-\epsilon,n}^k < \bar{c}_{0,n}^k \text{ and } \bar{c}_{+\epsilon,n}^k < \bar{c}_{0,n}^k \\ \bar{c}_{+\epsilon,n}^k - \bar{c}_{-\epsilon,n}^k & \text{otherwise} \end{cases} \tag{2.16}$$

The adjustment vector A^k is normalized and multiplied by a constant step size η to update the parameter vector at the end of each step k (lines 22–23).

Unlike other policy gradient methods that rely on within-episode reward signals to search for an optimal policy, or those in which the agent must learn the policy with no prior knowledge of a reasonably performing starting policy [for example, El Bsat et al. (2017) and Levine and Abbeel (2014)], in the method employed in this report, the policy is parameterized with a finite set of parameters and the overall system performance at each episode is optimized using an empirical estimate of the policy gradient based on finite differences. This approach is well-suited for the

traffic optimization problem for two reasons. First, the agent can leverage an existing policy with reasonable system performance. Second, the agent is required to proceed towards the optimal policy only by slight changes of the policy parameters in contrast to approaches in which randomized exploration policies can be executed more freely. Our empirical study suggests that considering such slight changes results in a total cost that is within an acceptable bound.

6.9. Enhanced Delta-tolling Mechanism

We now present the main contribution of this report: the enhanced Δ-tolling mechanism for solving the micro-tolling assignment problem. Enhanced Δ-tolling extends the Δ-tolling mechanism.

Algorithm 1: Finite Difference Policy Gradient RL uses two global variables to set tolls on every link in the network. Since different links possess different attributes—e.g., capacity, length, speed limit, etc.—optimizing the β and R parameters per link can potentially yield greater benefits (higher social welfare). However, doing so would require optimizing a set of $2|E|$ parameters instead of only two. Optimizing such a high dimensional function cannot be done in a brute force way.

$k \leftarrow 0$;
while *improving* **do**
  $k \leftarrow k + 1$;
  generate $\Pi^k = \{\pi_1^k, ..., \pi_M^k\}$,
    $\pi_m^k = [\theta_1^{k-1} + \delta_{1,m}^k, \ldots, \theta_N^{k-1} + \delta_{N,m}^k]^\intercal$,
    $\delta_{n,m}^k \sim Uniform\{-\epsilon_n, 0, \epsilon_n\}$;
  for *each* $m \in \{1, \ldots, M\}$ **do**
    $c_m^k \leftarrow run(\pi_m^k)$;
  for *each* $n \in \{1, \ldots, N\}$ **do**
    partition Π^k to
      $\Pi_{-\epsilon,n}^k = \{\pi_m^k : \delta_{n,m}^k = \epsilon\}$,
      $\Pi_{0,n}^k = \{\pi_m^k : \delta_{n,m}^k = 0\}$,
      $\Pi_{+\epsilon,n}^k = \{\pi_m^k : \delta_{n,m}^k = -\epsilon\}$;
    $\overline{c}_{-\epsilon,n}^k \leftarrow average(c_m^k : \pi_m^k \in \Pi_{-\epsilon,n}^k)$;
    $\overline{c}_{0,n}^k \leftarrow average(c_m^k : \pi_m^k \in \Pi_{0,n}^k)$;
    $\overline{c}_{+\epsilon,n}^k \leftarrow average(c_m^k : \pi_m^k \in \Pi_{+\epsilon,n}^k)$;
    if $\overline{c}_{-\epsilon,n}^k < \overline{c}_{0,n}^k$ *&* $\overline{c}_{+\epsilon,n}^k < \overline{c}_{0,n}^k$ **then**
      $a_n^k \leftarrow 0$;
    else
      $a_n^k \leftarrow \overline{c}_{+\epsilon,n}^k - \overline{c}_{-\epsilon,n}^k$;
  $\pi^k \leftarrow \pi^{k-1} - \eta \frac{A^k}{|A^k|}$,
    $A^k = [a_1^k, \ldots, a_N^k]^\intercal$;

$\pi_0 \leftarrow [\theta_{10},...,\theta_{N0}]$;

This report introduces enhanced Δ-tolling, which extends Δ-tolling by first considering unique β and R parameters per link and then incorporating policy gradient RL for optimizing these parameters.

In order to apply policy gradient RL (specifically FD-PGRL, as described in Section 6.8), the traffic assignment policy that maps the current state of the traffic to the appropriate actions, which are assigning tolls to each link of the network, should be parameterized. Since the Δ-tolling scheme inherently implemented a policy that takes into account the real-time state of the traffic by assigning the toll proportional to the current travel time, we only use RL policy gradient method to optimize the performance metric at the end of each traffic cycle. Therefore, we define the cost to be the total travel time at the end of each day and consider following three parametrizations of Δ-tolling:

$$\pi_R = [R_1,\ldots,R_n]$$
$$\pi_\beta = [\beta_1,\ldots,\beta_n]$$
$$\pi_{R,\beta} = [R_1,\ldots,R_n,\beta_1,\ldots,\beta_n] \quad (6.17)$$

The experimental results presented by Sharon et al. (2017b) suggest some correlation between the optimally performing β and R values. They state "For any fixed value of one parameter (β or R), there is a near-optimal value for the other parameter. This observation leads us to suspect that there are dependencies between the two parameters." This conjecture, if correct, means that optimizing only one of these parameters while keeping the second one constant would still allow optimized performance.

As the relation between the β and R parameters remains unclear, we consider three variants of enhanced Δ-tolling based on the parametrized policies listed in (6.17):

- **E**Δ-tolling$_\beta$ - this variant uses a global R parameter and link-specific β parameters ($|E|+1$ parameters in total). It should perform well under the assumption that there is a correlation between the best-performing β and R values and when FD-PGRL estimates the gradient over link-specific β parameters more accurately than it does for link-specific R parameters.

- **E**Δ-tolling$_R$ - this variant uses a global β parameter and link-specific R parameters ($|E| + 1$ parameters in total). It should perform well under the assumption that there is a correlation between the best-performing β and R values and when FD-PGRL estimates the gradient over link-specific R parameters more accurately than it does for link-specific β parameters.

- **E**Δ-tolling$_{\beta,R}$ - this variant uses link-specific β and R parameters ($2|E|$ parameters in total). It should perform best if there is no correlation between the best-performing β and R values and if sufficient computation time is given (converting on $2|E|$ parameters is usually slower than on $|E|+1$).

6.10. Empirical Study

Our experimental evaluation focuses on real-life road networks. Traffic is evaluated using the CTM (Daganzo, 1994a, 1995a), which is a discrete, explicit solution method for the hydrodynamic theory of traffic flow proposed in Lighthill and Whitham (1955) and Richards (1956).

CTM is frequently used in DTA. The time step used in this model is typically short, on the order of a few seconds. When used with enhanced Δ-tolling, this allows for a truly adaptive toll that can be updated based on observed traffic conditions.

6.10.1. Experiments and Results

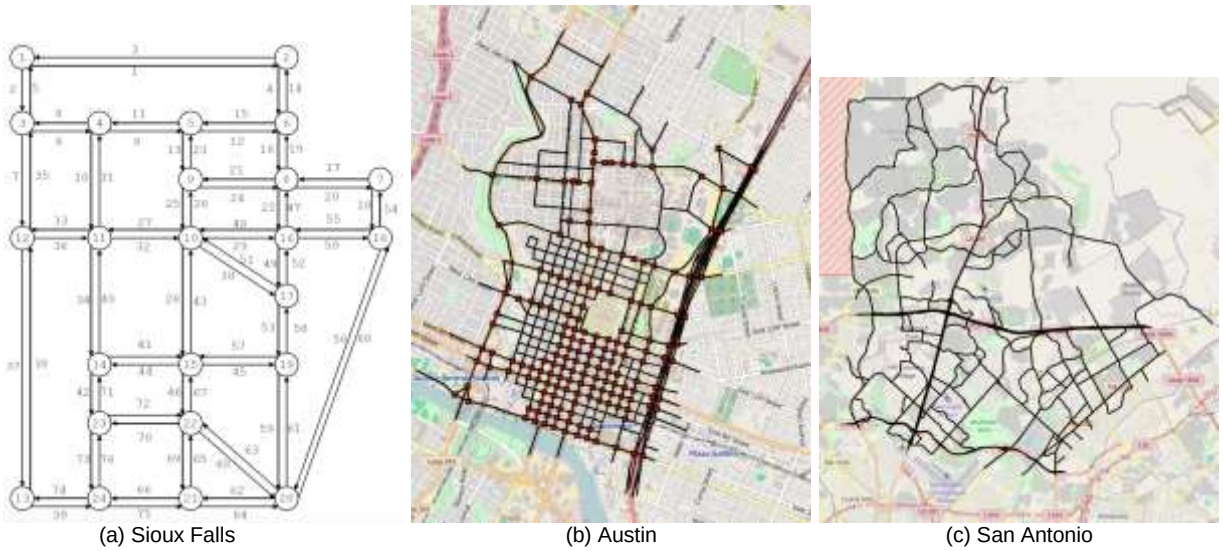
For running CTM we used the DTA simulator Chiu et al. (2011) implemented in Java. Whenever a vehicle is loaded onto the network, it is assigned a value of time (VOT) randomly drawn from a Dagum distribution with parameters $a\hat{} = 22020.6$, $\hat{}b = 2.7926$, and $\hat{}c = 0.2977$, reflecting the distribution of personal income in the United States (Lukasiewicza et al., 2012; Gardner et al., 2013).[19]

The step size in FD-PGRL , η, is 0.4. The policy perturbation parameter is set to 0.01 and the number of policy runs at each step, M, is 60 for all the experiments. These values presented best performance overall. Our empirical study focuses on three traffic scenarios:

- **Sioux Falls** - LeBlanc et al. (1975) — this scenario is widely used in the transportation research literature (Levin and Boyles, 2015), and consists of 76 directed links, 24 nodes, (intersections), and 28,835 trips spanning 3 hours.
- **Downtown Austin** - Levin et al. (2015b) — this network consists of 1,247 directed links, 546 nodes, and 62,836 trips spanning 2 hours during the morning peak.
- **Uptown San Antonio** — this network consists of 1,259 directed links, 742 nodes, and 223,479 trips spanning 3 hours during the morning peak.

The networks affiliated with each scenario are depicted in Figure 6.9. All of these traffic scenarios are available online at https://goo.gl/SyvV5m.

[19] The simulation settings were chosen to be identical to those presented in Sharon et al. (2017).



(a) Sioux Falls (b) Austin (c) San Antonio

Figure 6.9 Maps of traffic networks used in the experiments

6.10.1.1. System Performance

Our first set of results aims to evaluate the performance of the different variants of enhanced Δ-tolling, by comparing them with each other and basic Δ-tolling.

Figure 6.10 presents normalized values of total travel time summed over all trips (top figure) and social welfare that is the summation of costs, i.e., travel time times VOT, over all agents (bottom figure). The values are normalized according to the system's performance when no tolls are applied. Table 6.2 presents the total travel time and social welfare performance when applying no-tolls (representing the value of 1.0 in Figure 6.10).

The results present a clear picture in which Δ-tolling improves on applying no tolls in both total travel time and social welfare. EΔ-tolling$_{\beta}$ further improves the system's performance and both EΔ-tolling$_{R}$ and EΔ-tolling$_{\beta,R}$ achieve the best performance.

The fact that EΔ-tolling$_{R}$ results in system performance that is similar to EΔ-tolling$_{\beta,R}$ suggests a correlation between the best-performing β and R values.

The fact that EΔ-tolling$_{\beta}$ performs worse than EΔ-tolling$_{R}$ suggests that policy FD-PGRL estimates the gradient over link-specific R parameters more accurately than it does for link-specific β parameters.

Convergence rate applying EΔ-tolling to real-life traffic raises two concerns:

1. Convergence rate - the system should converge to a good solution with as few learning iterations as possible.
2. Worst case performance - during the learning process EΔ-tolling should perform at least as well as Δ-tolling.

Table 6.2 Average total travel time and total generalized cost when applying no tolls

	Sioux Falls	Austin	San Antonio
Travel time (hr)	11,859	21,590	26,362
Cost ($)	353,169	637,086	780,739

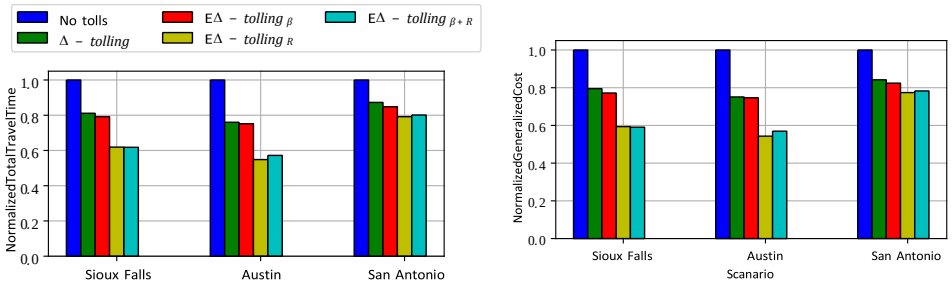


Figure 6.10 Total travel time and total generalized cost for different tolling schemes and scenarios

Figure 6.11 presents the system performance with regard to total travel time (y-axis) versus learning iteration step (x-axis) for each of our three scenarios and every EΔ-tolling variant. The error regions are obtained using 10 different runs of the algorithm for each example and EΔ-tolling variant and they show the standard error of the average performance in each iteration. Results for basic Δ-tolling are also included for comparison. The results are consistent with each other, showing that EΔ-tolling$_R$ performs best overall w.r.t convergence rate.

Table 6.3 presents the area under the curve for each scenario and EΔ-tolling variant. These results give a quantitative comparison of the convergence rates. We learn that EΔ-tolling$_R$ has the best overall performance with a total AUC of 4,285,353. Nonetheless, EΔ-tolling$_{\beta,R}$ performs better on the Sioux Falls scenario.

All the experiments are initialized with $\beta = 4$ and $R = 10^{-4}$ for all the links. A set of experiments (not presented) with different starting parameter values show that the performance is sensitive to the initial settings. However, the mentioned default starting values ($\beta = 4$ and $R = 10^{-4}$) perform relatively well across all scenarios and EΔ-tolling variants.

Table 6.3 Area under the convergence curves from Figure 6.11

Scheme	S. Falls	Austin	S. Antonio	Total
Δ-tolling	962,000	1,640,900	2,300,700	4,903,600
EΔ$_\beta$	943,076	1,619,928	2,257,830	4,820,834
EΔ$_R$	779,990	**1,360,861**	**2,144,502**	**4,285,353**
EΔ$_{\beta+R}$	**777,469**	1,415,094	2,162,006	4,354,569

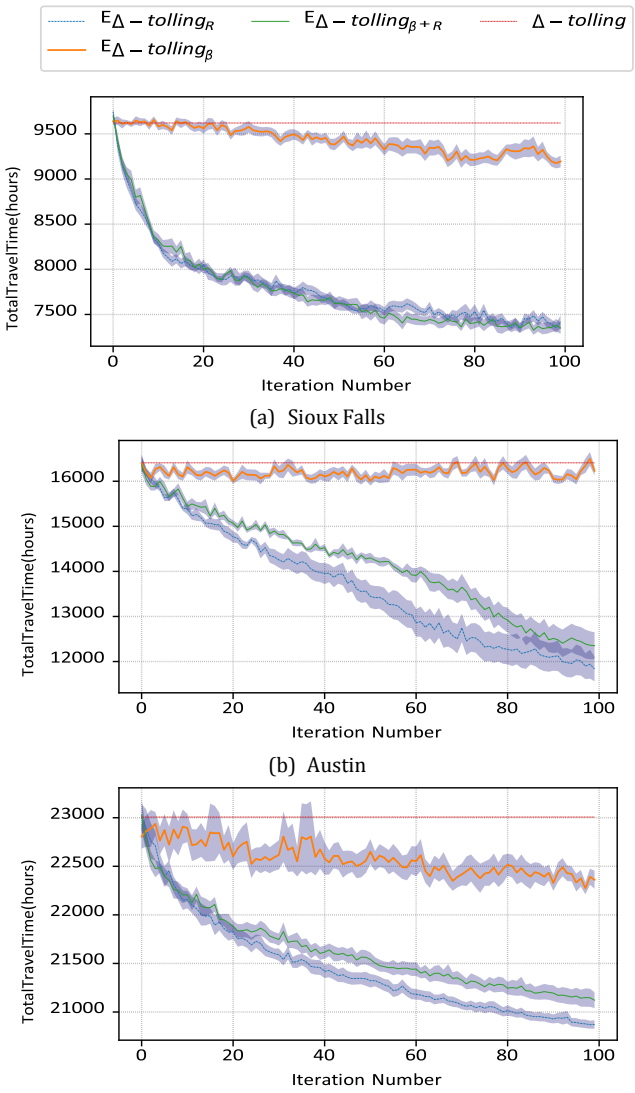


Figure 6.11 Total travel time and total generalized cost for different tolling schemes and scenarios

6.11. Discussion and Future Work

The promising experimental results reported suggest that EΔ-tolling can have practical applications where traffic optimization is performed constantly and in real time through manipulations to the R and or β parameters. Nonetheless, implementation of EΔ-tolling raises several practical issues that must first be addressed.

- Limitations - EΔ-tolling is limited in its convergence rate. General traffic patterns might change frequently, preventing EΔ-tolling from advancing in a promising direction. Practitioners must evaluate the convergence rate of EΔ-tolling versus the rate in which traffic patterns change in order to determine the applicability of EΔ-tolling in a specific network.

- Assumptions - EΔ-tolling, as presented in this report, assumes that all agents traversing the network are self-interested and responsive to tolls in real time. Real-world scenarios might violate these assumptions and the trends observed in our results cannot be assumed in such cases.

Practical aspects of EΔ-tolling present many promising directions for future work. Since the convergence rate of EΔ-tolling plays an important role in determining its applicability, one promising direction for future work is developing heuristics and utilizing advanced RL methods to guide the gradient exploration towards promising directions in order to facilitate faster learning.

Examining the effects of partial compliance to tolls is another promising direction. Building on recent theoretical study that examines the effects of partial compliance on similar micro-tolling schemes (Sharon et al., 2018), studying the practical impacts of partial compliance on EΔ-tolling is a promising direction to pursue.

Lastly, another promising direction is examining variants of EΔ-tolling that are adapted to scenarios where traffic congestion is predictable to some extent.

6.12. Conclusions

This chapter presented Δ-tolling, a simple road pricing scheme that makes minimal assumptions on the traffic flow model or driver behavior. This scheme involves only two parameters, and only requires link travel time and free flow travel times measurements to set tolls. The flexibility of Δ-tolling was demonstrated by applying it in three very different contexts: a day-to-day pricing framework where delay is determined by link performance functions and a static equilibrium model; a within-day adaptive tolling framework using the CTM for dynamic network loading, with adaptive route choice but no equilibrium; and an adaptive tolling application using a new reservation-based intersection scheme for automated vehicles, evaluated in microsimulation. In all of these cases, the Δ-tolling scheme was able to achieve significant benefits (measured in average travel time or social welfare) over the no-toll case, even without knowledge of the different traffic models being used, or the different assumptions on driver behavior. Benefits were seen both in small, artificial grid networks with randomized parameters as well as in larger networks representing real-world cities. We also note that Δ-tolling does not necessarily require a computer-controlled vehicle; it only requires computer-controlled route choice. Current smartphone software already provides navigation to human drivers, and such software could be modified to interact with tolling systems.

The Δ-tolling scheme represents an advance over previously suggested toll schemes, by not requiring any of the following assumptions: that demand is known or fixed, that roadway capacity is known or fixed, that the value of time is homogeneous; that the traffic model is known. Furthermore, Δ-tolling is applicable across large networks and aims to optimize social welfare. As discussed in Section 6.1, all previous work we are aware of makes one or more of these assumptions, all of which have significant practical implications. As a few examples, drivers are unlikely to voluntarily report all of their trips to the tolling agency (so that demand is not fully

known), all traffic models are approximations to real traffic flow, and what is optimal for a single corridor may not be optimal over a larger network. Our aim in presenting Δ-tolling is to show that substantial benefits can be obtained even without knowing all of this information.

Δ-tolling is simple to implement since it requires measuring only two variables: current travel time and free flow travel time, both measurements are feasible with current technology. It is robust to the underlying traffic model and does admit optimality results under certain assumptions (Model A). Even when optimality is not provable, Δ-tolling results in significant average travel time reduction in Models B and C. The fact that Δ-tolling gains significant improvements over three different traffic models suggests it may be beneficial in other models as well, including real-life traffic.

This chapter also introduced enhanced Δ-tolling, a micro-tolling assignment scheme that builds on the previously suggested Δ-tolling scheme. The previously suggested Δ-tolling scheme makes use of two global parameters, β and R, to tune the system for optimized performance (minimal total travel time or maximal social welfare).

Enhanced Δ-tolling generalizes Δ-tolling in two complementary ways. First, recognizing that different links in the network have different attributes (length, capacity, speed limit), enhanced Δ-tolling considers individual β and R parameters per link. Second, given the resulting large parameter set (twice the number of links), enhanced Δ-tolling suggests a policy gradient RL approach for tuning and optimizing these parameters.

Experimental results suggest that tuning the R parameter while keeping a global β parameter performs best overall (with regard to total travel time, social welfare, worst case performance, and convergence rates).

Chapter 7. Partial Compliance with Tolls

This chapter examines the impact of tolls on the travel time cost of a transportation network when only a portion of the agents traversing the network take tolls into account when planning their route. More specifically, we address the question: which subset of agents provides the most system benefit if they are compliant with an approximate marginal cost tolling scheme? Since previous work suggests this problem is NP-hard, we present three different heuristic methods as possible answers to this question. Our experimental results on three real-life traffic scenarios suggest that evaluating the marginal impact of a given agent serves as a particularly strong heuristic for selecting an agent to be compliant. Results from using this heuristic for selecting 7.6% of the agents to be compliant achieved an impressive increase of up to 10.9% in social welfare over not tolling at all. The presented heuristics and conclusions can help practitioners target specific agents to participate in an opt-in tolling scheme.

In this chapter, we consider a routing scenario in which a subset of agents is controlled centrally (*compliant agents*), while the remaining are *self-interested agents*. We model the system as a Stackelberg routing game (Yang et al., 2007) in which the decision-maker for the centrally controlled agents is the leader, and the self-interested agents are the followers. In this chapter, we provide a computationally tractable methodology for:

1. Determining the maximum number of self-interested agents that a system can tolerate at optimal flow.
2. Determining whether a given subset of centrally controlled agents is sufficient to achieve system optimum (*SO*).
3. Computing the actions the leader should prescribe to a sufficient set of compliant agents in order to achieve *SO*.

A known fact in routing games is that agents seeking to minimize their private travel time need not minimize the total system's travel time (Pigou, 1920b; Roughgarden and Tardos, 2002). That is, self-interested agents may reach a user equilibrium (UE) that is not optimal from a system perspective. However, if all agents are assigned paths with minimum system marginal cost then the system will achieve optimal performance (Pigou, 1920b; Beckmann et al., 1956; Dietrich, 1969).

Therefore, from a system manager perspective, it is desirable that all agents traversing a network would strictly utilize minimal marginal cost paths, even if such paths are not of minimum travel time for an individual agent. However, in many important scenarios, it will not be possible to enforce path assignment on all agents, but it may be possible to affect the behavior of a subset (the compliant agents). As a motivating example, consider an opt-in tolling system where drivers are given positive incentives to enroll but, in exchange, they will be subject to tolls that affect their route choice (Sharon et al., 2017a).

We show that, in the general case, computing the optimal assignment of compliant agents is NP-hard. Therefore, we focus on the specific scenario where the portion of compliant agents is sufficiently large to achieve SO. We present a novel *linear program* (LP) representation for computing the maximal portion of self-interested agents that allow the system to achieve SO and to determine whether a given set of compliant agents is sufficient to achieve SO. Furthermore, we provide a method to tractably compute the flow assignment for the compliant agents such that SO performance is guaranteed.

Experimental results, obtained using a standard traffic simulator, are provided and demonstrate that the number of compliant agents necessary to achieve SO can be a relatively small percentage of total flow (between 13% and 53%).

7.1. Motivation

Political factors deter public officials from allowing micro-tolling scheme such as Δ-tolling to be realized. Road pricing is known to cause a great deal of public unrest and is thus opposed by governmental institutions Schaller (2010). To tackle this issue and avoid public unrest, we suggest an *opt-in* micro-tolling system where, given some initial monetary sign-up incentive, drivers choose to opt in to the system and be charged for each journey they take based on their chosen route. The vehicles belonging to such drivers would need to be equipped with a GPS device as well as a computerized navigation system. Given the toll values and driver's value of time (VOT), the navigation system would suggest a minimal cost route where the cost is a function of the travel time and tolls.

While addressing the issue of political acceptance, an opt-in system would result in traffic that is composed of a mixture of self-interested and compliant agents (compliant in the sense that the system manager can influence their route choice). Such a scenario raises some practical questions which are the focus of this chapter, namely, what portion of self-interested agents can the system tolerate while still reaching optimum performance? The answer to this question can help practitioners to determine both the level and the targeting of incentives in an opt-in system.

7.2. Problem Definition and Terminology

The terminology in this chapter follows that presented in the previous chapter. We review the relevant concepts and notation in this section.

7.3. The Flow Model

The flow model in this work is composed of a directed graph $G(V,E)$, and a demand function $d(s,t) \rightarrow R^+$ mapping a pair of vertices $s,t \in V^2$ to a non-negative real number representing the required amount of flow between source, s, and target, t.[20] An instance of the flow model is a $\{G,d\}$ pair.

[20] The demand between any source and target, $d(s,t)$, can be viewed as an infinitely divisible set of agents (also known as a non-atomic flow) (Rosenthal, 1973).

$\Pi_{s,t}$ denotes the set of acyclic paths from s to t. Define π as the collection of all $\Pi_{s,t}$ (i.e., $\cup_{s,t \in V^2} \Pi_{s,t}$). The variable f_π represents the flow volume assigned to path π (in contrast to the previous section). Similarly, f_e is the flow volume assigned to link e. By definition, the flow on each link (f_e) equals the summation of flows on all paths of which e is a part. Define the system flow vector as f = vect$\{f_\pi\}$. f is said to be *feasible* if for all $s,t \in V^2$, $\sum_{\pi \in \Pi_{s,t}} f_\pi = d(s,t)$.

Each link $e \in E$ has a travel time function $l_e(f_e)$ which, given a flow volume (f_e), returns the travel time (travel time) on e. Following Roughgarden and Tardos (2002) we make the following assumption:

Assumption 1. *The travel time function $l_e(f_e)$ is non-negative, differentiable, and non-decreasing for each link $e \in E$.*

The travel time of a simple path π for a given flow f, is defined as $l_\pi(f) = \sum_{e \in \pi} l_e(f_e)$. A feasible flow f is defined as a *user equilibrium* (*UE*) if for every $s,t \in V$[21] and $\pi_a, \pi_b \in \Pi_{s,t}$ with $f_{\pi_a} > 0$ it holds that $l_{\pi_a}(f) \leq l_{\pi_b}(f)$ [see Lemma 2.2 in Roughgarden and Tardos (2002)]. In other words, at *UE*, no amount of flow can be rerouted to a path with lower travel time when the rest of the flow is fixed.

Define the system cost associated with link e as $c_e(f_e) = l_e(f_e)f_e$, the cost of a path π as $c_\pi(f) = \sum_{e \in \pi} c_e(f_e)$ and the cost of a flow f as $c(f) = \sum_{e \in E} c_e(f_e)$. Define $c'_e(x) = \frac{d}{dx} c_e(x)$ and $c'_\pi(f) = \sum_{e \in \pi} c'_e(f_e)$. A feasible flow f is defined as a *system optimum* (*SO*) flow if for every $s,t \in V^2$ and $\pi_a, \pi_b \in \Pi_{s,t}$ with $f_{\pi_a} > 0$ it holds that $c'_{\pi_a}(f) \leq c'_{\pi_b}(f)$ [see Lemma 2.5 in Roughgarden and Tardos (2002)]. In other words, at *SO*, the benefit from reducing the flow along any path is always less than or equal to the cost of adding the same amount of flow to a parallel, alternative path. We follow Roughgarden and Tardos (2002), and make the following assumption:

Assumption 2. *The cost function $c_e(f_e)$ is convex for each link $e \in E$.*

Assumptions 1 and 2 imply that the set of *SO* flows corresponds to the set of solutions of a convex program where the objective is to minimize $c(f) = \sum_{e \in E} c_e(f_e)$ [see Roughgarden and Tardos (2002) Corollary 2.7].

7.3.1. Problem Definition

The focus of this chapter is a scenario where the demand is partitioned into self-interested and compliant agents. We define two types of controllers that assign paths to all of the agents. These controllers are viewed as players in a Stackelberg game (Yang et al., 2007).

- *SO*-controller—Stackelberg leader: the *SO*-controller aspires to assign paths to the compliant subset of agents that, taking into account the self-interested agents' reaction,

[21] The *UE* enforced by the *UE*-controller applies only for the self-interested subset of agents. That is, no **self-interested** agent can benefit from unilaterally deviating from its assigned path.

optimizes the systems performance (i.e., minimizes total travel time). We refer to flow assigned by the *SO*-controller as *compliant flow*.

- *UE*-controller—Stackelberg follower: considering the compliant agents' path assignment as fixed, the *UE*-controller assigns paths to the self-interested agents, the *UE flow*, such that a state of UE (as defined above) is achieved.[2]

This report addresses these questions:

1. Given an instance of the flow model $\{G,R\}$, what is the maximum number of self-interested agents that can be assigned to the *UE* controller and still permit the optimal flow?
2. Given a set of compliant agents and an instance of the flow model $\{G,R\}$, can the *SO* controller assign paths to them in such a way that the system achieves *SO*?
3. If *SO* is achievable, how should the *SO*-controller assign the compliant flow? Equivalently, what is the optimal Stackelberg equilibrium?

To the best of our knowledge, this work is the first to answer these questions in a general setting.

7.4. Related Work

Previous work examined mixed equilibrium scenarios where traffic is composed of *UE* and Cournot-Nash (*CN*) controllers. A *CN*-controller assigns flows to a given subset of the demand with the aim of minimizing the total travel time only for that subset. For instance, a logistic company with many trucks can be viewed as a *CN*-controller.

It was shown that the equilibrium for a mixed *UE*, *CN* scenario is unique and can be computed using a convex program (Haurie and Marcotte, 1985; Yang and Zhang, 2008). On the other hand, no tractable algorithm is known for computing the optimal Stackelberg equilibrium for scenarios that also include a *SO*-controller.

Korilis et al. (1997) examined mixed equilibrium scenarios that do include a *SO*-controller. In their work, a technique for computing a solution for the above questions #1 and #3 was suggested for specific types of flow models. Their technique was proven to work for networks with a common source and a common target with any number of parallel links. Moreover, the travel time functions were assumed to be of a very specific form (linear function with a capacity bound). As a result, their solution is not applicable when general networks with arbitrary travel time functions are considered.

Other work (Roughgarden, 2004; Immorlica et al., 2009) studied a variant of the scheduling problem where infinitesimal jobs must be assigned to a set of shared machines each of which is affiliated with a non-negative, differentiable, and non-decreasing travel time function that, given the machine load, specify the amount of time needed to complete a job. When considering a scenario where some of the jobs are assigned to machines by a *UE*-controller while the rest are assigned by a *SO*-controller, they show it is NP-hard to compute the optimal Stackelberg

equilibrium (Roughgarden, 2004). Their problem can be viewed as a special case of our problem, specifically a network with a single source and target with multiple parallel links between them. Given that in this more restrictive setting, computing the optimal Stackelberg equilibrium is intractable, the same yet general question in our setting will also be computationally intractable.

7.5. Computing the Maximal *UE* Flow

Given that finding the optimal Stackelberg equilibrium is NP-hard for an arbitrary size of compliant flow, this work focuses on scenarios where the size of the compliant flow is sufficient to achieve *SO*. As we will show, finding the optimal Stackelberg equilibrium can be done in polynomial time for such cases. In this section, we will present a computationally tractable method to compute the maximal *UE* flow given an instance of a flow model {*G,R*}, and we will provide a method to check, for a given level of compliant flow, whether *SO* is achievable.

We define d^*_{UE} as the maximal amount of demand comprised of self-interested agents that the system can tolerate and still achieve *SO*. Additionally, we define $d^*_{s,t}$ as the amount of demand from source *s* to target *t* that is assigned to the *UE*-controller. That is, computing d^*_{UE} is equivalent to maximizing $\sum_{s,t} d^*_{s,t}$.

We can cast the problem of maximizing $\sum_{s,t} d^*_{s,t}$ as an optimization problem, specifically a *linear program* (*LP*). Assigning values to all variables of type $d^*_{s,t}$ must follow some constraints. Specifically, the *UE* flow from each origin to each destination must be both a *subflow* of the *SO* flow and must follow a least travel time path.

> **Definition 7** (Subflow of flow *f*). *For a directed graph G(V,E) and demand function d, a flow f^* is a* subflow of flow *f if for all links $e \in E$, $0 \le f^*_e \le f_e$ and for each pair of nodes $s,t \in V^2$, there exists $0 \le d_{s,t} \le d(s,t)$ such that*
>
> $$\sum_{e \in out(s)} f^*_e - \sum_{e \in in(s)} f^*_e = \sum_t d_{s,t}$$
>
> *and*
>
> $$\sum_{e \in in(t)} f^*_e - \sum_{e \in out(t)} f^*_e = \sum_s d_{s,t}$$
>
> .

> **Definition 8** (Zero reduced cost path). A path π, leading from vertex *s* to vertex *t*. *For a flow model {G, d}, a* zero reduced cost path *with regard to flow assignment f is a path $\pi \in \Pi_{s,t}$ such that $\forall \pi^0 \in \Pi_{s,t}$: $l_\pi(f) \le l_{\pi^0}(f)$ and $c'_\pi(f) \le c'_{\pi^0}(f)$. A link, e, is defined as a* zero reduced cost link, *with respect to source s, if it is part of any zero reduced cost path originating from s and terminating at t for some origin-destination pair (s, t) $\in V^2$. We denote the set of zero reduced cost links with respect to source s as E^s_{RC}.*

We require that the UE flow (flow routed by the UE-controller) is routed solely via zero reduced cost links/paths. This is because the UE controller can only assign flow to minimal travel time

paths (otherwise self-interested agents would deviate). The UE flow is also required to follow minimal marginal cost paths else it cannot be a subflow of the SO flow.

Note that it is sufficient to only consider whether or not a link *e* is part of a reduced cost path from the origin s to *some* destination *t* (not a specific *t*) because either link *e* is along a reduced cost path from (*s*, *t*), or there is no path only along links in E_{RC}^{s} that includes *e*.

We can efficiently compute the set of zero reduced cost links for any origin-destination pair (*s*, *t*) by applying uniform cost search from *s* to *t* and marking all links that are part of optimal paths, once with regard to minimal total travel time ($\underset{\pi \in \Pi_{s,t}}{argmin}\left(l_{\pi}(f^{SO})\right)$), and second with regard to minimal marginal cost ($\underset{\pi \in \Pi_{s,t}}{argmin}\left(c'_{\pi}(f^{SO})\right)$).

Let the constant f^{SO} denote the flow vector at a *SO* solution.[22] The *SO* flow is not unique when travel time functions are non-decreasing, and the maximal amount of *UE* flow permitted may, in general, depend on the specific *SO* flow. Therefore, we must efficiently search over the space of *SO* flows. This is possible due to the following lemmas.

Lemma 1. *For any two flows that achieve SO,* f^{SO} *and* $\hat{f}^{SO}$, $l_e(f_e^{SO}) = l_e(\hat{f}_e^{SO})$.

Proof. Given Assumption 2, a *SO* flow is the solution to a convex program (Roughgarden and Tardos, 2002). The solutions to a convex program form a convex set. Suppose that there are two flows that both achieve SO, but for which $f_a^{SO} \neq \hat{f}_a^{SO}$. Then $c_e(f_e) = l_e(f_e)f_e$ must be a linear function between f_a^{SO} and $\hat{f}_a^{SO}$ (to see this, note that any convex combination of f^{SO} and $\hat{f}^{SO}$ is also an *SO* solution, but if $c_e(f_e)$ is not linear, then the total system travel time would be strictly less, a contradiction). Since $l_e(f_e)$ is a non-decreasing function, the only way for $c_e(f_e)$ to be linear is for $l_e(f_e)$ to be constant between f_a^{SO} and $\hat{f}_a^{SO}$. □

Lemma 2. *The set of zero reduced cost paths is identical for all SO solutions.*

Proof. By Lemma 1, all *SO* flows have the same travel time on each link, so the *SO* solutions can differ by at most flows along a set of links with constant travel time over the range of which the two flows differ on those links. Since we assume that the travel time functions are differentiable, the derivatives of the travel time function are zero over the range at which they are constant. Therefore, $c'_e(f_e) = l_e(f_e) + f_e l'_e(f_e)$ is constant over the range as well. This implies that any path that is reduced cost in one flow is also reduced cost in the other flow, since the travel time functions and $c'_e(f_e)$ are constant for every link *e*. □

Define the constant $\bar{f}_e^{SO} = \sup\{f : l_e(f) = l_e(f_e^{SO})\}$, i.e., $\bar{f}_e^{SO}$ is the largest flow value such that the travel time on link *e* is equal to the travel time at a *SO* solution. Note that if l_e is strictly increasing at f_e^{SO}, then $\bar{f}_e^{SO} = f_e^{SO}$. However, if l_e is constant at f_e^{SO}, then $\bar{f}_e^{SO} > f_e^{SO}$.

[22] A *SO* flow can be efficiently computed as a solution to a convex program (Roughgarden and Tardos, 2002; Dial, 2006b).

Given that the zero reduced cost paths are the same for *all SO* flows (Lemma 2), and any *SO* flow has the same travel time on all links (Lemma 1), it will be sufficient to only search over flows that are less than $\bar{f}_e^{SO}$ on each link $e \in E$.

For each vertex, *s*, and link, *e*, define variable x_e^s denoting the amount of *UE* flow originating from source *s* that is assigned to link *e*. Let *in*(*v*) denote the set of links for which *v* is the tail vertex and *out*(*v*) the set of links for which *v* is the head vertex.

Definition 9 (UE linear program). *For a given flow model* {*G*,*d*}, *the UE* linear program *is the following:*

$$\max_{d_{s,t}^*, x_e^s} \sum_{s,t \in V^2} d_{s,t}^* \tag{7.1}$$

subject to

$$d_{s,t}^* \leq d(s,t) \qquad \forall s,t \in V^2 \tag{7.2}$$

$$\sum_{e \in out(s)} x_e^s = \sum_{t \in V} d_{s,t}^* \qquad \forall s \in V \tag{7.3}$$

$$\sum_{e \in in(t)} x_e^s - \sum_{e \in out(t)} x_e^s = d_{s,v}^* \qquad \forall s,t \in V^2 \tag{7.4}$$

$$\sum_s x_e^s \leq \bar{f}_e^{SO}$$

$$x_e^s \geq 0,\ d_{s,t}^* \geq 0 \qquad \forall e \in E, s \in V \tag{7.5}$$

$$x_e^s = 0 \qquad \forall s,t \in V, e \in E \tag{7.6}$$

$$\forall s \in V, e \in E \setminus E_{RC^s} \tag{7.7}$$

The flow $f_e^{UE} = \sum_v x_e^v$ defined by a feasible solution to the UE linear program [given constraints (7.2)–(7.7)] is a UE subflow. The flow defined by an optimal solution to the UE linear program is an optimal UE subflow.

Note that the number of variables is $|\{\forall s \in V, \forall t \in V, \forall e \in E : d^*_{s,t}, x_e^s\}| = O(|V|^2 + |V||E|)$, and the number of constraints is also $O(|V|^2 + |V||E|)$. Therefore, since the number of variables and constraints are polynomial in the flow model, the optimal solution to the UE linear program can be computed in polynomial time (Karmarkar, 1984).

Theorem 1. A UE subflow, f^{UE}, defined by a feasible solution to the UE linear program is a subflow of a SO flow.

*Proof***.** First, note that by equations (7.2)–(7.4), the UE subflow, f_e^{UE}, satisfies flow conservation constraints. Equation (7.2) states that the flow along all zero reduced cost paths from origin *s* to destination *t* must be less then total demand for (*s*,*t*). Then Equations (7.3) and (7.4) state that the flow out of node *v* must either be due to the demand generated by node *v* or the flow

into it, minus the flow that reaches v as a destination. Therefore, f_e^{UE} is a subflow of a feasible flow.

What must be shown is that there must exist a *SO* flow, f^{SO}, such that $f_e^{UE} \leq f_e^{SO}$ for all e. If e is such that l_e is strictly increasing at an *SO* solution, and therefore will be strictly increasing at all *SO* solutions by Lemma 1, then $f_e^{SO} - f_e^{SO}$ and constraint (7.5) guarantees this claim. Let E^0 be the set of links such that the travel time function is constant at a *SO* flow. Therefore, it only needs to be shown that there exists a *SO* solution, f, such that for $e \in E'$, $f_e^{UE} \leq f_e^{SO}$.

Suppose that there existed a set of links $e \in E^0$ such that for all *SO* flows f^{SO}, $f_e^{UE} > f_e^{SO}$. Let f^{SO} be an *SO* flow. Then there must exist an origin-destination pair (s,t) such that there are two sets of paths $\Pi_>, \Pi_< \subset \Pi_{s,t}$ for which for all $\pi \in \Pi_>$, $f_\pi^{UE} > \hat{f}_\pi^{SO}$, and for all $\pi' \in \Pi_<$, $f_{\pi'}^{UE} < \hat{f}_{\pi'}^{SO}$ and all paths only differ by links in E^0. This is because the total flow between any origin-destination is larger in the *SO* flow by Equation (7.2). Moreover, $\sum_{\pi \in \Pi_>} (f_\pi^{UE} - \hat{f}_\pi^{SO}) \leq \sum_{\pi' \in \Pi_<} (\hat{f}_\pi^{SO} - f_\pi^{UE})$ since the flow along non-constant travel time links constrains the total flow. Move $\sum_{\pi \subset \Pi_>} (f_\pi^{UE} \quad \hat{f}_\pi^{SO})$ units of flow from paths in set $\Pi_>$ to paths in set $\Pi_<$ in the *SO* flow f^{SO}. Denote the new flow by f^0. The total travel time for f^0 cannot increase because the flow has only increased on constant travel time links, and the new flow does not exceed f_e^{SO} on any link. The total travel time also cannot have decreased because f^{SO} was an *SO* flow, so f^0 is also an *SO* flow. Continue this procedure until there does not exist a link $e \in E^0$ for which f_e^{UE} exceeds the transformed *SO* flow. Then we have constructed an *SO* flow, f, in which, for all links $e \subset E$, $f_e^{UE} \leq f_e$, a contradiction. □

Lemma 3. *For a network {G,d}, let f^* be a subflow of a feasible flow f. Then the flow f^0 such that $f'_e = f_e - f_e^*$ is also a subflow of f.*

Proof. First, $0 \leq f'_e \leq f_e$, by the definition of a subflow. Now set $d^0_{s,t} = d(s,t) - d^*_{s,t}$. Then for all $s,t \in V$ [2], $\sum_{e \in out(s)} f'_e - \sum_{e \in in(s)} f'_e = \sum_t (d(s,t) - d^*_{s,t}) = \sum_t d'_{s,t}$, and similarly for $\sum_{e \in in(t)} f'_e - \sum_{e \in out(t)} f'_e$ □

Theorem 2. *The optimal value of the UE linear program for a network instance {G,d} is the maximum amount of UE agents that the network can support and achieve SO.*

Proof. First, by Theorem 1, there exists an *SO* flow such that the optimal *UE* subflow, f^{UE}, is a subflow of the *SO* flow, and by Lemma 3, there exists a subflow of compliant agents that can achieve the *SO* solution. Moreover, by the definition of the *UE* linear program and Lemma 2, the *UE* flow is only along zero reduced cost paths. By the definition of zero reduced cost paths, all *UE* agents are willing to take the assigned paths. Therefore, the *SO* solution is achievable with the *UE* flow, and there is some volume of *UE* flow that is equal to the objective of the *UE* linear program.

Now, suppose that there was another *UE* flow assignment, f^0, for which compliant flow could be assigned in such a way that the *SO* total system travel time was achieved and the total *UE* flow volume was larger than the value returned by the *UE* linear program. Note that this flow assignment (f^0) must be a subflow of some *SO* flow, f. Moreover, by the definition of *UE* flow and the fact that all paths in a *SO* solution are minimum marginal cost paths, all paths assigned with a *UE* flow

greater than zero must be a zero reduced cost path. Therefore, the flow f^0 satisfies the equations (7.2)-(7.6), and since the *UE* linear program returns the optimal *UE* flow assignment under these constraints, this is a contradiction. □

While we've demonstrated that we can compute the *maximal UE* flow that permits an *SO* solution given the appropriate assignment of the compliant flow, it is likely that a more common problem would be to determine, for a given set of compliant agents, whether or not it is possible to achieve *SO* with that set. Our methodology also provides an answer to this question, as the following Corollary demonstrates.

Corollary 1. *For a given network instance {G,d} and given a set of compliant demand, $d^C_{s,t}$, from each origin-destination pair s, t ∈ V², there exists a compliant flow f^C such that the network achieves SO if and only if there exists an x^s_e for all s ∈ V and e ∈ E such that $d_{s,t}{}^{UE} = d(s,t) - d_{s,t}{}^C$ and x^s_e are a solution to the UE linear program.*

Proof. By Theorem 1, any solution to the *UE* linear program defines a subflow of an *SO* flow. Therefore, if $d^{UE}_{s,t}$ and x^s_e is a solution, there exists an assignment of the compliant flow that achieves *SO*.

Moreover, if there exists an assignment of the complaint flow, f^C, such that a *UE* subflow with demands $r_{s,t}{}^{UE}$ achieves SO, then the *UE* flow is only along zero reduced cost paths by definition of *UE* flow and *SO*, and the *UE* subflow is feasible. Therefore, the decomposed *UE* flow satisfies the constraints of the linear program. □

7.6. Flow Assignment for Compliant Agents

Given that we can now determine both the maximal amount of *UE* flow that a system can tolerate and achieve SO and, for a given set of compliant agents, whether or not a system can achieve optimum, we are only left with assigning the compliant flow to paths. This section tackles the question of how to assign paths to a, sufficiently large, set of compliant agents such that *SO* is achieved.

The methodology from the previous section immediately suggests a solution. Given a network instance {*G,d*}, suppose that we have compliant demand equal to $r_{s,t}{}^C$ for all *s, t* ∈ *V*². Then we must find a *SO* flow, f^{SO}, such that $d^C_{s,t}$ and $d^{UE}_{s,t} = d(s,t) - d^C_{s,t}$ permit subflows of the *SO* solution. Such a *SO* flow must exist by Theorem 1 and Corollary 1.

The first step is to compute the *UE* subflow, f^{UE}, given *UE* demand. From the previous section: this exists and is computationally tractable. Any feasible subflow, f^C, with demand $d^C_{s,t}$ such that the total flow along link *e* satisfies $f^C_e + f^{UE}_e \leq \bar{f}_e^{SO}$ has travel time equal to the *SO* solution, and the flow $f^C_e + f^{UE}_e$, by Lemma 1, is an *SO* solution.

We can compute f^C with the following linear program:

$$\max_{f_e^C} 1$$

subject to

$$\sum_{e \in out(v)} f_e^C - \sum_{e \in in(v)} f_e^C = \sum_t (d_{v,t}^C) \qquad \forall v \in V$$

$$\sum_{e \in in(v)} f_e^C - \sum_{e \in out(v)} f_e^C = \sum_s (d_{s,v}^C) \qquad \forall v \in V$$

$$0 \leq f_e^C \leq \bar{f}_e^{SO} - f_e^{UE} \qquad \forall e \in E$$

We know that a solution to the above linear program exists and it can be computed tractably.

The final step is to decompose the compliant flow, f^C, into a per path assignment for each origin-destination pair (s,t) in order to assign individual agents to a path. This can be done in time $O(|V||E|)$ using standard flow decomposition algorithms [see Section 3.5 of Ahuja, Magnanti, et al. (1993) for a discussion].

7.7. Experimental Results

We are interested in the viability of *opt-in* micro-tolling schemes to more efficiently utilize road networks. As such, we have undertaken an empirical study to investigate the minimal amount of compliant flow required for *SO* (d^*_{UE}) in six realistic traffic scenarios over actual road networks.

7.7.1. Scenarios

Each traffic scenario is defined by the following attributes:

1. The road network, $G(V,E)$, specifying the set of vertices and links where each link is affiliated with a length, capacity and speed limit. Networks are, following standard practice, partitioned into traffic analysis zones (TAZs) and each zone contains a node belonging to V called the centroid. All traffic originating and terminating within the zone is assumed to enter and leave the network at the centroid.
2. A trip table which specifies the traffic demand between pairs of centroids. The demand function R between nodes other than centroids is set to zero.

The following benchmark scenarios were chosen both for their diversity of topology and traffic volume and their widespread use within the traffic literature: Sioux Falls, Eastern Massachusetts, Anaheim, Chicago Sketch, Philadelphia, and Chicago-regional. All traffic scenarios are available at https://github.com/bstabler/TransportationNetworks. Figure 7.1 depicts three representative network topologies (the three smallest networks).

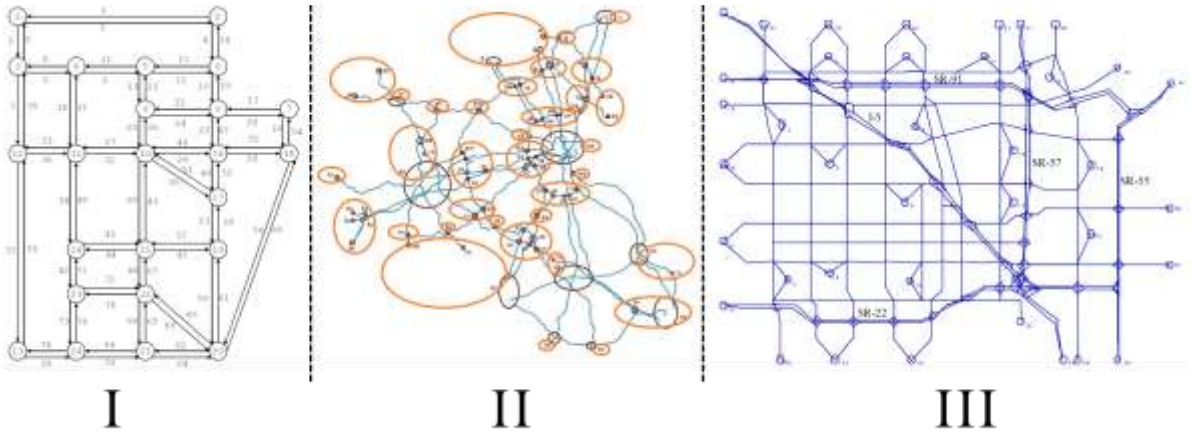


Figure 7.1 Three representative network topologies: I - Sioux Falls, SD, II - Eastern Massachusetts (ellipsoids represent different zones), III - Anaheim, CA

7.7.2. Results

Our results were obtained using a macroscopic simulator (see Section 7.11). Table 7.1 presents the percentage of flow that must be compliant in order to guarantee an *SO* solution for six different traffic scenarios. Each scenario is affiliated with the number of vertices, links, and zones comprising the affiliated road network as well as the number of trips that make up the affiliated demand.

Table 7.1 Required fraction of compliant agents given as "% compliant" for different scenarios

Scenario	Vertices	Links	Zones	Total Flow	*UE* TTT	*SO* TTT	% Improve	Threshold	% compliant
Sioux Falls	24	76	24	360,600	7,480,225	7,194,256	3.82	6.19E-11	**13.04**
Eastern MA	74	258	74	65,576	28,181	27,323	3.04	3.04E-13	
Anaheim	416	914	38	104,694	1,419,913	1,395,015	1.75	8.05E-11	
Chicago S	933	2,950	387	1,260,907	18,377,329	17,953,267	2.31	9.14E-10	19.73 19.76 27.29 49.59 53.34
Philadelphia	13,389	40,003	1525	18,503,872	335,647,106	324,268,465	3.39	4.20E-09	
Chicago R	12,982	39,018	1790	1,360,427	33,656,964	31,942,956	5.09	4.14E-07	

Note: The required fraction of compliant agents is given as "% compliant" for different scenarios along with network specifications for each scenario: number of vertices, links and zones followed by the Total Travel Time (TTT) at UE (0% compliant agents) and SO (100% compliant agents). The percentage of improvement of the SO TTT over the UE TTT is given as "% improve."

The columns "*UE* TTT" and "*SO* TTT" represent the total travel time (in minutes) over all agents for the case where 100% of the agents are controlled by the *UE* controller (*UE* solution) and when 100% of the agents are controlled by the *SO* controller (*SO* solution) respectively. The percentage of improvement in total travel time between *UE* TTT and *SO* TTT is also shown under "% improve."

The percentage of required compliant flow (formally $r_{UE^*}/|R|$ where $|R| = \sum_{s,t} R(s,t_1)$) as computed by the *UE* linear program (Definition 9) is presented for each scenario under "% compliant."

The results suggest that as the size of the network (i.e., the number of nodes and vertices) increases, a greater fraction of compliant travelers is needed to ensure the network achieves SO. This appears to be due to an increasing number of used paths at the *SO* solution as the network size increases. As the number of paths grows, the set of zero reduced cost paths grows more slowly, and, therefore, a higher percentage of compliant agents is required.

7.8. Targeting the Compliant Drivers in Non-stylized Traffic Models

A key problem towards implementing an opt-in micro-tolling system in real-life setting is the problem of identifying the set of agents who provide the most benefit to the system if they opt in. While it was shown above that it is possible to achieve system-optimal performance with partial compliance (Sharon et al., 2018), it is an open question as to how the set of compliant agents should be selected in the general case. Solving this problem would allow practitioners to identify the set of agents that maximizes system benefit and target them with specific incentives to become compliant. Next, we address this problem by answering the question, "Given that we can select *n* compliant agents, how should we select the *n* agents to maximize system social welfare?"

Prior work on a related problem suggests that computing the optimal set of compliant agents is NP-hard (Sharon et al., 2018). Thus, we develop heuristic methods for determining the set of compliant agents that will maximize social welfare for arbitrary *n*. These heuristics estimate the system's marginal benefit from assigning a given agent as compliant. In particular, our proposed **Time Evaluation** heuristic selects agents with lower *VOT*; the **Path Travel Time** heuristic selects agents with longer routes; and the **Difference in Marginal Cost Paths** heuristic selects agents according to the difference between the expected marginal impact of their chosen path if non-compliant and that of the path chosen if they are compliant. We present experimental results obtained from a dynamic traffic assignment simulation of three real-world traffic scenarios. The results show that assigning the compliant set according to the Difference in Marginal Cost Paths heuristic results in the best overall performance over the different scenarios. Moreover, the results suggest that a significant improvement in traffic flow can be achieved when as little as 7.6% of the agents are compliant

7.9. The Traffic Model

We consider a scenario where a set of agents, A, must be routed across a traffic network given as a directed graph, $G(V,E)$. Each link $e \in E$ has a travel time, l_e, defined to be the amount of time needed to traverse e. While l_e may change with the number of agents using e, we use l_e to denote the travel time assuming current conditions remain the same. A path, π, is an ordered set of adjacent links. The travel time of π is defined to be $l_\pi = \sum_{e \in \pi} l_e$. Each link e is also assigned a toll

value τ_e that may change at every discrete time-step k. For any path π we define the total tolls along π as $\tau_\pi = {}^{P}{}_{e\in\pi}\,\tau_e$.

Each agent $a \in A$ begins from a source node, $s_a \in V$ at time t_a and travels towards a target node, $t_a \in V$. A path, π, is *valid* for a given agent, a, if it leads from s_a to t_a. We denote the *VOT* for agent a as v_a, i.e., the agent's valuation of a delay of one time unit. Agents are assumed to be self-interested and, hence, follow the least cost path leading from s_a to t_a. In this work, we define two types of agents:

- **Compliant** - compliant agents are subject to tolls. As a result, a compliant agent, a, seeks to minimize the generalized cost of its route: $Cg(a,\pi) = l\pi \cdot va + \tau\pi$.
- **Non-compliant** - non-compliant agents are not subject to tolls. As a result a non-compliant agent, a, seeks to minimize only the travel time component of its route: $Cl(a,\pi) = l\pi \cdot va$.

Since travel time and toll values change, we assume agents continually re-optimize their chosen route according to current conditions.[23] As a result, an agent might change its planned route at every node along its path.

In addition to agents traveling the network, this work considers a system manager that selects the set of compliant agents. We assume selected agents always opt in to the system. The compliant set must be chosen in a way that minimizes travel cost, defined to be

$$-\sum_{a\in\mathcal{A}} l_a \cdot v_a .$$

where l_a is the actual travel time experienced by agent a. Toll costs are considered as transfer payments and thus excluded from total travel cost.

7.10. Selecting Compliant Agents

This section focuses on traffic scenarios where a subset of the agents is compliant with Δ-tolling and are thus traveling on a path, π, that minimizes $C_g(a,\pi)$ over all valid paths. The rest of the agents are considered as non-compliant with tolls and are thus traveling on a path, π, that minimizes $C_l(a,\pi)$ over all valid paths. In contrast to the research presented above (in Section 7.4), the following research also considers scenarios where the set of compliant agents is insufficient to achieve a SO flow. Specifically, we address the question, "Given limited resources that allow recruiting n agents to be compliant. Which set of n agents will the system benefit most from them being compliant?"

We propose three heuristic methods for selecting the compliant set of agents. These methods all fit into a family of methods that first assign each agent, $a \in A$, a value from a heuristic function and then select n compliant agents using the inverse of the cumulative distribution function (CDF)

[23] In principle, agents may predict changing latencies and toll values. However, including prediction in our work requires assuming a model for how agents would predict travel time and tolls.

of the heuristic value for a single agent selected uniformly from A. More formally, each method is defined by a heuristic function, $h : A \rightarrow R$, that maps agents to real values. Let $A \sim A$ be an agent sampled from A with uniform probability on all agents and define H as the random variable giving the value $h(A)$. Let F be the CDF of H, i.e., F(x) is the probability that h(A) < x. Each method selects n compliant agents by selecting all agents with

$$h(a) < F^{-1}(\frac{n}{|A|})$$

to be compliant where F^{-1} is the inverse of F. In practice, the true inverse CDF is likely unknown and will need to be estimated empirically.

We propose three instantiations of this general method that differ in how they define h: Time Evaluation, Path Travel Time, and Difference between Marginal Cost Paths.

7.10.1. Time Evaluation

The *Time Evaluation* (TE) heuristic gives higher preference to agents with lower VOT (v_a). The intuition behind this heuristic is that agents with low VOT are more influenced by tolls. As a result, such agents are more likely to change their route to one that has lower tolls and, consequently, less marginal impact on the system. The TE heuristic function is defined as $h_{TE}(a) = -v_a$. This function value can be computed in time $O(1)$ for any agent.

7.10.2. Path Travel Time

The *Path Travel Time* (PTT) method gives higher preference to agents that are traveling for longer time. The intuition behind this heuristic is that agents traveling for longer time tend to have more alternative routes to choose from. Consequently, they are more likely to have the option to reroute to a path with less marginal impact. The PTT heuristic function is defined as $h_{PTT}(a) = l_\pi$ where $\pi = \text{argmin}_\pi C_l(a,\pi)$ and π is valid for a. Computing this heuristic requires computing the shortest path [time complexity of $O(|V|^2)$] (Dijkstra, 1959).[24]

7.10.3. Difference between Marginal Cost Paths

Our final heuristic, denoted DMCP, uses the difference between the tolls along the compliant path and tolls along the non-compliant path. The marginal cost tolls for agent, a, along any path, π, represent the system's utility loss from agent a when routing through path π. The system's benefit from assigning an agent as compliant can, therefore, be computed as the difference between marginal cost tolls along its compliant and non-compliant routes. For agent a, denote the compliant path $\pi_g = \text{argmin}_\pi C_g(a,\pi)$ and the non-compliant path $\pi_l = \text{argmin}_\pi C_l(a,\pi)$. The DMCP heuristic function is defined as $h_{DMCP} = (1-\alpha)(\tau_{\pi_g} - \tau_{\pi_l}) + \alpha h_{TE}(a)$ where α is a small, positive constant (we use 0.01). The effect of the second term is to act as a tie-breaker when many agents have similar values for the first term. Unlike our previous two methods, preliminary experiments showed that approximately 80% of the agents had similar DMCP values on one of our tested scenarios. When

[24] Optimizations may be applied to lower the complexity. In any case, the heuristic is feasible to compute.

a large number of agents have a similar h value, it becomes difficult to differentiate which ones we want to select as compliant. The weighted combination allows us to prefer lower VOT agents when agents have similar values for $\tau_{\pi_g} - \tau_{\pi_l}$. Computing this heuristic requires computing the shortest path with each cost function [time complexity of $O(|V|^2)$]. In our experimental setting, a marginal cost toll (MCT) is infeasible to compute in practice. Thus, we use Δ-tolling to compute the tolls for the DMCP heuristic.

7.11. Empirical Study

We compare the relative performance of the proposed heuristics in several simulated traffic scenarios. In contrast to prior work on partial compliance presented in Section 7.4, we use a more realistic cell-transmission model simulator and use Δ-tolling as a real-time approximation method to MCT. We design our empirical study to address the following questions:

1) Do the proposed heuristics improve over a random assignment of compliant and non-compliant agents?
2) Which of the proposed heuristic methods performs best and under what compliance levels?

Analyzing the results of the initial experiments led us suspect that the optimal R parameter in Δ-tolling changes as a function of the compliance levels. This understanding, in turn, led us to a second set of experiments, aiming to address the question:

3) How does the compliance level relate to the optimal R value in Δ-tolling?

In all experiments, our metric of interest is total travel cost as defined in Section 7.9.

We compare the relative performance of the proposed heuristics within a dynamic traffic assignment simulator (Chiu et al., 2011), which models traffic through the *cell transmission model* (Daganzo, 1994a, 1995a). See Section 7.4 for more details.

7.11.1. Traffic Scenario Specification

We evaluated the performance of the different heuristics using three traffic scenarios: Sioux Falls, Austin, and San Antonio. Each scenario is specified by a network and a demand table that provides the source node (s_a), start time (t_a), and destination node (d_a) for each agent. Following are the network and demand table sizes for each scenario:

- **Sioux Falls** – (LeBlanc et al., 1975) — this scenario is widely used in the transportation research literature (Bar-Gera et al., 2013; Levin and Boyles, 2015), and consists of 76 directed links, 24 nodes (intersections), and 28,835 agents spanning 3 hours.
- **Austin** – (Levin et al., 2015b) — this network consists of 1,247 directed links, 546 nodes, and 62,836 agents spanning 2 hours during the morning peak.
- **San Antonio** - this network consists of 1,662 directed links, 864 nodes, and 10,858 agents.

The networks for all scenarios are depicted in Figure 7.2. The traffic scenarios are available online at https://goo.gl/SyvV5m.

(a) Sioux Falls (b) Austin (c) San Antonio

Figure 7.2 Traffic scenarios used in the experiments

During simulation, agents respond to changing link travel times and toll values by adapting their routes at each node. In particular, agents compute the minimum cost path from their current node n to their target t_a according to their cost function (C_g if compliant; C_l if non-compliant).

The simulation settings were identical to those presented earlier. For each compliance level and heuristic method we run the simulator 10 times and average the resulting total social welfare values.

7.11.2. Determining Heuristic Thresholds

Our three proposed heuristic methods require the empirical CDF of heuristic values over agents. For the TE heuristic, we simply use the inverse Dagum distribution. For the PTT and DMCP heuristics we estimate the inverse CDF by running the simulation with all vehicles as non-compliant.[25] When an agent, a, enters the system we compute $h(a)$ and sort the heuristic's values of all agents once the simulation is complete. If the sorted h values are indexed as $h_0 \ldots h_i \ldots h_{|A|}$ then the empirical CDF is defined as $F^{-1}(x) = h_{|A| \cdot x}$ for $0 \leq x \leq 1$.

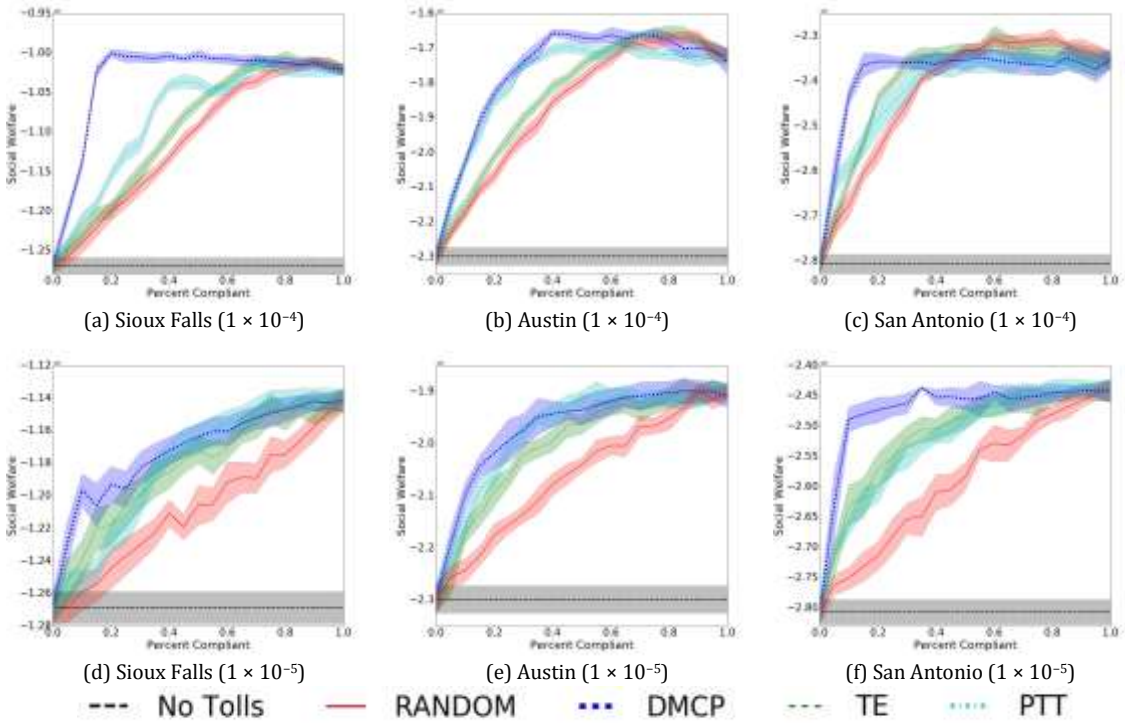
Due to stochasticity in the VOT of agents, using the empirical inverse CDF may result in greater or fewer than n compliant agents. When plotting results we use the true compliance level but then aggregate results to the nearest 5% of compliance level when averaging performance for each compliance level. For example, if a threshold results in 16% agents being compliant then we record and present the compliance level as 15% when averaging results.

[25] In real-life scenarios, a CDF function can be approximated for the PTT and DMCP heuristics through sampling of real-life observations.

7.11.3. Heuristics Comparison

Our main empirical analysis compares our three heuristic methods for various levels of compliance. We also include a baseline (denoted RANDOM) that selects compliant agents randomly. The top row of Figure 7.3 shows results for each heuristic as we vary the compliance level with $R = 1 \times 10^{-4}, \beta = 4$ [the parameter settings used by Sharon et al. (2017a)].

We first note that in all scenarios and for all heuristics (Figure 7.3, top row), the system's performance increases to an optimum and then remains constant or decreases. We hypothesize that the decrease in performance is most likely related to an R value that is too high—causing performance to deteriorate as more agents become susceptible to spiking toll values and oscillation. We test this hypothesis by repeating the same set of experiments with $R = 1 \times 10^{-5}$. We display results for these experiments in the bottom row of Figure 3.3. These results suggest that the system can benefit from a higher R value when less agents are compliant. In the following subsection we will revisit this observation.



(a) Sioux Falls (1×10^{-4}) (b) Austin (1×10^{-4}) (c) San Antonio (1×10^{-4})

(d) Sioux Falls (1×10^{-5}) (e) Austin (1×10^{-5}) (f) San Antonio (1×10^{-5})

The x-axis gives the fraction of agents who are compliant and the y-axis gives the total social welfare: $^{P}_{a \in A} -v_a \cdot l_a$. The "No Tolls" baseline corresponds to zero compliant agents ($|C| = 0$). The ideal result is to have as high a social welfare value as possible with a small number of compliant agents.

Figure 7.3 Each figure shows the average social welfare for each heuristic method

We observe the DMCP heuristic to perform best—in Sioux Falls and San Antonio it reaches the maximal or near maximal observed performance with approximately 20% of agents compliant when $R = 1 \times 10^{-4}$. In Austin ($R = 1 \times 10^{-4}$), DMCP requires 40% of agents to be compliant to reach optimal social welfare—half as many as TE or the baseline. With $R = 1 \times 10^{-5}$, DMCP also leads to

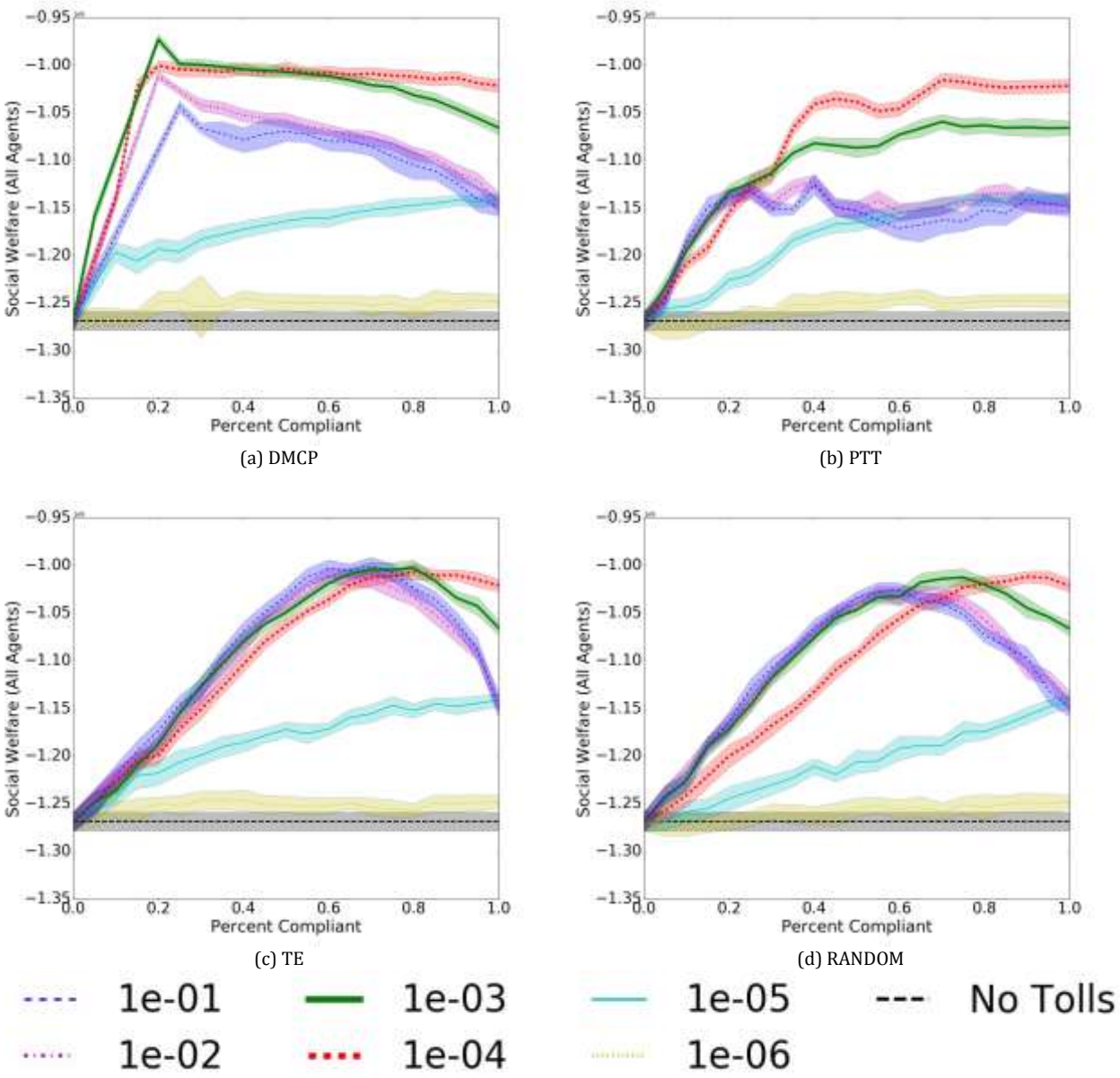
greater social welfare with less agents compliant. The PTT heuristic performs second best until approximately 60% of agents are compliant at which point the performance gap between all heuristics is small. Using agent's VOT (TE) is a small improvement over randomly selecting compliant agents (RANDOM). We also note that RANDOM perform slightly better than the proposed heuristics in the San Antonio ($R = 1\times10^{-4}$) experiment for high compliance levels. This result may indicate that it is possible for our heuristics to find local optima since they are selecting compliant agents greedily.

7.11.4. Setting Delta-tolling Parameters

As noted in the previous subsection, it may be better to have less compliant agents for certain values of the R parameter. Sharon et al. 2017b reported $\beta = 4$ and $R = 10^{-4}$ as the best performing static values overall. However, these values were reported for a scenario where 100% of the agents were compliant with Δ-tolling. Lower R values mitigate negative effects due to rapidly changing tolls. When tolls change quickly, many agents may switch to a new path which causes the travel time (and tolls) on the new path to spike causing demand to swing back the other way. When only a subset of agents complies with tolls, we hypothesize that a higher R value ($> 10^{-4}$) will perform better. The reasoning behind this hypothesis is that only agents that are affected by the tolls are susceptible to oscillation, and so fewer compliant agents would result in less oscillation of traffic. Moreover, a higher R value contributes to a toll value that is more reactive to observed traffic.

To test this hypothesis, we evaluate different values of R for each of our heuristics at different compliance levels. We also compare different R values for our RANDOM baseline. $\beta = 4$ in all experiments. Figure 7.4 contains the results for each method.

Across heuristics we see that higher R values lead to worse social welfare as the number of compliant agents increases. In Figure 7.4(a), we see that the maximal performance obtained by the DMCP heuristic is sensitive to the R parameter. For $R \geq 1 \times 10^{-4}$, social welfare peaks at approximately the 20% compliance level and then remains constant or decreases. The height of the peak is greatest for $R = 1 \times 10^{-3}$. In Figure 7.4(b), we see that PTT is less sensitive to the R parameter. $R = 1 \times 10^{-4}$ performs the best across all compliance levels. With $R = 1 \times 10^{-4}$, performance does not decrease as the number of compliant agents increases. Finally, we see similar performance between TE and RANDOM: higher R values lead to better performance with less compliant agents. However, as the compliance level increases performance decreases more for higher R values.



(a) DMCP

(b) PTT

(c) TE

(d) RANDOM

Each figure shows seven curves, each representing a different R value. The x-axis represents different compliance levels (between 0 and 1) while the y-axis gives the social welfare (higher values are better).

Figure 7.4 Compliance level (x-axis) vs. social welfare (y-axis) for different R values and different heuristics

7.12. Discussion and Future Work

This chapter discussed a scenario where a set of agents traverse a congested network, while a centralized network manager is seeking to optimize the flow (minimizes total travel time) by influencing the route assignment of a set of compliant agents. For a stylized, macroscopic traffic model a methodology was presented for computing the minimal volume of traffic flow that needs to be compliant in order to reach a state of optimal traffic flow. Moreover, the methodology extends

to inferring which agents should be compliant and how exactly the compliant agents should be assigned to paths. Experimental results demonstrate that the required percentage of agents that are compliant is small for some scenarios but can be greater than 50% in others.

For non-stylized traffic models, we consider the problem of how to select compliant agents. Since selecting the optimal set of compliant agents has been suggested to be NP-hard, we proposed three heuristic methods for doing so. In experiments with a dynamic traffic assignment simulator we demonstrate that across all traffic scenarios and all heuristics, any number of compliant agents is better than none (assuming that the parameters for Δ-tolling are correctly tuned). This result indicates that if even a small number of agents can be incentivized to participate in a marginal cost tolling system (such as Δ-tolling) we may see an improvement in the system's performance. Furthermore, this result demonstrates feasibility of an opt-in micro-tolling system when only a subset of agents opts in.

While any number of compliant agents is better than none, we show that our proposed heuristic methods lead to further improvements in system performance compared to assigning a random subset of agents to be compliant. In particular, across all traffic scenarios we see that the DMCP heuristic can obtain close to the performance of 100% compliance. In fact, in the San Antonio scenario with 7.6% compliant agents, we see an improvement of 10.9%, and in Sioux Falls we see an improvement of 21.1% with 18.7% compliant agents.

In our empirical analysis we make two assumptions that may not be reflective of opt-in micro-tolling systems in practice. First, we assume that agents selected by one of our heuristic methods become compliant with probability 1. In the real world it is unlikely that all selected agents will decide to opt in. Future work should consider the robustness of our proposed heuristic methods when selected agents may remain non-compliant with some probability. Second, we considered traffic scenarios where each agent makes a single trip through the network while in the real world, people may make multiple trips every day. In such a setting, it may be possible to obtain better performance by considering the frequency of trips that an agent makes.

Finally, it is also important to consider how to incentivize agents to participate in a micro-tolling system. A first step towards addressing this problem could be to investigate how travel times differ between compliant and non-compliant agents.

Chapter 8. Performance Guarantees for Micro-Tolling

As was discussed above, charging *marginal cost tolls* (MCT) from self-interested agents participating in a congestion game leads to optimal system performance, i.e., minimal total travel time. This chapter studies the impact of charging MCT with some fixed factor error on the system's performance. We prove that underestimating MCT results in a system performance that is at least as good as that obtained by not applying tolls at all. This result might encourage adoption of MCT schemes with conservative MCT estimations. Furthermore, we prove that no local extrema can exist in the function mapping the error value, r, to the system's performance, TSTT(r). This result implies that accurately calibrating MCT for a given network can be done by identifying an extremum in TSTT(r) which, consequently, must be the global optimum. Experimental results from simulating several large-scale, real-life traffic networks are presented and provide further support of our theoretical findings.

8.1. Preliminaries

This chapter assumes a standard flow model that is common in the routing and congestion games literature (Yang et al., 2007; Pigou, 1920b; Roughgarden and Tardos, 2002). The terminology for this model follows the previous two chapters.

8.1.1. The Flow Model

Recall that the flow model is assumed to be composed of a directed graph $G(V,E)$, and a demand function $d(st) \rightarrow \mathsf{R}^+$ mapping a pair of vertices, $s,t \in V^2$, to a non-negative real number representing the required amount of flow between source, s, and target, t.

As before, the variable f_π represents the flow volume assigned to a path, π. Similarly, f_e is the flow volume assigned to link e. A flow is defined as *valid* if:

- $f_\pi \geq 0$ for all paths π; that is, no path is assigned negative flow.
- the flow on each link (f_e) equals the summation of flows on all paths of which e is a part. That is, $f_e = \sum_{\pi \in \Pi_e} f_\pi$ where Π_e is the set of acyclic paths that include link e.

A valid flow is defined as feasible if it satisfies $d(st)$.

> **Definition 10** (Feasible flow). *A flow is defined as feasible if it is valid and the traffic demand is satisfied, that is,* $\sum_{\pi \in \Pi_{st}} f_\pi = d(st)$ *for all node pairs* (s,t).

For this work we make the following regularity conditions on the travel time functions, which are a standard assumption in the transportation literature (Patriksson, 1994).

Assumption 3. *The travel time function* $l_e(f_e)$ *is non-negative, convex, and non-decreasing for each link* $e \in E$.

Define the total travel time associated with link e as $T_e(f_e) = l_e(f_e)f_e$. The total system travel time (TSTT), for a given flow f, is TSTT(f) = $\sum_{e \in E} T_e(f_e)$.

A feasible flow f is defined as a *system optimum* (*SO*) if TSTT(f) is minimal over the set of feasible flows. We use TSTT(*UE*) to denote the total travel time at the UE solution. Similarly, TSTT(*SO*) denotes the total travel time at the SO solution.

It is easy to show that Assumption 3 implies that TSTT(f) is strictly convex in f. As a result, unique *UE* and *SO* flows exist[26] (Beckmann et al., 1956; Dafermos and Sparrow, 1969).

8.1.2. Applying Tolls

A recent body of work (Yang et al., 2004; Zhou et al., 2015; Chen et al., 2018; Sharon et al., 2017a) examined mechanisms for assigning tolls to links with the goal of affecting the route choice of self-interested agents. Such work assumes that drivers are willing to sustain time delays in return for monetary gain (or avoiding monetary loss). This line of work requires translating time delays into monetary value using the agents' *value of time* (VOT). VOT represents the agents' monetary evaluation of a single unit of time. Following this line of work, we make the following definition and assumption.

> **Definition 11** (generalized-cost *UE* (*GUE*)). *Let τ_π be the toll associated with path π (the sum of the tolls on its constituting links). A feasible flow f is a GUE if for every $s,t \in V^2$ and $\pi_a, \pi_b \in \Pi_{st}$ with $f_{\pi_a} > 0$ it holds that $l_{\pi_a}(f) \times VOT + \tau_{\pi_a} \leq l_{\pi_b}(f) \times VOT + \tau_{\pi_b}$. In other words, at GUE, no amount of flow can be rerouted to a path with lower generalized cost (travel time times VOT plus toll) when the rest of the flow is fixed.*

Assumption 4. *A solution for a traffic scenario follows the GUE principle.*

Note that Definition 11 makes a latent assumption of homogeneous VOT. Nonetheless, Assumption 4 does not require homogeneous VOT. Dealing with heterogeneous VOT, however, requires a different definition for *GUE*, one that addresses a set of agents instead of a flow of agents. This, in turn, would require a discrete traffic model. Though we expect that the main contributions of this report ought to extend naturally to that case, for clarity of presentation, we leave consideration of discrete models for future work.

A traffic scenario is said to be *toll-optimized* if the set of tolls (τ) causes the *SO* and *GUE* solutions to align. Specifically, a sufficient (yet not necessary) condition for an optimized system is that τ equals the set of marginal cost tolls, τ^{MCT} (Beckmann et al., 1956; Dafermos and Sparrow, 1969).

> **Definition 12** (Marginal cost toll). *In marginal cost tolling, each agent (infinitesimally portion of the flow) is charged a toll equivalent to the damage it inflicts on the system. When the travel time functions are differentiable, the MCT for link e equals*

[26] A flow, f, is considered unique if it maps to a single assignment to all f_e variables. The f_π variables, by contrast, might have non-unique values in f.

$$f_e \frac{\partial l_e}{\partial f_e}$$

That is, the increase in travel time caused by adding one more unit of flow to link e (i.e., $\frac{\partial l_e}{\partial f_e}$) multiplied by all the flow that suffers from this increase (i.e., f_e).

Assuming that the travel time functions are known and differentiable is not practical in many traffic models, e.g., the cell transmission model (Daganzo, 1994b, 1995b) or microsimulation models (Yang and Koutsopoulos, 1996; Dresner and Stone, 2008; Krajzewicz et al., 2002). Such an assumption is certainly not practical for real-life traffic networks. Consequently, Sharon et al. (2017a; 2017b) introduced Δ-tolling, a model-free method for approximating MCT when the travel time function is unknown. Despite showing reductions in TSTT across markedly different traffic models, Δ-tolling, or any mechanism that approximates MCT for that matter, is not guaranteed to be toll-optimized. This fact poses a major problem since applying tolls that are different than MCT might result in arbitrarily worse TSTT compared to that at the *UE* (such a scenario is discussed later). This chapter makes a first attempt to examine the impact of applying inaccurate MCT. Specifically, it provides conditions under which the system performance (TSTT) will be no worse than that at the *UE* solution.

8.2. Inaccurate Marginal Cost Tolls

We consider a scenario where the tolls assigned to all links in a network are off by some factor from the MCT. Such a scenario might represent a systemic error in evaluating the β parameter in Δ-tolling [see Sharon et al. (2017b) for exact details]. Another relevant scenario is one in which MCT can accurately be computed in units of time delays (e.g., by computing $\int_p \frac{\partial l_e}{\partial f_e}$). In such cases, a systemic error in the evaluation of the VOT would result in a constant factor, MCT error.

> **Definition 13** (Imperfect MCT scenario). *A scenario is said to be Imperfect MCT if the toll affiliated with every link, $e \in E$, equals $r \cdot \tau_e^{MCT}$ for some error factor $r \geq 0$. Where τ_e^{MCT} is the true MCT for link e.*

Define the *GUE* flow for an Imperfect MCT scenario with error r as f^r. As a result, TSTT(f^r) denotes the TSTT for the *GUE* flow. Since f^r is a function of r, we use TSTT(r) instead of TSTT(f^r) for brevity.

8.3. Bounding the System's Performance

The following section presents the main contribution of this work, i.e., provable bounds on the system's performance (TSTT) as a function of the error factor r. We begin with several supporting lemmas.

Lemma 4. *A GUE flow, f, for an Imperfect MCT system minimizes*

$$r \sum_{e \in E} [f_e l_e(f_e)] + (1 - r) \sum_{e \in E} \left[\int_0^{f_e} l_e(z) dz \right] \tag{8.1}$$

subject to f being feasible (see Definition 10).

Proof. The Karush-Kuhn-Tucker (KKT) optimality conditions for the above optimization problem (see, for instance, Bertsekas, 1999) include the following:

$$f_p \geq 0 \qquad \forall p \tag{8.2}$$

$$l_p(f_p) + r f_p l_{p0}(f_p) \geq c_{st} \qquad \forall s,t \in V^2,\ p \in \Pi_{st} \tag{8.3}$$

$$f_p[l_p(f_p) + r f_p l_p^0(f_p) - c_{st}] = 0 \qquad \forall s,t \in V^2,\ p \in \Pi_{st} \tag{8.4}$$

The condition given in Equation 8.2 enforces non-negative path flows. The condition given in Equation 8.3 enforces that c_{st} is the minimal generalized cost over all paths leading from s to t. The condition given in Equation 8.4 enforces that if a path is used ($f_\pi > 0$) its travel time must be equal to c_{st}. These conditions are met if and only if the solution is a *GUE* flow (see Definition 11).□

Theorem 3. A unique GUE flow for an Imperfect MCT scenario exists.

Proof. In order to prove this lemma it is sufficient to show that the objective function given in Lemma 4 (Equation 8.1) is strictly convex. The Hessian matrix for Equation 8.1 ($H \in \mathbb{R}^{|E|\times|E|}$) is diagonal, where each entry on the diagonal (representing one edge, e ∈ E) equals:

$$(r+1)\frac{\partial l_e}{\partial f_e} + r f_e \frac{\partial l_e^2}{\partial f_e^2} \tag{8.5}$$

For any edge, *e*, the value of Equation 8.5 is strictly positive since:

- $r \geq 0$, see Definition 13.
- $\frac{\partial l_e}{\partial f_e} > 0$, see Assumption 3.
- $f_e \geq 0$, see Definition 10.
- $\frac{\partial l_e^2}{\partial f_e^2} \geq 0$, see Assumption 3.

A diagonal matrix with strictly positive entries along its diagonal is positive definite. As a result, Equation 8.1 is strictly convex.

Given that a unique *GUE* flow that minimizes Equation 8.1 exists, we now turn to evaluating the TSTT value for three key *r* values: 0, 1, and ∞.

Lemma 5. TSTT(0) = TSTT(*UE*)

Proof. Setting $r = 0$ in Equation 8.1 results in the minimization of

$$\int_0^{f_e} l_e(z)dz$$

$_{e\in E}$ ^{0}

subject to the feasibility constraint. This minimization problem results in the UE flow Beckmann et al. (1956). □

Lemma 6. TSTT(1) = TSTT(*SO*)

Proof. Setting $r = 1$ in Equation 8.1 results in the minimization of

$$\text{X} \quad f_e l_e(f_e)_{\,e\in E}$$

subject to the feasibility constraint. This minimization problem translates to minimizing TSTT, i.e., an SO flow Beckmann et al. (1956). □

Lemma 7. TSTT(∞) = TSTT(f^{∞}) *where f^{∞} is a UE solution for a scenario in which the travel time affiliated with every path, π, equals*

$$f_{\pi} \frac{\partial l_{\pi}}{\partial f_{\pi}}$$

Proof. Dividing Equation 8.1 by a constant (specifically r) preserves the minimal assignment and yields

$$\sum_{e\in E}[f_e l_e(f_e)] + \frac{1-r}{r}\sum_{e\in E}[\int_0^{f_e} l_e(z)dz \quad] \tag{8.6}$$

Since $\lim_{r\to\infty}(1-r)/(r) = -1$, Equation 8.6 converges to

$$\sum_{e\in E}[f_e l_e(f_e)] - \sum_{e\in E}[\int_0^{f_e} l_e(z)dz \quad] \tag{8.7}$$

The KKT optimality conditions for minimizing Equation 8.7 under the feasibility constraints include:

$$f_{\pi} \geq 0 \qquad \forall \pi \tag{8.8}$$

$$f_{\pi} l_{\pi 0}(f_{\pi}) \geq c_{st} \qquad f_{\pi}(f_{\pi} l_{\pi 0} \qquad \forall st,\ \pi \in \Pi_{st} \tag{8.9}$$

$$(f_{\pi}) - c_{st}) = 0 \qquad \forall st,\ \pi \in \Pi_{st} \tag{8.10}$$

from which the *UE* definition (see Section 8.1.2) holds if the travel time function for any path π is replaced by

$$f_{\pi} \frac{\partial l_{\pi}}{\partial f_{\pi}}. \square$$

Lemma 7 implies that at r = ∞ the system performance (TSTT) can be arbitrarily worse than TSTT(SO) or TSTT(UE). As an example, consider the network depicted in Figure 8.1. The travel

time on the bottom link equals the fraction of flow that is assigned to it. If, for instance, 25% of the flow is assigned to the bottom link then the travel time on the top link equals 0.25. The travel time on the top link equals a constant, *C*, regardless of the amount of flow that is assigned to it. For *C* ≥ 2 the SO and UE align and TSTT(SO) = TSTT(UE) = 1 · R(st). Since the travel time on the top link is not a function of the flow, $MCT = x\frac{\partial l_a}{\partial x} = 0$ for the top link while MCT = x∂l∂xπ ≥ 0 for the bottom link. As a result, at r = ∞, 100% of the flow from *s* to *t* would travel the top link while 0% would travel the bottom link. Such a flow would result in TSTT = *C* ·R(st). It is easy to see that as *C* increases so does the difference between TSTT(∞) and TSTT(SO) or TSTT(UE), potentially to infinity.

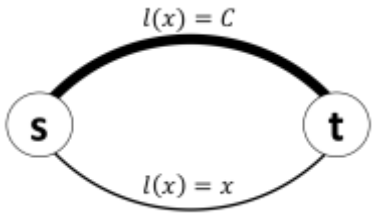


Figure 8.1 A network where setting r = ∞ results in an arbitrary worse system performance compared to both the UE and SO solutions

Given that no bound on the system's performance can be given for $r = \infty$ we turn to examine bounds on other values of *r*. We start by examining values of *r* that fall between zero and one.

Lemma 8. *Any two error values* $0 \le r_1 < r_2 < 1$ *satisfy TSTT(*r_1*)* ≥ TSTT(r_2).

Proof. For simplicity of presentation we use *U*(*r*) to denote

$$\sum_{e \in E} \left[\int_0^{f_e^r} l_e(z)dz \right]$$

Any *GUE* flow f^r must minimize Equation 8.1 (Lemma 4). That is, subject to being feasible, f^r minimizes the expression $rT(r) + (1 - r)U(r)$. Minimizing Equation 8.1 under r_1 requires that

$$r_1T(r_2) + (1 - r_1)U(r_2) \ge r_1T(r_1) + (1 - r_1)U(r_1)$$

and as a result

$$r_1(TSTT(r_2) - \text{TSTT}(r_1)) \ge (1 - r_1)(U(r_1) - U(r_2)) \tag{8.11}$$

Similarly, minimizing Equation 8.1 under r_2 requires that

$$r_2(TSTT(r_2) - \text{TSTT}(r_1)) \le (1 - r_2)(U(r_1) - U(r_2)) \tag{8.12}$$

Assume, in contradiction to the lemma, that TSTT(r_2) − TSTT(r_1) > 0. Since $1 - r_2 > 0$ and $r_2 > 0$, Equation 8.12 would require $U(r_1) - U(r_2) > 0$. Since all the components of Equations 8.11 and 8.12 are strictly positive, we can rewrite them as:

$$\frac{r_1}{1 - r_1} \ge \frac{U(r_1) - U(r_2)}{T(r_2) - T(r_1)} \tag{8.13}$$

$$\frac{r_2}{1-r_2} \leq \frac{U(r_1)-U(r_2)}{T(r_2)-T(r_1)} \tag{8.14}$$

From Equations 8.13 and 8.14 we obtain

$$\frac{r_1}{1-r_1} > \frac{r_2}{1-r_2} \tag{8.15}$$

Since the function $f(r) = r/(1-r)$ is continuous and strictly increasing for $r < 1$ then Equation 8.15 must satisfy $r_1 \geq r_2$ in contradiction to the lemma's premise. □

Next we turn to examine the behavior of error values that are greater than one.

Lemma 9. *Any two error values* $1 < r_1 < r_2$ *satisfy TSTT(*r_1*)* ≤ TSTT(r_2).

Proof. Assume, in contradiction to the lemma, that TSTT(r_2) − TSTT(r_1) < 0. Since $1 - r_1 < 0$ and $r_1 > 1 > 0$, Equation 8.11 requires $U(r_1)-U(r_2) > 0$. Even though the signs of $(T(r_2)-T(r_1))$ and $(1-r_1)$ and $(1-r_2)$ are in contrast to the case presented in Lemma 8, rearranging Equations 8.11 and 8.12 still result in Equations 8.13 and 8.14, which leads to the inequality in Equation 8.15. Since the function $f(r) = r/(1-r)$ is continuous and strictly increasing for $r > 1$ then Equation 8.15 must satisfy $r_1 \geq r_2$ in contradiction to the lemma's premise. □

Following Lemma 8 and 9, we can now provide bounds for an Imperfect MCT system.

Theorem 4. *If* $0 \leq r \leq 1$ *then TSTT(*r*)* ≤ TSTT(*UE*).

*Proof. TSTT(*0) = TSTT(*UE*) (Lemma 5) and TSTT(r) is non increasing in the interval [0,1) (Lemma 8). Also TSTT(1) = TSTT(*SO*) ≤ TSTT(*UE*) (Lemma 6). □

Theorem 5. *If* $r \geq 1$ *then TSTT(*r*)* ≤ TSTT(f^∞) *when* f^∞ *is a UE solution for a scenario where the travel time on every path,* π*, equals* $f_\pi l'_\pi(f_\pi)$.

Proof. TSTT(∞) = TSTT(f^∞) when f^∞ is a UE solution for a scenario where the travel time for every path, π, equals $f_\pi l_\pi^0(f_\pi)$ (Lemma 7). TSTT(r) is non decreasing for $r > 1$ (Lemma 9). Also TSTT(1) = TSTT(*SO*) ≤ TSTT(∞) (Lemma 6). □

Theorem 4 implies that when underestimating MCT by a constant factor, $0 \geq r < 1$, the systems performance cannot be worse that the one obtained by the UE solution, TSTT(*UE*).

Theorem 5 implies that when overestimating MCT by a constant factor, $r > 1$, the systems performance cannot be worse than TSTT(∞). However since TSTT(∞) can be arbitrary worse than TSTT(*UE*) and TSTT(*SO*), this bound is not as useful as the one provided for the previous case, $0 < r < 1$.

8.4. Empirical Study

In order to validate our theoretical findings, we simulated different traffic scenarios while varying the MCT error factor (r). The total system performance (TSTT) was measured for each setting and the trends were compared to the above theoretical claims.

The traffic scenario, traffic model, and experimental settings were chosen to be identical to those used in Section 8.1.

8.4.1. Results

Table 8.1 presents the six scenarios' specifications and also the system's performance (TSTT) for five different error values ($r = \{0, 0.5, 1, 2, \infty\}$). The SO solution ($r = 1$) provides the best performance (minimal TSTT), as expected. The performance for $r = \infty$ is slightly better than that at the UE solution ($r = 0$) in some cases, e.g., Sioux Falls and Philadelphia, but might be significantly worse in others, e.g., Eastern Massachusetts where TSTT(∞) was outperformed by TSTT(UE) by 15%.

Table 8.1 The system performance (TSTT) given as "T(x)" for different scenarios along with network specifications, for each scenario: number of vertices, links, zones, and total demand (Pst R(st))

Scenario	Vertices	Links	Zones	Total Demand	$T(UE)$	$T(0.5)$	$T(SO)$	$T(2)$	$T(\infty)$
Sioux Falls	24	76	24	360,600	7,480,223	7,205,048	7,194,256	7,198,091	7,222,857
Eastern MA	74	258	74	65,576	28,181	27,411	27,324	27,392	32,460
Anaheim	416	914	38	104,694	1,419,913	1,397,216	1,395,015	1,398,631	1,549,075
Chicago S	933	2,950	387	1,260,907	18,377,331	17,991,235	17,953,268	17,994,192	19,630,440
Chicago R	12,982	39,018	1790	1,360,427	33,656,969	32,078,668	31,942,957	32,096,038	38,190,675
Philadelphia	13,389	40,003	1525	18,503,872	335,647,096	325,211,099	324,268,465	325,176,216	335,296,306

Results for applying half and double the true MCT are also provided (T(0.5) and TSTT(2) respectively). Results for these values are mixed where in some cases TSTT(0.5) performs slightly better than TSTT(2) and vice versa in others. Nonetheless, r = 0.5 has a clear advantage over r = 2 since, unlike TSTT(2), the value of TSTT(0.5) is bounded by TSTT(UE) for any scenario (Theorem 4).

Figure 8.2 presents normalized values for TSTT as a function of the error factor *r*. The TSTT values (y-axis) for each curve are normalized according to TSTT(*SO*), e.g., a TSTT value of 2 correlates to double TSTT(*SO*) for the relevant curve (scenario). Consequently, TSTT(1) = TSTT(*SO*) = 1 in all the curves. The data points were computed for the range $r = [0,20]$ with a step size of 0.1. Each of the curves starts with a dot representing TSTT(*UE*). Additionally, dots on the right border of the plot represent TSTT(∞). Such dots are presented only for the Sioux Falls and Philadelphia scenarios as TSTT(∞) is out of the presented TSTT range for the rest (exact values are available in Table 8.1).

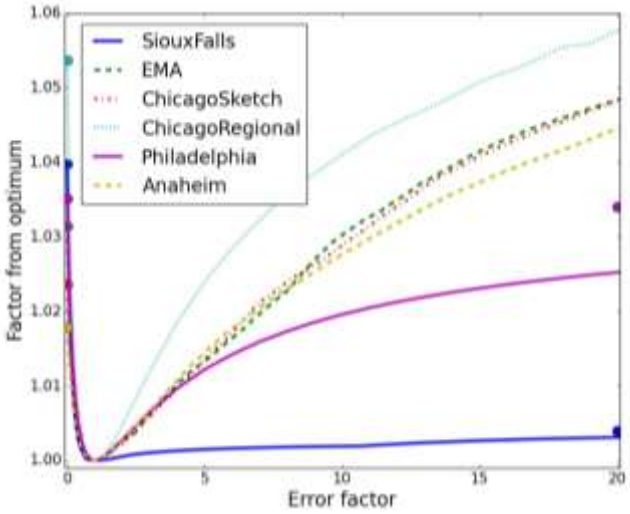


Figure 8.2 Normalized TSTT (factor from optimal TSTT) as a function of the error factor (r) for six benchmark traffic scenarios

As predicted by Lemmas 8 and 9, the curves are non-increasing in the range [0,1] and non-decreasing in the range [1,∞].

8.5. Discussion

Lemmas 8 and 9 and Theorems 4 and 5 as well as the presented experimental results lead to the following general conclusions:

- Underestimating MCT by a constant factor across a traffic network would result in a system performance that is not worse than not applying tolls at all.
- When calibrating a parameter that is a multiplier of the true MCT, a value that is locally optimal is guaranteed to be globally optimal.

The implication of these conclusions might be substantial when installing a new tolling scheme with a tunable parameter, θ, where the value of θ correlates to a fixed error in MCT. As stated in Section 8.2, this can occur when calibrating the expected drivers' VOT or the β parameter in Δ-tolling (Sharon et al., 2017a).

A simple, yet effective, approach for tuning θ would be to set it to zero initially. Then, as long as the system's performance does not deteriorate, θ can be safely increased by a small enough ϵ. Once the system's performance deteriorates this process is stopped and θ is reduced by ϵ. The above conclusions suggest that, for a small enough ϵ, traffic congestion will not deteriorate along the tuning process (excluding the step before the last) and system-optimal performance will be achieved at the final step.

8.6. Summary and Future Work

This chapter considers a traffic scenario in which MCTs with some fixed factor errors are imposed on all drivers. The system performance is analyzed with regards to the error rate and performance bounds are provided as a function of the error value.

Three main claims are proven:

1. If the error factor is lower than 1 (MCT is underestimated), the system will not perform worse than if no tolls were applied.
2. As the error factor increases from 0 to 1, the system's performance will not deteriorate.
3. As the error factor increases from 1 to infinity, the system's performance will not improve.

These claims can allow the tuning of MCT-based tolling schemes while ensuring quality of service along the tuning process.

There are many other conceivable errors besides a multiplicative, system-wide factor on the true MCT. Consequently, future work will examine scenarios with other assumptions on the toll error, such as when the assessed toll is within some bounded interval around the MCT. Finally, another promising direction for future work, inspired by Sharon et al. (2018), is to examine traffic scenarios where only a subset of the flow is compliant with errored MCT.

Chapter 9. CAV's Impacts on Long-distance Household Travel across the U.S. and Texas

9.1. Background

As the U.S. population grows, it is expected that the demand for inter-city travel will rise, running up against the limited capacity of existing infrastructure. The federal government and states continuously seek to improve long-distance mobility; however, national-scale passenger travel demand modeling is still an emerging area of research. In efforts to enable proactive planning, the Federal Highway Administration (FHWA) commissioned several studies. One of the studies produced a passenger travel demand model called rJourney that models all long-distance travel in 2010 for the entire United States (Federal Highway Administration, 2015).

While the rJourney model surpasses the limitations of traditional travel demand forecasting methods by rigorously incorporating several forms of travel behavior, the prospect of applying the model to an increasingly automated future is challenged by the fact that automated vehicles (AVs) were not a mode of choice in 2010, and therefore are not represented in the model. While traveler behavior may gradually change as the future emerges and AVs continue to enter the marketplace, the most feasible and best-validated future-looking models at hand are inevitably based upon today's knowledge.

9.1.1. AVs and Long-distance Travel

While there have been several simulations of AVs' and shared AVs' effects on intra-regional travel [e.g., Fagnant and Kockelman (2014) and Childress et al. (2015)], there is little research on inter-regional travel and how longer-distance destination and mode choices will change. LaMondia et al. (2016) explored mode choices in Michigan for trips over 50 miles in length and forecasted that over 25% of airline trips under 500 miles will shift to AVs. Such changes will have important impacts on airlines, infrastructure planning and future land use (especially around long-distance transportation facilities), highway congestion, and the travel industry more generally.

Long-distance travel is common in many countries and regions. Mercedes-Benz responded to the Google challenge in August 2013 with the S500 Intelligent Drive Autonomous Car long-distance test drive between Mannheim and Pforzheim without any driver input. Automated public vehicles may provide much of the long-distance travel between European countries (Heinrichs, 2016). Nineteen percent of Americans with disabilities report leaving their homes relatively infrequently, and are less likely to take long-distance trips (BTS, 2003). However, Meyer and Deix (2014) noted that if AVs allow disabled individuals to make the same length and number of car trips, their vehicle-miles traveled (VMT) would probably increase by more than 50%.

AVs reduce the burden of travel for drivers and may improve the quality of travel for passengers, who can now focus on more meaningful interactions with those previously focused on driving. Thanks to easier "driving," the value of travel time (VOTT) of the driver (or his/her willingness to

pay to save travel time) is expected to fall, by 20 to 50% or more, so the generalized cost of travel can fall by several dollars per hour to $6 or more per hour, for many travelers. Auld et al. (2017) applied an integrated transportation system model to analyze the impact of hypothesized CAV scenarios, varying the market penetration, capacity changes, and travel time valuations, on performance of the transportation network and changes in mobility patterns for Chicago region. The results show that an increase in capacity of 80% can be achieved with only 4% induced additional VMT. Changes in travel time cost, or VOTT savings, have a significant impact, especially at very low levels of VOTT, increasing VMT by up to 59%.

9.1.2. Extensions of Prior Models

With the impending introduction of AVs as a viable mode choice in the near future, it is necessary for today's future-looking travel demand forecasting models to incorporate them. Childress et al. (2015) used a Seattle, Washington activity-based travel model (including short-term travel choices and long term work-location and auto-ownership choices) to anticipate the impacts of AV technology introduction on regional travel (attributed to higher roadway capacities, lowered VOTT, reduced parking costs, and increased car-sharing). They estimated that higher income households are more likely to choose the AV mode, as costly technology and VOTT reductions for higher-VOTT travelers are likely to be more significant. When shared automated vehicles (SAVs) are modeled to cost $1.65 per mile (similar to costs of current ride-sharing taxi services, like Lyft and Uber), drive-alone trips were estimated to be reduced by one-third and transit shares increased by 140%, as modeled households did away with traditional vehicles and bought AVs or shifted to SAVs as well as other travel options.

Other existing projects introduced AVs as a new mode in mode choice or destination choice models. Gucwa (2014) used an activity-based model approach to simulate the travel decisions of individuals in the nine-county San Francisco Bay Area. The autonomous vehicle scenarios are modeled under different values of travel time and road capacity, using the Bay Area's Travel Model One. The mode choice confirms to a random utility model. The result showed that the automation can expect a short-run increase of 4 to 8% in daily VMT. Zhao and Kockelman (2017) extended the Austin, Texas six-county region local municipal planning organization's conventional travel demand model with new connected and autonomous vehicle (CAV) and SAV modes. The gravity model for trip distribution was replaced with a multinomial logit (MNL) model to allow destination choice to be influenced by the new modes. The mode choice model was also simplified and extended to support the new modes. Simulations varied the assumed operating and parking costs. Results suggested that by the year 2020, the introduction of these modes would add 20% demand to the region's current VMT. An added consequence is a reduction of transit system usage. Both of these were attributed to the relative value of time (VOT) of CAV and SAV travelers as well as an anticipated competitive SAV pricing scheme. Results of this report suggest that without full realization of other anticipated benefits of CAVs and SAVs (e.g., smaller headways, shared rides), overall congestion would worsen from that of today.

Energy and environment can also be affected by travel demand that corresponds to the introduction of AVs. Wadud et al. (2016) used a coherent energy decomposition framework to combine automation effects on travel and energy demand greenhouse gas emissions. Through illustrative scenarios, they found that autonomous vehicles provide potential, but not assured reductions in energy consumption and emissions, because the reductions are not directly a consequence of automation. Automation is considered to play an active role in vehicle operations, vehicle design, or transportation system design. The reductions are also related to connectivity, even without full automation. They also predicted that if the autonomous vehicles have a dominant market penetration, they could sharply reduce the energy consumption. Fagnant and Kockelman (2014) worked with an agent-based model for SAVs that simulated environmental benefits of such a fleet as compared to that of traditional, personally-owned vehicles, focusing on a dense urban core area. Simulation results indicated that each SAV may replace 11 conventional privately owned vehicles while increasing travel distances by up to 10%. When the simulation was extended to a case study of low market penetration (1.3% of trips) in Austin, Texas, each SAV was able to replace nine conventional vehicles and generated 8% more VMT on average due to empty, unoccupied travel (Fagnant et al. 2015).

This chapter investigates a possible use of rJourney to forecast traditional personal car, commercial air carrier, and personal AV mode and destination choice offers insight on future U.S. inter-city travel. Since aircraft will still travel much faster than AVs between long-distance city pairs (e.g., New York City to Los Angeles), it is intuitive that those markets could be largely immune to this new mode alternative. However, looking at what routes will be significantly changed lacks research and is important for airline and infrastructure planning. If for example the 240-mile (385 km) route between Houston and Dallas is largely dominated by AVs, interstate planners should expect higher traffic on Interstate 45 and the airport managers should expect less short-distance travel between the two cities.

This remainder of this chapter is organized as follows. First, the rJourney data set used in this research is introduced, followed by the preliminary methodology. Then, results of the research model are identified, as well as an exploration of how the model can be used to estimate how the introduction of AVs may affect overall airline industry revenue. Finally, this chapter concludes and offers future research directions.

9.1.3. Data Set Description

The rJourney data that is leveraged in this research is part of an extensive, nationwide tour-based long-distance travel model created by RSG for the FHWA. The motivation for the creation of rJourney is to study intercity travel and to enhance interstate, long-distance modeling efforts. As noted earlier, long distance travel is modeled among almost all pairwise combinations of 4,486 National Use Microdata Area (NUMA) zones as shown in Figure 9.1. As part of the rJourney effort, NUMAs are derived from both Census Bureau Public Use Microdata Areas (PUMAs) and county boundaries. The 1.17 billion rJourney tours are generated from a synthesized household population of 31.5 million, representing all long-distance travel in 2010. Destination and mode

choice are modeled with cross-nested logit, supporting four modes: automobile, bus, rail, and airlines. Trip models are organized among five purposes: business travel, commuting, personal business for shopping and relaxation, visiting friends and family, and leisure travel (Outwater et al., 2014).

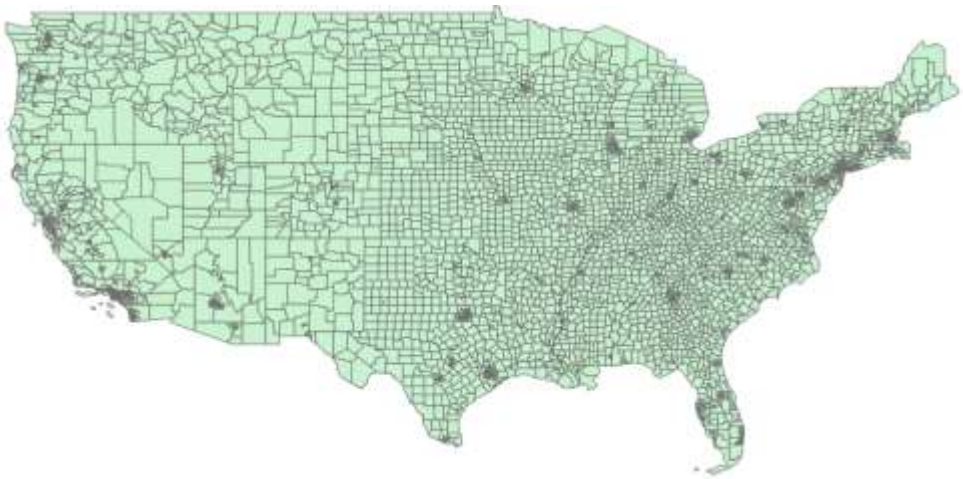
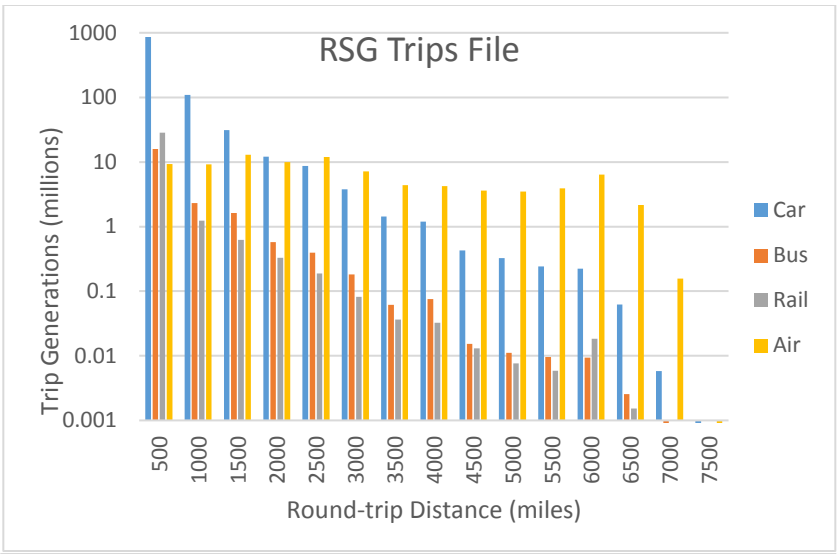


Figure 9.1 NUMA boundaries within the continental United States

The generated tours provided in the rJourney set across all trip types are distributed as shown in Figure 9.2. Distances for all modes are measured as round-trip driving distance. All tours consist of one outbound and one return trip over the same path. Important aspects to note about this distribution are that no round-trips shorter than 100 miles (161 km) are expressed in the rJourney tours data set since rJourney only looks at longer-distance trips that involve originating in one NUMA and arriving at a distant NUMA. As expected, car usage largely dominates shorter trips (less than or equal to 500 miles, or 805 km), while air travel dominates for longer ranges. Bus and rail consistently account for a small portion of all trips. The average party size in a tour is 2.15 people.

a.



b.

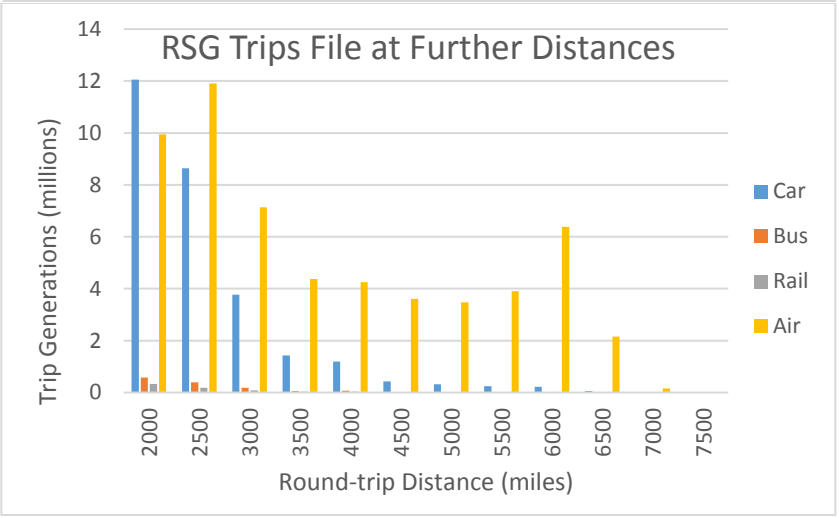


Figure 9.2 Distribution of rJourney trips for all trip types for a. all distances (shown logarithmically), and b. further distances

The rJourney set also provides a skim file that includes mode statistics of traveling between most possible pairs of NUMAs. These include estimated travel time by car or air, access and egress times, traveling toll or cost, and other factors that would influence a traveler's choice of transportation mode. Summary statistics of the skim files are shown in Table 9.1. Corresponding to these are mode choice and destination choice coefficients. In these coefficients, VOTT for car drivers is $12/hour (in 2010 dollars). These skims and data are used in this research for evaluating the effects of adding a new AV mode.

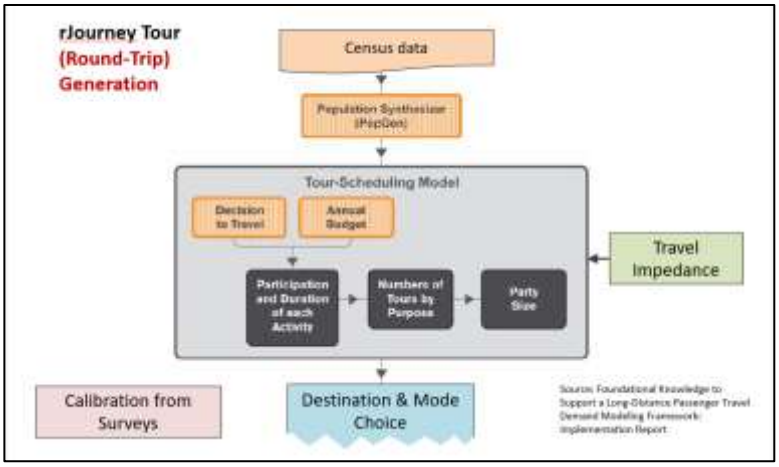
Table 9.1 Summary statistics for the rJourney skim file

Variable	Mean	Std	Min	Max
Air File, *N* = 18,424,925				
Time	218.73	97.94	25.00	812.00
Transfers	82.37	50.19	0.00	200.00
FreqDirect	10.58	24.07	0.00	339.00
Freq1Stop	145.41	258.47	0.00	2,286.00
Freq2Stop	348.81	932.69	0.00	10,968.00
OnTime	88.79	4.00	0.00	100.00
EconomyFare	519.13	327.69	0.00	50,776.00
BusinessFare	1,199.61	955.59	0.00	152,328.00
AccessDistance	38.15	25.99	0.00	101.00
EgressDistance	38.22	26.34	0.00	102.00
Rail File, *N* = 8,010,759				
Time	2,167.24	1,269.59	4.00	6,270.00
Transfers	134.57	111.05	0.00	800.00
Frequency	7.77	10.41	3.00	93.00
EconomyFare	131.75	39.51	9.00	181.00
BusinessFare	340.56	132.40	18.00	605.00
AccessDistance	22.82	14.65	0.00	50.00
EgressDistance	22.16	15.14	0.00	50.00
Road File, *N* = 19,727,179				
CarTime	1,161.72	668.14	1.00	3,613.00
Distance	1,185.41	706.49	1.00	3,582.00
Toll	67.15	137.85	0.00	1,344.00
BusTime	1,313.12	1,249.89	0.00	5,617.00
BusFare	94.71	85.72	0.00	383.00

9.2. Model Specification

Implemented model leverages a subset of rJourney data and models and also uses pre-existing parameters as a means to quickly characterize the trip distributions for each mode, while leaving the opportunity to add a new mode such as AVs. The subset of data and coefficients were used to closely reproduce the mode choice results, and then a new AV mode was added. For this analysis, the model was set up as a nested logit model, where mode choice was a nest within an overarching destination choice model. Figure 9.3 shows how the rJourney model operates.

a.



b.

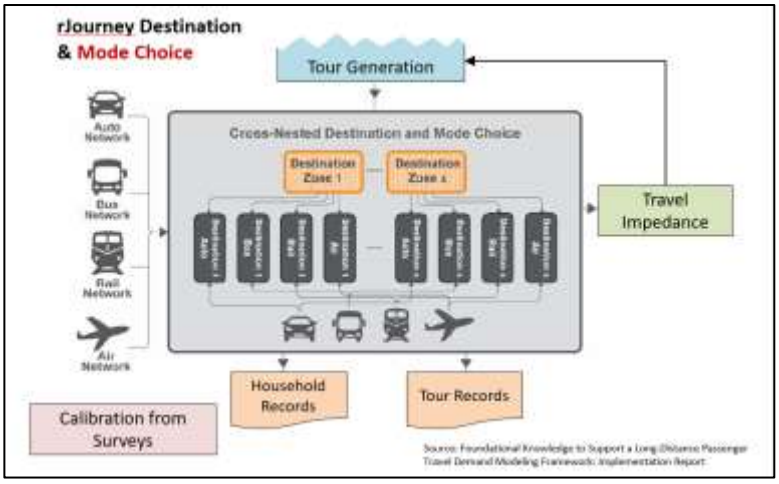
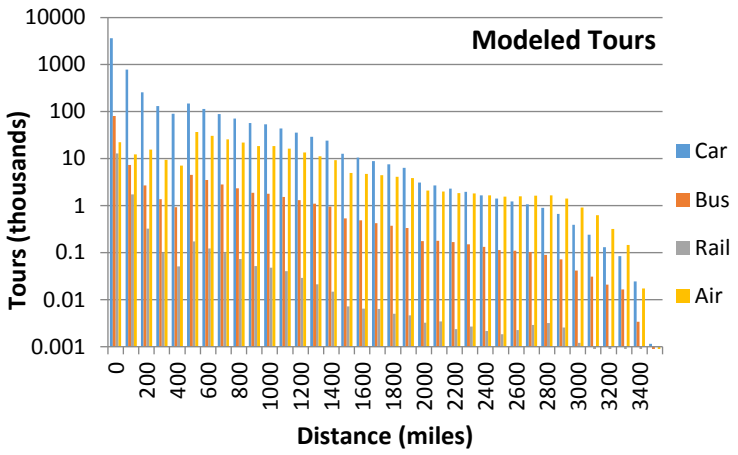


Figure 9.3 rJourney model summary

For finding mode choice from each origin to each destination, parameters include direct costs (VOT, tolls, and fares), NUMA household density, service frequency, transfer frequency, and rail station/airport access and egress penalty. For simplicity, unavailable data and insignificant parameters in mode choice (e.g., with low T-stats) are not represented in utility functions as they are in the rJourney model, including household size, party size, and number of nights staying. Party size is currently assumed to be 1.

In fact, the model subset does not produce an exact replication of the rJourney tours data set. Furthermore, the attempted addition of the AV mode inherently lacks supporting data, already necessitating the use of a subset of existing parameters. Although model subset results show a similar distribution to that of the rJourney tours data set, air travel in particular was underrepresented, showing a correlation of 0.71 overall, shown in Figure 9.4. To establish a closer representation, a strategy for adjusting (or "pivoting") the results off of the rJourney tours data set is introduced.

a.



b.

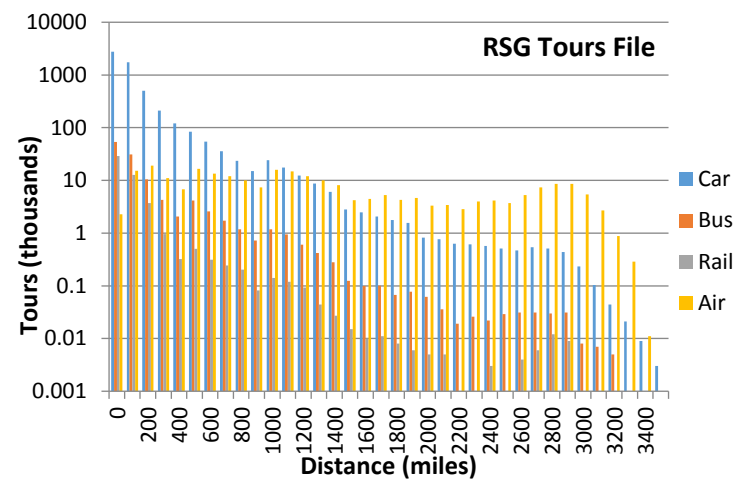


Figure 9.4 Air travel comparison between model and rJourney data

While future work related to this research will continue to improve upon the rJourney model usage, the preliminary exercise discussed in this chapter illustrates the kinds of analyses that are possible with such a model. These are the mode choice utilities, functions of NUMA zone, destination NUMA zone, and trip purpose. Refer to Table 40 in the long-distance passenger travel demand modeling framework report (FHWA, 2015) that contains the coefficient values and T-stats for each of the trip purposes identified by coefficient subscript number. In this analysis, the data series pertaining to cost of traditional vehicle operation was drawn using the estimated value of $0.17/mile. Because this model focuses on mode choice at the time of travel, the ownership cost is not incorporated as in (AAA, 2015). While this serves as a rough estimate, it would be possible with further research to better quantify operation costs as a function of each trip-maker's annual driving distance. The results of the initial analysis shall inform how this function can be evaluated in the future. The rJourney data includes 285,579 NUMA pairs that lack car mode statistics. These NUMA pairs and corresponding trips are omitted from this analysis because of lack of car-distance data, which is needed in estimating the distance of all modes of travel.

The introduction of AVs into the model presents challenges in implementation, mainly in that the rJourney models and results obviously do not consider the presence of AVs, and little data currently exist to specifically justify model parameters. For AVs to be considered as a new modal alternative, existing data and coefficients are leveraged to arrive at a "best-guess" parameter set. In initially designing how the new modal alternative is integrated, the following assumptions are made: a) a future time is modeled where AVs cost on average $0.20 per mile to operate; b) the $6.00 VOT to the occupant is half of that of traditional car; and c) all other parameters are that of traditional cars.

Probability splits for mode choice given each origin, destination, and purpose are found by the utility of choosing each mode. The destination choice portion of the model incorporates the logsum of the mode choice utility functions along with indicators pertaining to distance ranges, as well as household and employment counts that come from the NUMA zone data set. Again, for simplicity as well as lack of access to data, parameters that are not strongly influential in mode choice and destination choice were omitted. However, as noted later, preliminary results are helpful in identifying investigations of the model in future work. As an observation, the rJourney model does not include gross domestic product per NUMA zone, which could possibly be helpful for future efforts in better representing destination attractiveness. The destination choice model uses coefficients drawn from (Federal Highway Administration, 2015) Table 39. Future research efforts will evaluate how more of the rJourney destination-choice model can be leveraged for arriving at an improved representation of attractiveness.

From this, joint mode/destination choice probabilities are found by combining the mode choice and destination choice conditional probabilities for each origin-destination (OD) pair. The last step is to use the joint probabilities to distribute trips that are generated from each origin across all modes and destinations. For this analysis, the number of generated trips is obtained from the rJourney tours data that was simulated from generated households across the United States. Because the idea is to study how mode choice and destination choice changes with the introduction of AVs, the mode choices represented in the rJourney tours dataset are ignored to allow the same number of generated tours to be redistributed according to the post-AV introduction model. The model implementation procedure is shown in Figure 9.5.

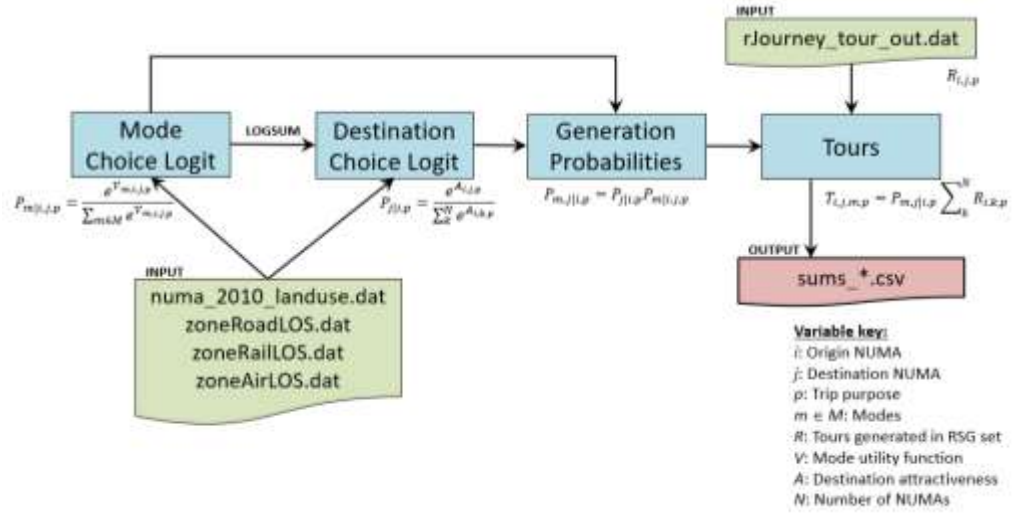


Figure 9.5 Model implementation

Computation of this model (shown in Figure 9.6) can be classified as a big data problem. In representing the expanded 1.17 billion trips, 38 million rJourney trip records over 2 million NUMA pairs constitute 4 GB of data, and files representing the intermediate and final computational results for all trip purposes amount to gigabytes of additional storage requirements. The Python Pandas library is used to perform the computations along with HDF5 file format support. With a number of considerations made for vectorized matrix operations, the entire set takes on the order of 30 minutes to run on a modern, high-end computer. Operations that read and write files from flash storage account for over half of the run time.

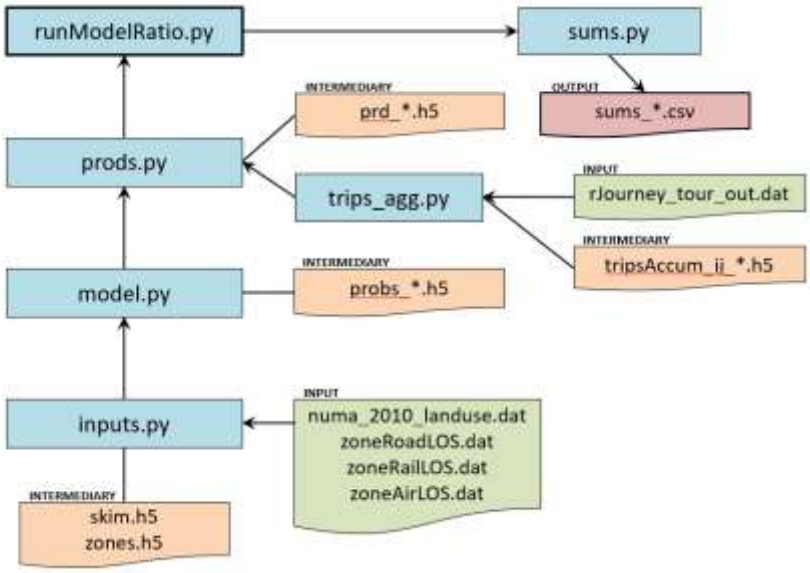
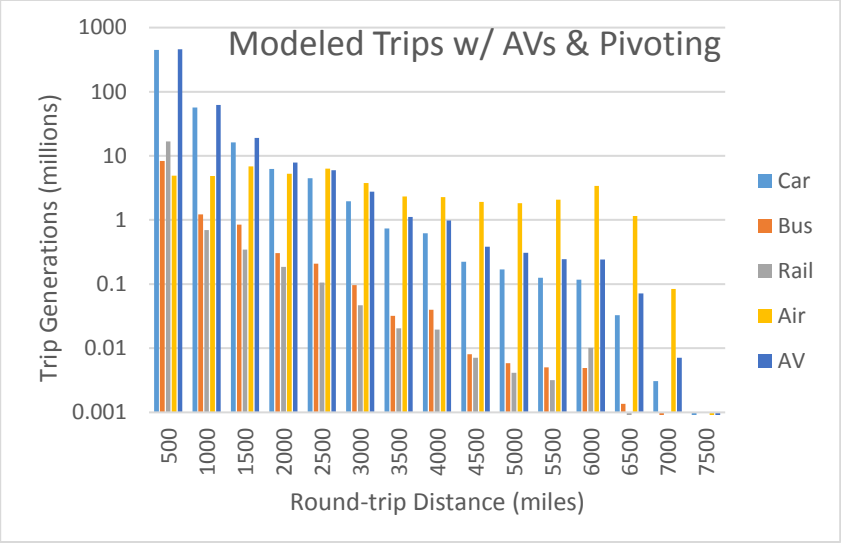


Figure 9.6 Computation procedure

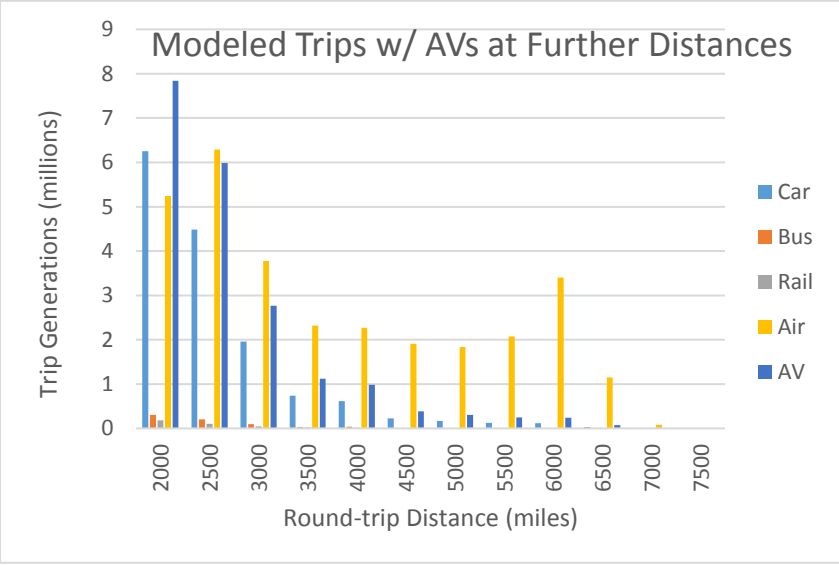
9.3. Impacts of CAV on Model Choice and Destination Choice

9.3.1. AV Trip Distribution

Figure 9.7 shows the resulting number of trips after the AV mode is added to the initial model as described in the methodology. This can be compared with the tours data set distribution in Figure 9.2. A notable observation is that the distribution of AV trips tracks the distribution of traditional vehicles with an increase in mode share at further distances. This can be attributed to high correlation of several parameters that are represented in the traditional vehicles. The key differences with AVs are the increase in operating cost, and reduced VOT driving. With similarity in parameters, this mode split is influenced by the independence from irrelevant alternatives (IIA) property (or, noted many times in the literature as the "red bus/blue bus paradox") inherent in multinomial logit models. This property causes highly correlated inputs to be treated as independent, which creates an artificial demand that may not necessarily happen in reality. The high degree of correlation and presence of IIA can best be addressed by creating a nest (e.g., "personal vehicles") that contains both of the AV and car results.



(a)



(b)

Figure 9.7 Number of trips from the mode choice/destination choice analysis, all purposes, at a. all distances (shown logarithmically), and b. further distances

There are two notable outcomes that offer insight on the possible effects of AV introduction to the market, as well as a shift in destination choice. First, results show that the introduction of AVs deeply cuts into the number of trips that had formerly been air trips. See Table 9.2 for results in terms of shorter and longer trips (e.g., < 500 miles (805 km) versus ≥ 500 miles). For shorter trips, the number is dramatically reduced, whereas for longer trips, the number is far less affected by AV introduction. As largely influenced by the Car coefficient for one-way distance greater than 500 miles as well as travel time, trips over 500 miles in length are penalized because of the negative "captivity factor" of remaining in a car for a long period of time possibly over several days. It is assumed in this model that this disutility would be similar for AVs as it would be for traditional cars. Note that in Table 9.2, "Car+AV" is shown as a means to represent respective totals of personally owned vehicles.

Table 9.2 Trip mode choice impact of AV introduction for all trip purposes

TOURS	AV Market Penetration	Car+AV < 500 mi. round trip	Car+AV ≥ 500 mi. round trip	Air < 500 mi. round trip	Air ≥ 500 mi. round trip
Before AV	0%	860.5 M	168.8 M	9.3 M	79.5 M
After AV	51%	906.9 M	189.0 M	4.9 M	42.0 M
% change	-	105.4%	112.0%	52.9%	52.8%

VEHICLE-MILES	Car+AV < 500 mi. round trip	Car+AV ≥ 500 mi. round trip	Car+AV Total	Air < 500 mi. round trip	Air ≥ 500 mi. round trip	Air Total
Before AV	400.8 B	821.0 B	1,221 B	6.4 B	437.9 B	444.3 B
After AV	425.2 B	913.7 B	1,339 B	3.4 B	232.3 B	235.7 B
% change	106.1%	111.3%	109.6%	52.9%	53.0%	53.0%

Second, among traditional cars and new AVs, more destinations are chosen after introduction of AVs that are further in distance from origins. However, if all modes are considered, the trend is reversed, possibly because of the severe reduction of air trips that dominate the longer-distance trips. Table 9.3 shows a change in distribution across overall trip distances. For both pre- and post-AV introduction the model uses the same number of trip generations per NUMA per trip purpose. The significant decrease of air travel may be a consequence of the aforementioned IIA property. In addition to treating cars and AVs as a single nest, further work on characterizing VOTT and operating cost, as well as specifying additional factors in the destination-choice portion of the model, may have the outcome of evolving how trip distances are biased among closer and further long-distance trips.

Table 9.3 Trip distance shift for all trip purposes

TOURS	Tours < 500 mi. one way	VMT for tours < 500 mi.	Tours ≥ 500 mi. one way	VMT for tours ≥ 500 mi.
Before AV	914.1 M	422.4 B	256.1 M	1,294 B
After AV	937.0 M	437.1 B	235.2 M	1,165 B
% change	102.5%	103.5%	91.8%	90.0%

9.3.2. Market Penetration of AV

The degree that AVs penetrate the market varies according to trip distance. Figure 9.8 shows market penetration both for AVs among the personal vehicle modes (e.g., car and AV), and also AVs among all mode choices. With respect to personal vehicles, the market penetration increases as distance increases because of the significance of lower VOTT. However, air travel continues to be preferred for longer distances and results in the AV mode share diminishing at further distances. The deviation in penetration for the 7000-mile bin is likely a result of fewer trip samples for that furthest distance.

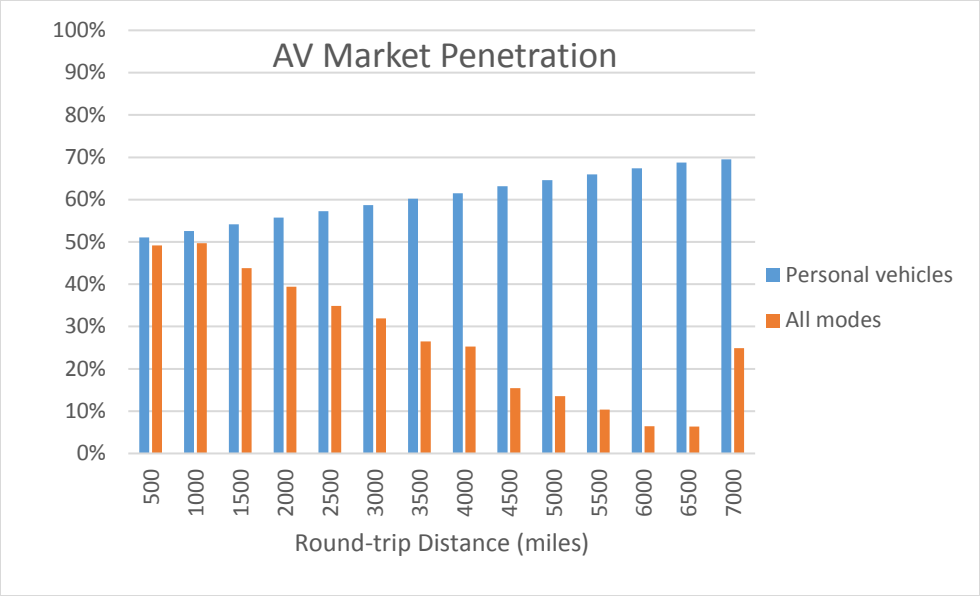


Figure 9.8 Penetration of AVs among personal vehicles (car+AV) and all modes

9.3.3. Passenger Airline Sales

Given that large-scale introduction of AVs has not yet happened and that no data can be collected directly from AV usage today, a model such as this rJourney subset with AVs added as a new mode can be helpful in roughly estimating market effects that could result from the widespread introduction of AVs. One question that can be addressed with this model is how much revenue the airline industry can possibly lose due to more travelers choosing AVs over air travel. The rJourney data set gives airfare estimates in USD for all NUMA pairs that have suitable access to airports served by commercial passenger carriers. Table 9.4 shows estimated airline sales before and after the addition of AVs for all modeled trips. Note that because these are based upon cost to the traveler, these sales figures include airport taxes.

Table 9.4 Passenger airline sales for all trip purposes

REVENUE	Tours < 500 mi. round trip	Tours ≥ 500 mi. round trip	Total revenue
Before AV	$16.0 B	$159.1 B	$175.1 B
After AV	$8.4 B	$83.9 B	$92.3 B
% change	52.7%	52.7%	52.7%

In this result, the percent changes between sales between shorter and longer long-distance trips are similar. This is counterintuitive because of the idea that AVs should have a more significant attractiveness for shorter trips and thus cut more into the shorter distance market. It may be here that the model is dominated by the IIA property in adding AVs as a separate mode rather than as a car+AV "personal vehicle" nest. Additionally, with refinements in the mode choice and destination choice models the split may improve in accuracy.

9.3.4. AV Parameter Sensitivity

As mentioned earlier, the parameters and assumptions given to AVs are largely unknown and must be estimated. Two notable parameters include cost of operating the vehicle, as well as personal VOTT. (Another parameter that is relevant but not yet analyzed includes a more pronounced representation of the 500-mile captivity factor, which may be different for car drivers than it is for AV passengers.) A thorough analysis should offer a set of scenarios that span a range of expected operational costs and personal VOTT, given the targeted years, expected AV market penetration, and socioeconomic classes of trip-makers that are being analyzed.

To further understand the sensitivity of these variables on the resulting mode split and destination choice, six new scenarios are created for the "leisure" trip purpose. Scenarios are presented in Table 9.5.

Table 9.5 Scenarios of sensitivity analysis

Scenario	Oper. Cost ($/mile)	VOTT ($/hr)	Notes
A	$0.20	$6.00	Base case
B	$0.10	$6.00	Operating cost is cheaper
C	$0.50	$6.00	Operating cost is more expensive
D	$0.20	$3.00	VOTT is decreased
E	$0.20	$9.00	VOTT is increased
F	$1.65	$6.00	AVs are modeled as shared vehicles

Recall that dollar amounts are expressed in year 2010 dollars. The scenario of AVs having the same operating cost and VOTT of cars has been omitted because there would be no distinction between the car and AV modes. Scenario F in particular has been included as a hypothetical scenario to roughly model all AVs on the roadways as shared autonomous vehicles (SAVs). With SAVs, passengers do not own their vehicles, but rather pay per mile for travel in a borrowed vehicle that others can use for other trips, in this case $1.65 per mile. As more data emerges, an improved model would likely offer SAVs as a mode choice that is separate from personally-owned AVs. Table 9.6 shows the results of each of these scenarios.

Table 9.6 Trip generations with varied AV parameters, for "leisure" trip purpose

Mode	Dist.	Scenario	A	B	C	D	E	F
Car+AV	Trips < 500 mi.	Before AV	253.5 M	253.5 M	253.5 M	253.5 M	253.5 M	253.5 M
		After AV	271.5 M	267.2 M	279.4 M	268.1 M	274.3 M	280.3 M
		% change	107.1%	105.4%	110.2%	105.7%	108.2%	110.6%
	Trips ≥ 500 mi.	Before AV	55.7 M	55.7 M	55.7 M	55.7 M	55.7 M	55.7 M
		After AV	63.4 M	65.7 M	57.8 M	65.3 M	61.7 M	46.7 M
		% change	113.8%	118.0%	103.7%	117.2%	110.7%	83.9%
Air	Trips < 500 mi.	Before AV	2.30 M	2.30 M	2.30 M	2.30 M	2.30 M	2.30 M
		After AV	1.23 M	1.20 M	1.28 M	1.21 M	1.24 M	1.40 M
		% change	53.2%	52.1%	55.7%	52.3%	54.0%	60.9%
	Trips ≥ 500 mi.	Before AV	18.11 M	18.11 M	18.11 M	18.11 M	18.11 M	18.11 M
		After AV	9.65 M	9.43 M	10.16 M	9.47 M	9.81 M	11.17 M
		% change	53.3%	52.1%	56.1%	52.3%	54.2%	61.7%

In observing Scenarios B, A, and C in order of increasing operation cost, it can be seen that closer trip generations increase, and longer trips decrease, possibly because of the significance of operating cost on longer trips. Meanwhile, the cut into the air market decreases as the operation cost increases. In the rough SAV Scenario F, the results coincide with a similar trend, where longer distance trips are more significantly curtailed. For Scenarios D, A, and E in order of increasing VOTT, a similar phenomenon occurs. The reduction of air trips decreases as VOTT increases.

In all cases, the variations that are evaluated do not show an extreme difference in outcomes. In considering travelers' expenses and VOTT, it is possible to reason that the results should be more distinct. Two factors may be dominating the models as these inputs are varied. First, the addition of the AV mode as an independent choice may be an inaccurate model structure that is highly correlated and represented too significantly in the results. As mentioned earlier, it may be more appropriate to treat cars and AVs as a "personal vehicle" nest and estimate the correlation that is to be expected among the mode choices of hypothetical travelers. Second, the representation of AVs in the model is somewhat indistinct from cars, as few parameters exist to offer better differentiation. The addition of new parameters to the car and AV modes can help with this and reduce the correlation between the two modes.

9.3.5. VMT Change

Based on the mode choice and destination choice for all NUMA zones, traffic assignment is performed for OD matrix before and after AV scenario over the U.S. network in TransCAD. Total VMT data is collected from the simulation. Before AV scenario, we have 14.4 billion VMT of car, while the total VMT of Car+AV increase to 15.8 billion. The increase of VMT for all trips happened within Texas region (Texas as origin, destination, or through Texas) is about 9.78%.

Table 9.7 shows the VMT change results of several states. VMT change across U.S. for each state showed an increment arranging from 8.09% in North Carolina to 13.52% in Kansas. VMT in border states like Texas and Illinois shows an increase approaching 10%. Coastal states like

California and Virginia have about 9% increase in VMT, while inlands states like Utah and Wyoming witness nearly 13% increase in VMT.

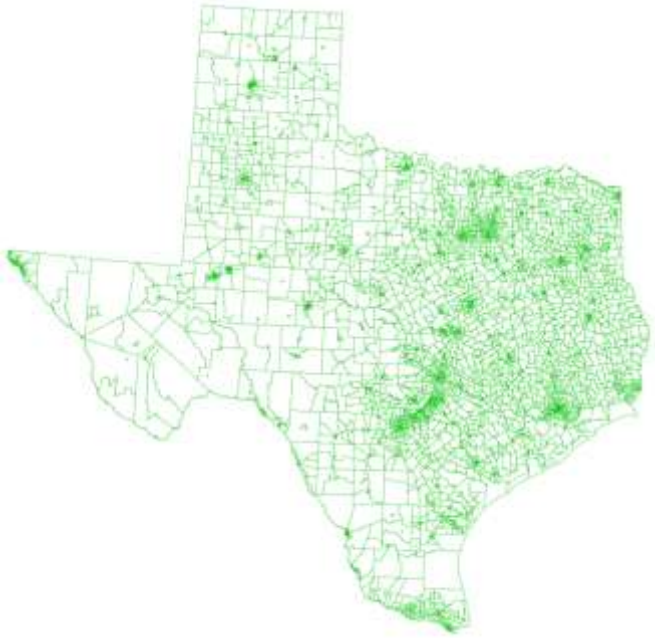
Table 9.7 VMT change in state for interstate long-distance trip

States	Texas	New York	Florida	California	Pennsylvania	Illinois
Before	14.44 B	8.05 B	9.55 B	12.74 B	11.81 B	11.14 B
After	15.85 B	8.73 B	10.50 B	13.91 B	12.94 B	12.21 B
Increase	9.78%	8.39%	9.94%	9.18%	9.55%	9.57%
States	**Washington**	**Virginia**	**Utah**	**Kansas**	**Wyoming**	**North Carolina**
Before	2.63 B	10.72 B	1.62 B	3.14 B	3.56 B	6.89 B
After	2.87 B	11.67 B	1.83 B	3.57 B	4.01 B	7.45 B
Increase	9.06%	8.82%	12.9%	13.5%	12.8%	8.09%

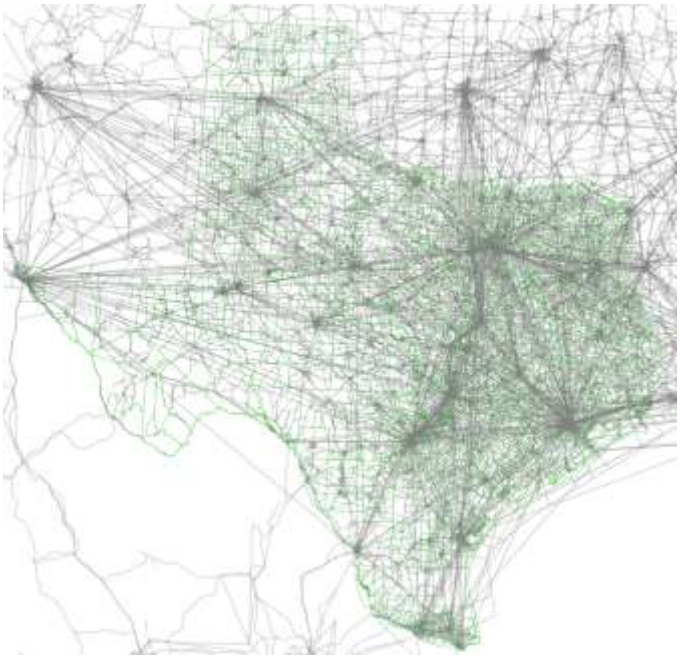
9.4. Summary of Anticipated Long-Distance Trips

9.4.1. Data Set

The Statewide Analysis Model (SAM) provides 4667 TAZs across the state of Texas, as shown in Figure 9.9(a). Figure 9.9(b) shows the highway, railway, and airline networks, which contain 200,445 links and 168,507 nodes. The links and nodes cover the entirety of North America, though the highest detail is in Texas. This study focuses on travel within Texas, while some trips would take routes outsides Texas and come back.



a. SAM TAZs



b. SAM Networks

Figure 9.9 SAM's geographic data

9.4.2. Methodology of Four-step Model

The four-step travel demand modelling process is used here to model traffic patterns across the entire state of Texas: trip generation, trip distribution, mode choice, and traffic assignment. The traditional trip distribution procedure is replaced in this study by a destination choice model, and a production-attraction matrix (PA) to origin-destination matrix (OD) procedure is conducted to

convert the PA matrix to OD matrix. The model is divided into four times of day, and AM peak results are reported here.

9.4.2.1. Trip Generation

Trip generation data is obtained from SAM generation results of its year 2040 scenario. Passenger trip generation of SAM is based primarily on the 2009 National Household Transportation Survey (NHTS). Of SAM's seven trip purposes, intra-city trips include home-based work, home-based other, home-based school, non-home-based other and non-home-based visitor. Intercity trips include infrequent long distance business trips and infrequent long distance other trips. This work assumes that all types of trips produced and attracted conform to the destination choice from random utility theory. However, the trip generation ignores external trips, either coming from or heading outside Texas. Freight trips are also ignored here, while freight trips in fact account for a great proportion of Texas traffic.

9.4.2.2. Trip Distribution

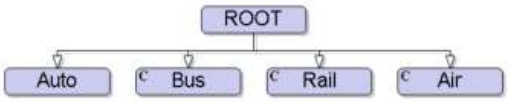
The general trip distribution procedure is replaced by a destination choice model in TransCAD for this study. The attraction of each destination depends on the mode accessibility and the destination's attraction factors. Therefore, the destination choice is assumed to be dependent on the population of each TAZ and the mode choice logsum. Based on parameter assumptions from Yong and Kockelman 2017, the parameter assumption for this work is shown in Table 9.8:

Table 9.8 Parameter assumption for destination choice model

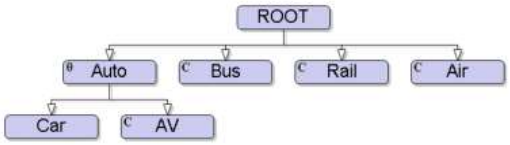
	Mode Choice Logsum	Log of Population
Parameter	0.855	1

9.4.2.3. Mode Choice Model

Four modes are considered in the base case scenario: auto, bus, rail, and air. Auto represents all modes utilizing a conventional vehicle (including driving alone, taxi, and transportation network company, e.g. Uber). SAM model considered 20 or more combination of mode choices, while this study considers only four basic modes for the purpose of easily accommodating the autonomous vehicle mode. Although bus and auto modes usually share the same network, the congestion caused between bus and auto are ignored in this study, since the SAM network employed here models the bus mode in the transit network instead of the highway network. Since these four modes share different networks, the base case scenario only assigns conventional automobiles on the highway and urban network. Fare and in-vehicle travel time of bus, rail, and air are obtained from the SAM model. Rail commuting time and fare are the average of all available rail modes in SAM model for a certain OD pair, including urban rail, intercity rail, and high-speed rail. When the AV mode is added, AV and conventional vehicle are nested under the auto mode, which itself is in the same level as bus, rail, and air, as shown in Figure 9.10(b).



a. Mode choice structure without AV



b. Mode choice structure with AV

Figure 9.10 Mode choice structure before and after AV

The mode choice parameters for base case scenario and AV scenario are shown in Table 9.9:

Table 9.9 Mode choice model parameters

Base Case	Auto		Bus	Rail	Air
Constant	-		-2.8	-2.8	-2.8
Operating Cost	-0.072		-0.14	-0.14	-0.14
In-vehicle Time	-0.019		-0.019	-0.019	-0.019
With AV case	Car	AV	Bus	Rail	Air
Nest Coefficient	0.7		-	-	-
Constant	-	-0.05	-2.8	-2.8	-2.8
Operating Cost	-0.072	-0.072	-0.14	-0.14	-0.14
In-vehicle Time	-0.019	-0.0095	-0.019	-0.019	-0.019
VOTT	15.83	7.92	8.14	8.14	8.14

9.4.2.4. Time of Day and PA to OD

The time-of-day intervals are shown in Table 9.10, and Table 9.11 is the PA to OD departure and return table. The PA to OD departure and return table generally gives the trip distribution over a day, for both departure and return trips.

Table 9.10 Time-of-day intervals

Time of Day	Time Interval
AM Peak (AM)	6:00 am to 9:00 am (3 Hours)
Mid-Day (MD)	9:00 am to 4:00 pm (7 hours)
PM Peak (PM)	4:00 pm to 7:00 pm (3 hours)
Hours Night (NT)	7:00 pm to 6:00 am (11 hours)

Table 9.11 PA to OD departure and return table

Hour	Departure	Return
0	0.35	0.35
1	0.1	0.1
2	0.4	0.4
3	0.05	0.05
4	0.05	0.05
5	0.5	0.5
6	1.6	1.6
7	4.45	4.45
8	2.05	2.05
9	1.6	1.6
10	1.95	1.95
11	2.05	2.05
12	2.6	2.6
13	2.4	2.4
14	2.45	2.45
15	3.35	3.35
16	4.65	4.65
17	4.25	4.25
18	3.2	3.2
19	3.95	3.95
20	2.95	2.95
21	2.4	2.4
22	1.6	1.6
23	1.05	1.05

9.4.2.5. Traffic Assignment and Feedback Loop

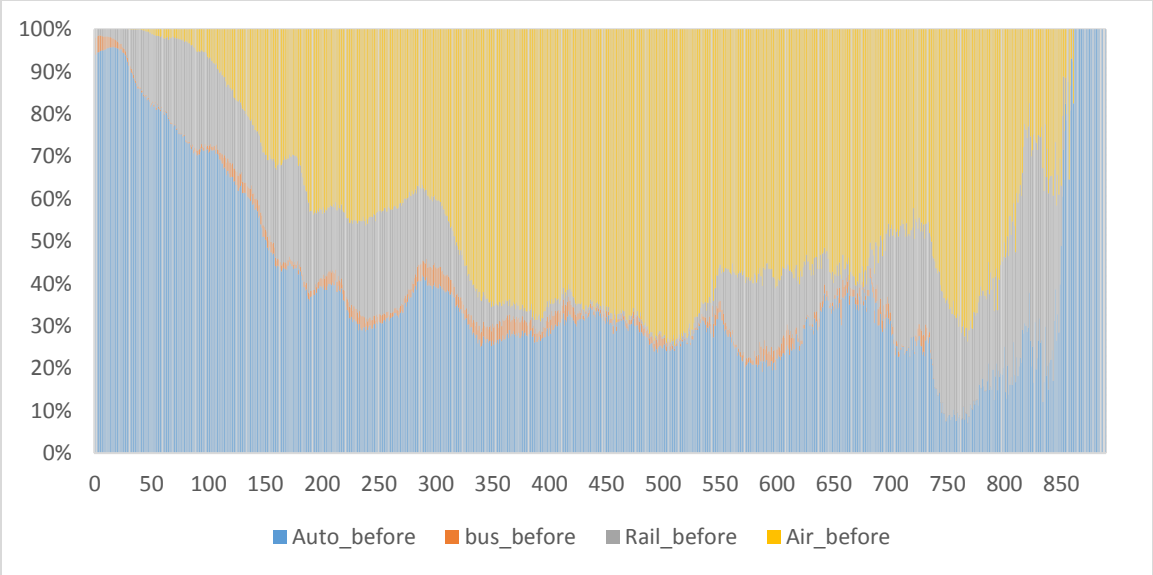
Feedback loops are performed, using the congestion travel time to provide feedback to each subsequent iteration. Different congestion time will lead to different destination choices and mode choices, thereby altering traffic assignments. The static traffic assignment is conducted for the base case, where only conventional automobiles are assigned to the network. Multi-modal, multi-class assignment is conducted for the AV case, where both conventional vehicles and AVs are assigned to the network. The feedback loop is set to perform 10 iterations, with the stop criteria of a relative gap below 10^{-4}.

9.4.3. Results

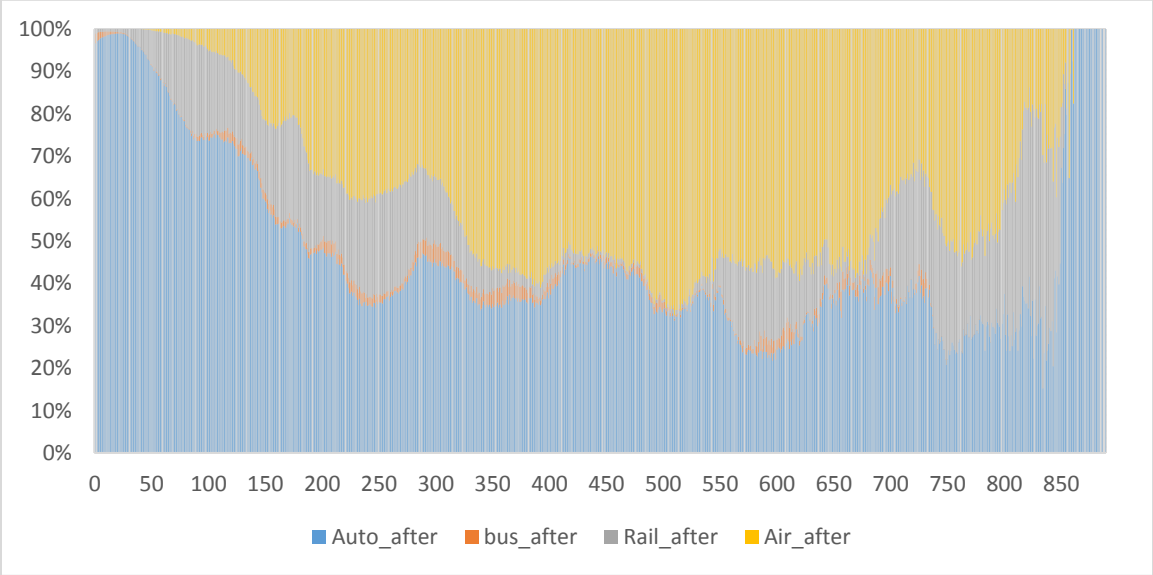
9.4.3.1. Mode Share

Figure 9.11 shows the mode share for four modes before and after AV implementation. The auto mode is the sum of conventional automobile trips and those via autonomous vehicles. Rail mode and bus mode both remain relatively stable before and after AV introduction. After AV

introduction, auto mode share increases for distances between 200 miles and 800 miles, shifted mostly from air travel. This is likely due to the lower VOTT for traveling by AV, compared to other modes. However, there are still more trips taken by air between 500 miles and 650 miles in length. Trips of greater than 850 miles are only traveled by auto, possibly because the spatial location of airports: the longest distances between commercial airports within Texas are approximately 750 miles. It is also possible that trips by air would be costlier due to connecting time.



a. Before AV implementation



b. After AV implementation

Figure 9.11 Mode share against trip distance

9.4.3.2. Trip Distribution Analysis

Figure 9.12 shows the trip distribution of thousand trips of auto before and after AV introduction. After AVs are introduced, trips of distance greater than 60 miles increase a little, while the shortest trips nearly double. The peak remains around 15 miles trip distance both before and after AV introduction. However, as Figure 9.13 shows, when comparing conventional vehicle travel ("Car" mode) with that of AVs, the peak in trip frequency for conventional automobiles is lower, occurring at 13 miles, compared to 18 miles for AV trips. AVs dominate travel among auto modes, especially for distance greater than 10 miles and less than 500 miles. This shows that people are more willing to choose AVs over conventional automobiles for long trips, where the VOTT plays a larger role.

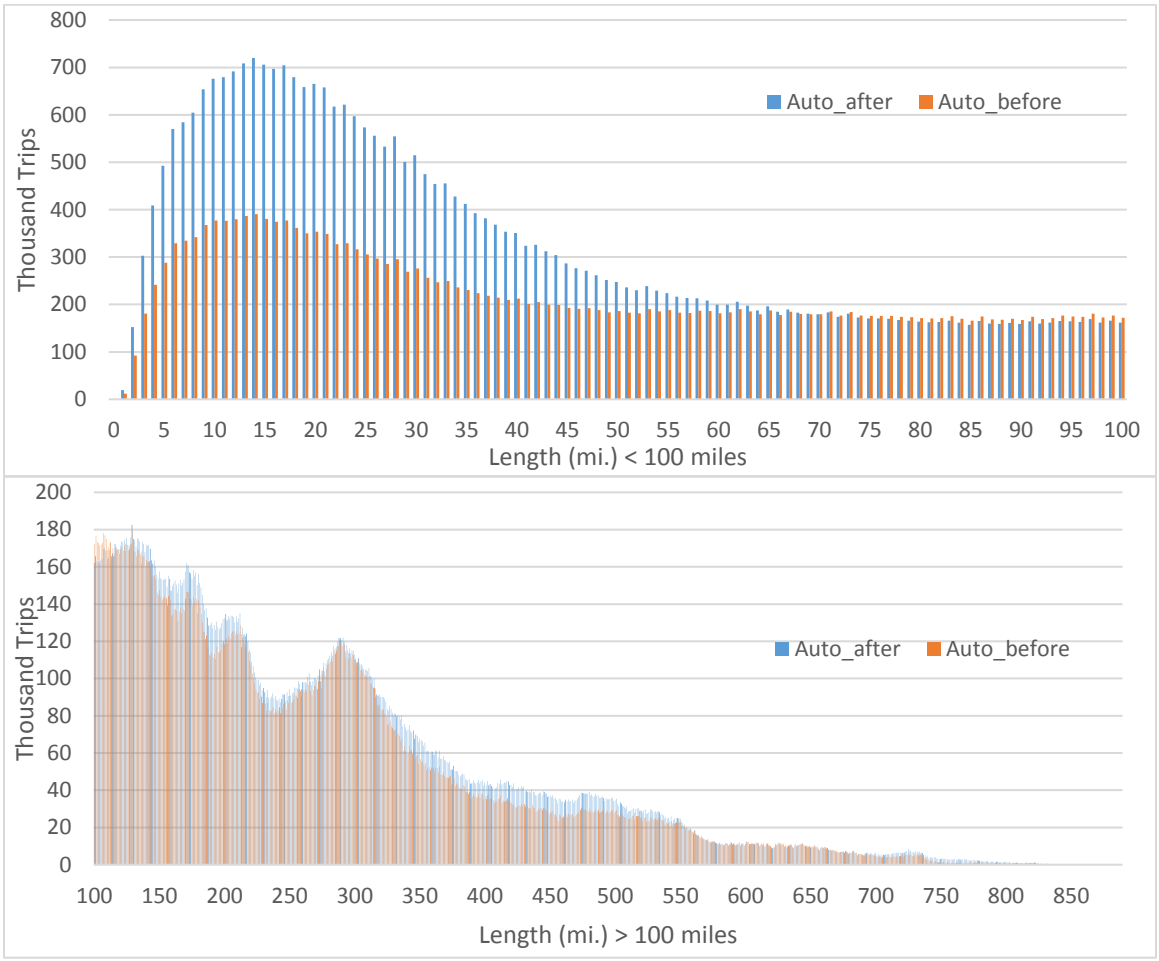


Figure 9.12 Trip distribution of Car mode before and after AV

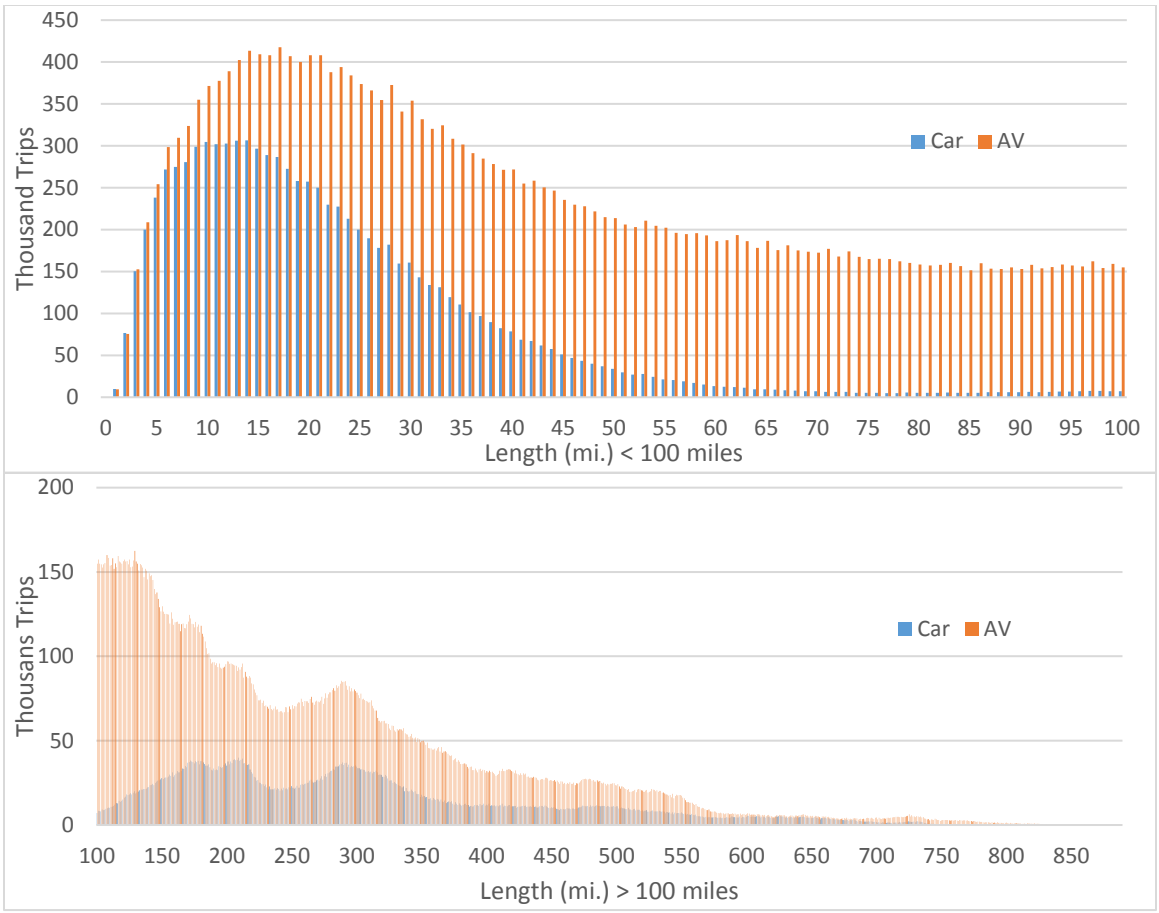


Figure 9.13 Distribution of AV and Car modes

9.4.3.3. VMT Change

Table 9.12 shows the VMT change for all modes after introducing AVs. Air mode has more VMT than auto mode before AV introduction, but they converge to similar VMT after AV introduction. Rail, bus, and air modes show a decrease in VMT, with rail travel decreasing by 19.3%, air travel by 11.7% and bus VMT shrinking by 1.8%. Overall auto modes show a 12.4% increase in VMT after AV introduction.

Table 9.12 VMT change of modes before and after AV

VMT (billion vehicle-mile)	Auto	Rail	Bus	Air
Before	10.97	0.69	3.60	13.95
After	12.33	0.56	3.54	12.31
Change	12.4%	-19.3%	-1.8%	-11.7%

Table 9.13 details the VMT changes in cities of interest, which the SAM model identifies as the TxDOT districts. The Dallas and Fort Worth area show an increase in total VMT of 20%, while San Antonio, Houston, and Austin each show a small decrease. More cities located in border areas of Texas become more attractive as they are probably more accessible through the AV mode. AV

introduction increases total VMT across Texas, as these districts show a total VMT increase of 3.2%.

Table 9.13 VMT change in million vehicle-miles traveled in districts

Area	VMT before	VMT after	Change	Area	VMT before	VMT after	Change
Abilene	72.59	85.88	18.3%	Laredo	91.98	87.76	-4.6%
Amarillo	64.64	73.21	13.2%	Lubbock	86.83	103.27	18.9%
Atlanta	45.92	59.01	28.5%	Lufkin	97.34	92.61	-4.9%
Austin	177.04	163.28	-7.8%	Odessa	73.88	75.62	2.3%
Beaumont	98.50	90.05	-8.6%	Paris	67.83	81.61	20.3%
Brownwood	88.96	90.51	1.7%	Pharr	41.20	41.18	-0.1%
Bryan	137.89	125.57	-8.9%	San Angelo	113.91	110.17	-3.3%
Childress	39.44	46.96	19.1%	San Antonio	147.41	142.38	-3.4%
Corpus Christi	66.87	63.77	-4.6%	Tyler	85.07	100.01	17.6%
Dallas	111.35	133.67	20.0%	Waco	134.24	120.17	-10.5%
El Paso	26.31	42.81	62.7%	Wichita Falls	58.22	73.83	26.8%
Fort Worth	92.62	109.15	17.8%	Yoakum	121.53	103.37	-14.9%
Houston	132.16	130.62	-1.2%	Total	2273.73	2346.43	3.2%

9.4.4. Discussion

This work utilizes TransCAD software and SAM data to perform four-step model and provide a nested logit model to accommodate AVs in destination choice and mode choice model. Auto mode (conventional vehicle plus AV) trips increase for distance between 200 miles and 800 miles, largely shifted from air trips. People would also like to go further distances after AV is introduced, with peak of distance frequency as 18 miles rather than 13 miles of conventional vehicles. Results also show the VMT change for each region in Texas. Predictions of reduced air trips and increasing demand for highway infrastructure should provide insights in preparations for the advent of autonomous vehicles for counties and regions in Texas. However, a restriction of this work currently is its focus on the changes of internal passenger trips within Texas, although trips going across Texas border and coming back are allowed. The ignorance of external trips and freight trips would largely underestimate the AV trips, so AV trips are expected to have greater impact than discussed in current results. Further work would be look at the external passenger trips as well as the freight trips, and also the sensitivity analysis for various parameter assumptions for AVs, so to achieve a comprehensive of AV prediction on long-distance trips.

9.4.5. Summary

This chapter questions of how the rJourney model and extensions to it can be applied to other regions beyond the U.S. is relevant, especially as several other nations are amenable to the introduction of AVs onto their roadways. Data similar to the rJourney skims and land use records, including travel distances, travel time, airfare, transfers, housing density, and employment, can be readily collected for other regions. However, the application of the rJourney model in other nations

must be taken with a significant amount of care and additional work, as much of rJourney portrays traveler behavior that is specific to the United States. Indeed, the nation-scale trip generation is heavily dependent upon extrapolations from long-distance surveys that were issued in only 5 of the 48 lower continental states. The sparseness of the survey coverage is identified as an area for improvement, as there is a strong recommendation for a long-distance passenger travel survey to be issued for the entire United States (Federal Highway Administration, 2015). Regardless, many of the survey results are inevitably influenced by socioeconomic status, land use patterns within household surroundings, and infrastructure capacity—aspects that are specific to the respective region and nation.

For an applicable model to be created for another region, it is advisable for one to acquire the region-specific data, both the quantitative regional facts as well as a sampling of qualitative results through surveying and research. Then, a process may be followed that is similar to the creation of rJourney's underlying logit models and the generation of households and trips. The coefficients of the mode and destination choice models, as well as the tours records, would then properly reflect the region being examined.

This chapter also addresses how distance traveled in Texas relates to the introduction of AVs. The area of Texas smaller than the rest of the U.S., which provides the possibility to obtain the detailed change of VMT for each county in Texas. However, a current restriction is its focus on the changes of internal passenger trips within Texas, although trips going across Texas border and coming back are allowed. The ignorance of external trips and freight trips would largely underestimate the AV trips, so AV trips are expected to have greater impact than discussed in current results. Further work would be look at the external passenger trips as well as the freight trips, and also the sensitivity analysis for various parameter assumptions for AVs, so to achieve a comprehensive of AV prediction on long-distance trips.

Chapter 10. Survey Analysis to Determine CAVs' Impact on Travel

10.1. Background

Public opinion regarding vehicle automation and fully automated, or autonomous, vehicles (AVs) are evolving rapidly. Past studies suggest that AVs, once a distant reality, are becoming more acceptable over time, and may be a real mode option in the relatively near future (see, e.g., Vujanic and Unkefer, 2011; Schoettle and Sivak, 2014; Bansal and Kockelman, 2016). Sommer (2013) reported that around half of Americans were concerned about riding in an AV, even though they admitted to the technology's many benefits, and this view was supported by respondents to Schoettle and Sivak's (2014) survey. A more recent U.S. survey, by Kelly Blue Book (2016), suggests that respondents believed conventional vehicles are still safer than AVs—at least for the time being. Schoettle and Sivak's (2016) second AV survey revealed similar reactions, with more than 35% of U.S. respondents very concerned about AVs, and partial autonomy less feared. Bansal and Kockelman (2016), MIT AgeLab (Abraham et al., 2016), Deloitte (2014), and Lee et al. (2017) have all concluded that younger people are more likely to use AVs, so demographic evolution is also important to consider, when anticipating the future use and adoption of advanced transport technologies. Until AVs are widely available in showrooms, at reasonably affordable prices, there will be regular fluctuations in public perceptions in any country or setting. Thus, regular survey efforts, and better surveys, with greater nuance, can make valuable contributions to transportation planning, policymaking, and vehicle production decisions.

With ride-hailing applications maintaining a steady increase in mode shares, especially in dense settings like San Francisco (SFMTA, 2015), and several studies illuminating the operational benefits of dynamic ride-sharing (DRS) (see, e.g., Agatz et al., 2010; Bischoff et al., 2016; Fagnant and Kockelman, 2016; Loeb et al., 2017; Farhan and Chen, 2017), a shift towards shared AVs (SAVs) with DRS options is expected. However, detailed studies on DRS have not yet been conducted. Bansal and Kockelman (2016) estimated SAV use for different pricing levels, but do not delve into ride-sharing. Quarles and Kockelman (2018) have recent, unpublished results that suggest about 16% of Americans are willing to share rides with strangers by paying about 40 percent less (e.g., 60 cents/mile rather than $1 per mile of SAV use). However, response-time or waiting-time analysis has not been carried out. To the best of the author's knowledge, only one study, in Australia (Krueger et al., 2016) captures such nuances, by modeling a discrete choice decision between SAVs without DRS, SAVs with DRS, and a respondent-specific travel alternative. They concluded that DRS is a preferred option among the young people and the people who regularly use car-sharing services, while recognizing the limitation of response bias that can emerge from posting such hypothetical questions. Similar studies in the United States have not yet been conducted.

Privacy and data security are another relevant topic, with one survey suggesting that privacy is Americans' top concern when choosing to not use AVs (Schoettle and Sivak, 2014). Existing work

in this area lacks many details: e.g., what are people willing to pay for privacy-enforcing measures? Related to this, automation can pose ethical dilemmas. Bonnefon et al. (2016) and Goodall (2017) believe that public opinion must be considered in crash-response programming and the like. Jenkins (2016) and Lin (2017) have described several possible outcomes of an inevitable crash scenario and Fleetwood (2017) censured algorithms that teach AVs to choose targets by force, arguing that they should not be readily allowed for public use. However, the public perception of what is most ethical in crash response contexts, and other situations, like who is to blame for a computer's decision or criteria to pass to be allowed to use SAVs, is yet to be determined. This survey adds new questions and public opinions to that discussion.

Finally, the long-distance (LD) travel implications of AVs are an important consideration. LaMondia et al. (2016) introduced AVs as new mode for LD trips originating in Michigan. Bansal and Kockelman (2016, 2017) suggested that LD-trip frequency may well double, and Perrine et al. (2017) are predicting major losses in U.S. airline revenues, long term, once AVs are widely available. However, many details are missing, especially questions that probe actual Americans on these topics. This study addresses many such investigative gaps.

10.2. Survey Design and Data Processing

The survey consists of 70 questions, tackling various aspects of AV and SAV use, including ride-sharing preferences, privacy and security concerns, ethical implications of crash response algorithms, LD travel shifts, and future travel choices, with each subject section having about 5 to 8 questions.

The section on current AV perceptions included questions on impressions of and WTP for AVs, SAV use, and DRS with strangers. Questions regarding an acceptable age for children/young people to travel individually or in a group were also asked, along with questions regarding opportunities for serving persons with disabilities. A slider response was used to obtain continuous responses on WTP, including for DRS with a stranger—by time of day (night vs. daytime) and assuming different time delays. The value of providing one's location en route (to a close friend or family member, to increase travelers' sense of security) was also addressed, when sharing an SAV ride with an unknown person.

To assess the ethical implications, three distinct ethical dilemmas were posed to the respondents: two regarding AV crashes with a pedestrian and other cars on the road, and one addressing crash responsibility. Questions on LD travel were based on mode-choice preferences for different types of trips and a respondent's typical LD trip. A demographic section was included towards the survey's end, to provide control variables and correct for various sampling biases, to better represent the U.S. population.

10.2.1. Data Collection

Survey Sampling International's (SSI's) panel of Americans was used to access respondents from across the United States. Nearly 10,000 Americans were targeted before the required sample

attributes were obtained, due to two screening procedures. The first screen blocked respondents from accessing the survey in its entirety if they failed to answer two initial basic questions regarding AVs and SAVs, after relevant information was provided. The second level of screening was done by removing respondents who took less than 15 minutes to complete the survey, since a low response time was deemed unrealistic for anyone going through this 70-question long survey. Both screens helped ensure respondents were intellectually engaged and paying attention.

Most questions contained a text input option as "Other: _____" for respondents to elaborate and expand response options. These inputs were manually mapped to an existing option or to a new option, as appropriate. After screening respondents and remapping responses, usable sample size was n = 2,588 respondents, from across the United States, with purposeful oversampling (n = 1258) of Texans, due to a strong interest in understanding Texans' preferences. Both sets of responses are given below, after a discussion on sample weighting or expansion.

10.2.2. Population Weighting

The 2,588 complete responses were associated with household and person-level weights to ensure that all reported statistics and regression analyses reflect the broader population of interest. The U.S. Census Bureau's Public Use Microdata Sample (PUMS) for years 2011–2015 provided national and state percentages across various classifications: location (Texas vs. U.S.), income and race, household size and worker count, vehicle ownership, age, gender, educational attainment and marital status. Certain demographics were under-represented (e.g., males who had not finished high school) and some others were over-represented (e.g., gender ratio was 47/53 rather than 49/51, 24% of the sample were people 65 years or older rather than 18%), resulting in slightly higher weights. A MATLAB code performed iterative proportional fitting over all the combinations of dimensions, ending once categorical percentages fell within 0.001% of the population percentages. Population-weighted sample characteristics are shown in Table 10.1. All of the following results reflect these adjustments to raw sample statistics.

Table 10.1 Survey data's population-weighted summary statistics

Sample Demographics	Mean	SD	Min	Max
Age (in yrs)	46.00	16.34	21	70
Gender (Male)	48.64 %	-	0	1
Employed Full-Time	37.59 %	-	0	1
Education – Bachelor's	17.56 %	-	0	1
U.S. License Holder	89.77 %	24.86 %	0	1
Disabled	7.91 %	-	0	1
HH Size	2.330	1.047	1	11
HH Annual Income	$70,340	$47,226	$5,000	$250,000
No. of Workers in HH	1.150	0.951	0	5
No. of Children in HH	0.535	0.917	0	9
No. of Vehicles in HH	1.750	0.960	0	6

10.3. Summary Statistics

10.3.1. Current AV Perceptions

As indicated earlier in this chapter, the survey's first section gauged perceptions of AVs. Table 10.2 summarizes the public's opinion on driving preferences, benefits offered with AV use, concerns in using them, and considerations at play in owning an AV. In general, Texans' responses do not differ by much, in any survey section, but there are some questions in which notable differences emerge. For example, 36.4% Americans enjoy driving conventional vehicles and do not plan on using AVs in the future, while just 26.7% of Texans give that response. In all, 31.8% Texans (versus 29.4% of Americans overall) want to keep the AV option open for their travel, even though they enjoy driving, while 15.0% of Texans (versus 11.6% of Americans) expect to prefer AV use to driving.

The great majority (92.9% Americans and 90.5% of Texans) believe that safety is a major AV benefit, yet over 60% are concerned that AVs may not be safe enough, with faulty software being a top concern. The mixing of AVs and conventional vehicles on public roadway is also an important concern. Top factors favoring AV ownership, instead of U.S. households relying more on SAVs, are the ability to store items in one's own vehicle and keeping one's own vehicle relatively clean or free of other's germs, while enjoying greater privacy and flexibility in their AV use decisions. It was unusual to find an AV's self-parking ability to be chosen by less than 2% of Americans as a major benefit. Proxy information about individuals with a disability was assessed, and 59.2% of Americans and 60.3% of Texans acknowledged that they knew at least one person among their immediate family, relatives, friends, or neighbors, who was disabled and would benefit from the use of SAVs.

Americans appear WTP, on average, $2073 more to own an AV as compared to a conventional vehicle, plus another $1078 if that new AV includes a human-driving mode option.

Table 10.2 Driving preferences and factors affecting AV ownership

Response Variable	U.S.	Texas	Response Variable	U.S.	Texas
Current driving preferences					
Enjoys driving and does not plan to use AVs	36.4%	26.7%	Does not like driving and will prefer AV use	11.6%	15.0%
Enjoys driving but will prefer some AV use	29.4%	31.8%	Prefers only non-motorized modes of travel	2.9%	0.9%
Prefers some driving as well as some AV use	17.5%	14.0%	Does not like driving but does not plan to use AVs	0.5%	1.3%
Expected major benefits of AVs & SAVs					
Safety improvement offered by AVs	92.9%	90.5%	Reliability	1.7%	5.1%
Congestion relief	2.8%	1.9%	Self-parking	1.4%	2.2%
Convenience of travel	2.6%	2.0%			
Expected major concerns of AVs					
Safety against crashes offered by AVs is still questionable				66.5%	62.5%
Faulty software in AVs				75.6%	71.1%
Confusion among human drivers and AVs on the streets				49.9%	51.9%
Privacy breaches inside AVs				16.9%	19.1%
Others tracking one's home or work location is easier with AVs				30.3%	39.3%
Factors causing one to own AVs instead of sharing SAVs					
Parking space availability	6.1%	7.4%	Privacy benefits of owning an AV	19.9%	15.4%
Relative cost of AVs over conventional cars	15.2%	11.0%	Hygiene concerns about SAVs that are not clean due to previous use and possible presence of germs	8.2%	11.6%
Availability of children's car seats in one's own AV	13.3%	14.7%	Security and safety	0.4%	1.3%
Ability to leave small items behind in one's own AV	21.4%	22.7%			
Storage space for large items	15.6%	15.4%			

10.3.2. Ride-Hailing and SAV Use

The survey's second section emphasizes ride-hailing applications and SAV use, including respondents' willingness to allow children to use AVs. Responses, shown in Table 10.3, suggest that only 32.5% of Americans (and 33.3% of Texans) have personal ride-hailing experience. Among these ride-hailing users, only 27.3% (across the U.S., and 14.7% from Texas) have shared their rides with strangers.

Texans appear to believe that children should be at least 17 years to use privately owned (household) AVs, while the average American appears comfortable with a 16-year-old threshold. However, 62.2% of Americans were against the idea of sending their own children, at any age, in an SAV, without an adult escort. Texans were slightly more comfortable in such private-AV-use behavior, with an acceptance rate of 45.7%.

Table 10.3 Americans' perspectives on ride-hailing and SAV use

Response Variable	U.S.	Texas	Response Variable	U.S.	Texas
Age appropriate for RIDE-HAILING services			*Age appropriate for children to use parents' AVs*		
Median age (in years)	16.0	16.0	Median age (in years)	16.0	16.0
Average age (in years)	16.0	16.3	Average age (in years)	16.4	17.4

Response Variable	U.S.	Avg. Age	Texas	Avg. Age
Is it acceptable to allow a group of children use an SAV without adult supervision?				
Yes, if there are all at least XX years old.	26.2%	16.2 yrs	27.9%	16.1 yrs
Yes, if any one child in the group is at least XX years old.	23.0%	16.8 yrs	30.9%	16.7 yrs
No, it is not acceptable to send children in SAVs.	62.2%		54.3%	

10.3.3. Ride-Sharing with Strangers and Willingness to Pay (WTP)

Public opinion on ride-sharing with strangers (while using an SAV) was assessed in detail. First, a hypothetical 5-mile SAV trip was presented and rising travel times (to reflect delay from adding another passenger) were added to this trip. Next, each respondent's willingness to share the same, hypothetical, 5-mile trip during the night was assessed. Maximum travel delays for sharing trips during the middle of the day and during the night were identified. Any added willingness to use DRS when their location was continuously available/broadcast to a family member (or friend) was also recorded, for both cases of day and nighttime trip-making. In addition to these preferences, the ideal cost of using an SAV in order to willingly let go of a currently owned household vehicle was obtained for different SAV response times (i.e., the time taken between a trip request and the SAV's arrival at the traveler's origin). All these results are summarized in Table 10.4.

As shown in the table, only 62.5% Americans and just 54.9% of Texans may be willing to share their ride with strangers when no delay accrues (i.e., no time is added to their 5-mile trip). This willing-to-share-rides pool of respondents reported an average WTP of 74¢ per trip-mile. Interestingly, all scenarios of added travel time returned a similar average. Americans (and Texans) may be more interested in their trip distance than their travel time, once they have opted to share their ride.

Table 10.4 Ride-sharing preferences during daytime

Response Variable	U.S.	Texas
Willingness to use SAV with strangers, no additional time		
Yes	22.5%	30.0%
Maybe	40.0%	24.9%
No	37.5%	35.1%
Average WTP (per mile)	$0.74	$0.71

Response Variable	U.S.	Texas	Response Variable	U.S.	Texas
Willingness to use SAV with strangers, 5 min. additional time			*Willingness to use SAV with strangers, 15min. additional time*		
Yes	18.5%	23.2%	Yes	6.0%	8.8%
Maybe	34.8%	31.9%	Maybe	19.1%	21.6%
No	46.7%	45.0%	No	75.0%	69.6%
Average WTP (per mile)	$0.73	$0.69	Average WTP (per mile)	$0.79	$0.65
Willingness to use SAV with strangers, 30 min. additional time			*Willingness to use SAV with strangers, 1 hr. additional time*		
Yes	2.8%	2.7%	Yes	2.2%	2.2%
Maybe	7.9%	15.6%	Maybe	4.2%	5.7%
No	89.4%	81.7%	No	93.6%	92.1%
Average WTP (per mile)	$0.77	$0.65	Average WTP (per mile)	$0.74	$0.62

Table 10.5 describes willingness to share rides (including trip durations, in DRS mode) during the day and the night. Very few Americans (just 4.4%, versus 11.0% of Texans) seem willing to share their rides at night (though this may well change, as people become more accustomed to SAV and DRS services in the future). Of those willing to use DRS during the middle of the day, 4.0% more Americans are willing if the service is offered only to people without a prior criminal record. Americans (and Texans) are willing to pay a 10¢-per–mile premium, on average, to share a ride during the night (presumably because they need more chauffeured trips at night [for consumption of alcohol, for example] or expect lower supply of SAVs at night). On average, respondents are more willing to tolerate trip delays at night, presumably because time constrains (on work and school arrivals, for example) are more severe during the daytime.

Table 10.5 Ride-sharing preferences at night

Response Variable	U.S.	Texas	Response Variable	U.S.	Texas
Willing to share a ride with a stranger in an SAV during the night?					
Yes				4.4%	11.0%
Maybe, if the stranger has no criminal record				8.0%	5.7%
Maybe, if the stranger's identifying information is given ahead of time				4.0%	5.0%
No				83.7%	78.3%
Average WTP for those willing to share (in $/mile)				$0.87	$0.85
Maximum trip duration for DRS (with a stranger) in an SAV during middle of day (in minutes)					
Mean	29.0	32.6	Median	25.0	26.0
Maximum trip duration for a shared ride in an SAV during the night (in minutes)					
Mean	34.8	35.4	Median	29.0	30.0
Maximum trip duration between day and night among those willing to share a ride both in the day and in the night					
Average during the day (in minutes)	40.4	47.5	Average during the night (in minutes)	34.8	35.4

Additional DRS features, like location information broadcast to family or friends for safety purposes, resulted in more people (roughly 15%) willing to share rides (during the day and at night). However, as seen in Table 10.6, more than 60% of Americans (and Texans) remained unwilling to ride-share in an SAV. And over 90% seemed hesitant about paying for such a service. Among those willing to pay for such a service, Texans appear to be more concerned about their safety than other Americans.

Table 10.6 Effects of ride-sharing trip locations being broadcasted

Response Variable	U.S.	Texas	Response Variable	U.S.	Texas
Willingness to use SAV when location is continuously broadcast to family member or friend					
During the middle of the day…			*During the night…*		
Yes, if the location is constantly broadcasted to family	43.0%	50.1%	Yes, if the location is constantly broadcasted to family	21.8%	30.9%
Yes, even without the location being broadcasted to family	16.4%	18.7%	Yes, even without the location being broadcasted to family	10.4%	7.4%
Not willing to share a ride with anyone	40.6%	31.2%	Not willing to share a ride with anyone	67.8%	61.7%
WTP for location to be broadcasted to family or friends (to enhance trip safety)					
During the middle of the day…			*During the night…*		
Yes	8.6%	7.9%	Yes	6.8%	14.3%
Maybe	18.1%	30.2%	Maybe	8.5%	8.0%
No	73.2%	61.8%	No	84.7%	77.7%
WTP to share a ride with unknown person during the night if trip locations are continuously broadcast to family or friends					
Average WTP (in $/mile)				$0.19	$0.23

Table 10.7 summarizes the cost that an SAV must be operated at, for different response times, so that the respondent is comfortable letting go of an existing household vehicle. The American Automobile Association (AAA, 2016) estimates that current vehicle ownership and operating costs average 50 to 80 cents per mile, once depreciation of purchase costs is reflected. Those costs can be higher or lower for vehicles driven fewer or more miles per year than the typical U.S. household vehicle. Interestingly, respondents are willing, on average, to pay about that same amount for SAV access—and Texans tend to offer more money than the average American. SAV users can avoid vehicle maintenance and parking costs and hassles, but they cannot guarantee how quickly SAVs will get to them, like they can when walking to their parked vehicle. Actual SAV system experiences will end up impacting everyone's WTP, and service times may vary a fair bit by location (e.g., urban vs. suburban trip ends). It is an interesting evolution of supply and demand that should one day play out around the world.

Table 10.7 Cost of SAVs at different response times to persuade reduction in current vehicle ownership

Response Variable	U.S.	Texas	Response Variable	U.S.	Texas
Cost of using SAV in order to replace vehicles that a respondent's household currently owns (in $/mile)					
Average response time under 1 minute	$0.75	$0.83	Average response time under 10 minutes	$0.52	$0.62
Average response time under 2 minutes	$0.71	$0.75	Average response time under 30 minutes	$0.38	$0.54
Average response time under 5 minutes	$0.64	$0.71			

10.3.4. Privacy Concerns using AVs and SAVs

Privacy is not on top of respondents' minds when AV-related concerns are requested at the survey start. However, when targeted as a separate topic, more privacy-related concern was observed. Table 10.8 demonstrates this, with 89% of Americans (and 83% of Texans) to at least some privacy concerns. However, many respondents (39.8% of Americans and 40.6% of Texans) appear unwilling to pay to anonymize their location while using SAVs. Respondents were also asked to rate their levels of comfort when their location data is used for different socially meaningful purposes. Nearly 48% Americans, on average, were comfortable or somewhat comfortable with this data being used for policing activities, managing traffic and for general community surveillance. However, more than half were against targeted advertising use.

Table 10.8 Privacy concerns related to AVs and SAVs and WTP for privacy

Response Variable	U.S.	Texas	Response Variable	U.S.	Texas
WTP for anonymizing user location for the entire trip while using an AV or SAV if they opt in					
Average (in $/trip)				1.10	1.19
Comfort level in allowing trip-location data usage…					
…to aid policing activities with a warrant			*…for general community surveillance*		
Very uncomfortable	17.7%	15.9%	Very uncomfortable	19.2%	26.1%
Somewhat uncomfortable	6.2%	9.1%	Somewhat uncomfortable	14.0%	15.1%
Unsure	22.4%	29.7%	Unsure	30.0%	26.3%
Somewhat comfortable	27.8%	23.6%	Somewhat comfortable	23.8%	21.6%
Very comfortable	25.9%	21.7%	Very comfortable	13.0%	10.9%
…to manage traffic & forecast travel conditions			*…to facilitate directed advertising*		
Very uncomfortable	15.4%	18.8%	Very uncomfortable	42.5%	49.2%
Somewhat uncomfortable	8.7%	12.6%	Somewhat uncomfortable	17.9%	21.3%
Unsure	22.4%	24.3%	Unsure	24.0%	15.9%
Somewhat comfortable	39.0%	30.2%	Somewhat comfortable	11.8%	10.2%
Very comfortable	14.5%	14.1%	Very comfortable	3.8%	3.4%

10.3.5. Crash Ethics While using AVs

Two distinct crash scenarios were presented in the survey, describing an AV crashing into a group of pedestrians in one case and crashing into other cars on the road in another. Respondents picked from a broad list of options to describe ethical and non-ethical crash outcomes. Table 10.9 outlines the opinions regarding the most ethical outcomes along with the person or business that should be held accountable for such events.

The most popular common believe is that AVs should not change course, once a crash is inevitable, and should crash into the first pedestrian or vehicle that crosses its path. Many others feel strongly that vehicle and pedestrian differences should be ignored while heading into a crash. Presumably, Americans recognize that there is not great solution to most crash situations and no new target (like a heavier vehicle or older adult) should be picked, leaving outcomes more to random chance and relatively similar to what humans may do under such difficult situations, with little response time available. Nevertheless, a strong share of respondents (about 20 percent) would like children to be avoided, when feasible, and more crash-hearty vehicles be selected, to minimize loss of life. More than 60% believe that AV manufacturers should be held responsible for such crashes.

Table 10.9 Crash choices and responsibilities

Response Variable	U.S.	Texas
Scenario 1: AV inevitably crashing into a group of pedestrians		
AVs must not change course, no matter what, and must crash into whoever is ahead.	54.2%	47.6%
The crash must should occur without any biases or preferences on age, race and gender of individuals in the group of pedestrians.	24.8%	26.4%
Children must be avoided under all circumstances.	19.2%	21.1%
Respondent is unsure if any of the options correctly describes an ethical outcome.	6.8%	9.2%
AVs must avoid crashing into friends identified in this group.	3.3%	4.2%
The AV must change into its human-driven operation mode so that the human can instinctively decide.	0.7%	0.3%
The occupant of the AV must be sacrificed for agreeing to use such a vehicle.	0.3%	0.2%
Scenario 2: AVs inevitably crashing into other vehicles on the road		
The crash must occur without any biases on vehicle-type, value or insurance.	38.4%	38.9%
AVs must not change course, no matter what, and must crash into the first vehicle it encounters.	31.8%	31.8%
The crash must occur such that the overall harm to human-life is minimized (e.g., AVs can crash into bigger vehicles.)	19.9%	19.5%
The crash must occur such that the harm to the AVs occupants is minimized.	11.4%	12.7%
Respondent is unsure if any of the options correctly describe an ethical outcome.	5.9%	6.5%
The crash must occur such that cars identified as belonging to a friend must not be damaged.	1.6%	2.7%

Response Variable	U.S.	Texas	Response Variable	U.S.	Texas
Who should take responsibility for all damages in an unavoidable crash?					
The AV manufacturer should take responsibility.	60.9%	59.7%	Respondent does not hold an opinion.	5.0%	4.8%
The programmer who built the AV's algorithm.	23.2%	23.2%	Should be decided by insurers.	1.4%	0.4%
Crashes will continue to occur; no one needs to take responsibility.	19.6%	22.2%	The courts should decide.	0.6%	1.7%
The individual who owns the AV and knows the risks that entail operating the vehicle should be held responsible for the crash.				0.4%	1.0%

10.3.6. Long-Distance Travel Choices

Various LD trip-making behaviors were investigated, including frequency of LD trip-making (per month), the longest trip made over the past year, share of LD trips with other persons (e.g., alone versus with friends, family, or colleagues), and mode preferences (across trip purposes and distance bands). Most LD trips occur with family members, and most respondents travel more LD often for personal trips than for business or vacation.

Over 80% of Americans (and Texans) prefer to use their own household vehicle for any non-business trip type under 500 miles. With the introduction of AVs and SAVs, conventional-

(human-driven) vehicle choice for non-business LD trips under 500 miles drops to 40%. AVs and SAVs enjoy a combined mode preference of 49.6% for business trips between 50 and 500 miles (one-way distance). For distances over 500 miles (one-way), air travel is preferred, for all trip types. Respondents may be expecting that they somehow can better afford air travel in the future, since this mode split is not consistent with current airline use splits. These results may be game-changers for travel demand forecasting when included in LD mode choice analyses in all statewide, national, and international travel models.

10.4. Model Results

10.4.1. Willingness to Pay for Dynamic Ride-Sharing

WTP for DRS in an SAV was estimated in two parts, to reflect the high number of respondents unwilling to share rides with strangers, as shown in Table 10.10.

Table 10.10 Respondents unwilling to share rides (in an SAV, for different added times)

Added Time	% Respondents not WTP to Share Rides
0 minutes	37.47%
5	46.70
15	74.99
30	89.37
60	93.63

The two-part model is motivated by Cragg's (1971) hurdle regression specification and was estimated using Stata software (StataCorp., 2015). This approach assesses the hurdle beyond which a particular event occurs. Here, the hurdle is one being WTP to share a ride and is estimated as a selection variable, s_i, using the maximum likelihood techniques while allowing for unobserved heteroscedasticity (across respondents) as a function of age. Correlation between responses from the same respondent was accounted for using data stratification in Stata, and an independent and identically distributed epsilon is assumed between respondents. A zero-dollar lower bound for each respondent's WTP was imposed as shown below., where $\mathbf{x_i}$ is the vector input of predictor variables affecting this $0 selection, $\boldsymbol{\beta_1}$ is the associated vector of model coefficients and $\varepsilon_{i,1}$ is (assumed to be) a normally distributed error term.

$$s_i = \begin{cases} 1 & if\ \mathbf{x_i}\boldsymbol{\beta_1} + \varepsilon_{i,1} > 0 \\ 0 & otherwise \end{cases}$$

The second part of the model estimates the specific amount that one is WTP using an exponential regression function, in cases where $s_i = 1$. Both equations are estimated simultaneously using maximum likelihood estimation (MLE). An exponential regression function ensures that WTP estimates can only be positive, with $\mathbf{z_i}$ serving as the vector of predictors or explanatory variables, $\boldsymbol{\beta_2}$ the vector of parameters to be estimated, and $\varepsilon_{i,2}$ as another set of independent, identically distributed normal error terms.

$$Y_i = \exp(\mathbf{z_i \beta_2} + \varepsilon_{i,2})$$

Table 10.11 shows the estimated parameters for both the selection model and exponential regression model. As expected, the travel time added via ride-sharing significantly affects respondents' decision to ride-share. Presence of a worker in the household reduces one's willingness, perhaps because workers have more constrained activity patterns, and so desire or need more independent travel. Interestingly, older people (everything else constant) and those with drivers' licenses are estimated to be less likely to share a ride. Those in households with annual incomes between $75,000 and $125,000 appear more likely to share a ride, as compared to other income brackets. It is possible that lower income brackets cannot simply afford to use an SAV, while those in higher income brackets prefer private rides.

Respondents with an associate's degree or higher are more willing to share rides (i.e., offer a non-zero valuation for such travel), everything else constant. Interestingly, those currently living in more densely populated but less densely employed neighborhoods appear less willing to share rides, and this could be people living close to downtown where walking gets you to most places.

While coefficients of the exponential regression model cannot be used directly to infer changes in one's expected WTP (due to the non-linear transformation that ensures non-negativity in this response variable), one finds that added travel time does not significantly affect WTP once a traveler is ready to share a ride. Older persons and those without any college education appear to be WTP a lower value, assuming they are already willing to share a ride, in this hurdle model specification.

Table 10.11 Model estimation results for WTP to share a ride

Selection Model		
Independent Variables	***Coefficients***	***T-stat***
Constant	1.14	4.86
Time added to the shared ride (in minutes)	-0.04	-13.80
Worker present in the household?	-0.30	-2.61
Age of respondent (in years)	-0.01	-3.83
Have U.S. driver's license?	-0.47	-2.59
Household income between $75k and $125k?	0.36	3.22
Has attended some college?	0.26	2.14
Population density (per square mile)	-0.3E-4	-2.99
Employment density (per square mile)	0.5E-4	3.08
Exponential Regression Model		
Independent Variables	***Coefficients***	***T-stat***
Constant	-0.68	-4.82
Age of respondent (in years)	0.01	3.13
Has attended some college?	-0.21	-2.66
Functional Variables for Heteroscedasticity		
Age of respondent (in years): Exponential model	-0.01	-8.00
Fit statistics		
Final log-likelihood		-1992.5321
Pseudo R-square		0.7034
Likelihood Ratio Chi-Square		9450.88
Number of observations (number of respondents)		12,940 (2,588)
F-test (2, 2586)		7.29

The change in response when each of the covariates was changed by one standard deviation was computed to understand how the expected WTP to share rides may change and this is tabulated in Table 10.12. For continuous variables, like respondent's age, the marginal expected value of WTP is calculated one standard deviation away from the mean age, in both directions. For indicator variables, the change in responses are determined by completely switching from a base level (like 0), to the next or subsequent levels (for example, 1, 2, or 3) and then calculating the marginal expected value of WTP at that point. Essentially, a continuous covariate's mean, plus/minus one standard deviation, is used to compute the new mean WTP for the sample and this percent change with respect to the previous mean is tabulated and for indicator variables, these percent change values are calculated by assuming all responses are at a high (that is, 1) or some intermediate point (like 2, 3 or 4 in a multi-level indicator) and then calculating the new mean. Computed changes in expected value of WTP with respect to the initial mean suggests that the lack of a driver's license affects mean values the most, by increasing it by 38%. When everything else is constant, a one standard deviation in average age of Americans can reduce the expected WTP by 27%. However, as Americans continue to age, the increase in average age will bring it down. As more people fall into the middle-class household income category, results suggest that there will be a 26% increase in average WTP to share rides.

Table 10.12 Covariate elasticities for WTP to share rides

Independent Variables		% Change in WTP
Worker present in the household?	Y	+19.6%
	N	-7.84%
Age of respondent (in years)	+1SD	-26.86%
	-1SD	+18.07%
Have U.S. driver's license?	Y	-4.73%
	N	+38.19%
Household income between $75k and $125k?	Y	+26.06%
	N	-6.61%
Has attended some college?	Y	+6.71%
	N	-10.02%
Population density (per square mile)	+1SD	-19.54%
	-1SD	+10.49%
Employment density (per square mile)	+1SD	+21.56%
	-1SD	-5.92%

10.4.2. Willingness to Pay to Anonymize Location while Using SAVs

A similar hurdle exponential regression was estimated to determine one's WTP to anonymize pick-up and drop-off locations while using SAVs. Table 10.13 shows the estimated coefficients for the two-part model. As expected, respondents who are concerned about privacy are more likely to be WTP to anonymize their location. Disabled people and females are more likely to be WTP, perhaps because they feel that they are relatively vulnerable and make an easier target for criminal behaviors. Vehicle ownership is also estimated to increase a respondent's WTP to a non-zero value for this anonymization benefit. Older people and those in smaller households are estimated to be less likely to pay to anonymize their locations. Household income is an interesting factor in this decision, since it oscillates back and forth between different income groups. In terms of one's level of payment, model results suggest that older persons and Caucasians are more WTP than those with a driver's license or those in households with more children.

Table 10.13 Model estimation results for WTP to anonymize location while using SAVs

Selection Model		
Independent Variables	***Coefficients***	***T-stat***
Constant	-0.40	-1.61
Concerned about privacy?	1.73	9.26
No disability?	-0.69	-5.75
Household owns 1 vehicle?	0.60	5.40
2 vehicles?	0.67	5.48
3 vehicles?	0.63	4.64
4+ vehicles?	0.66	4.14
Household size equal to 2?	0.16	2.02
equal to 3?	0.27	2.67
equal to 4+?	-0.11	-1.13
Household workers equal to 1?	-0.12	-1.54
equal to 2?	-0.10	-1.07
equal to 3?	-0.47	-3.14
equal to 4+?	-0.51	-1.89
Age of respondent (in years)	-0.02	-11.14
Is Male?	-0.35	-6.35
Household income: < $20,000	0.72	5.51
Or < $30,000	0.13	1.06
Or < $40,000	-0.02	-0.14
Or < $50,000	0.18	1.31
Or < $60,000	0.17	1.19
Or < $75,000	0.33	2.41
Or < $100,000	0.25	1.87
Or < $125,000	0.17	1.19
Or < $150,000	0.68	3.96
Or < $200,000	0.14	0.84
Or > $200,000	0.70	4.06
Exponential Regression Model		
Independent Variables	***Coefficients***	***T-stat***
Constant	-0.86	-7.23
Age of respondent (in years)	-0.4E-2	-3.24
Have U.S. driver's license?	0.26	3.72
Caucasian?	-0.14	-3.10
Household has 2 or less children?	0.48	6.11
Household income: < $20,000	0.23	2.45
Or < $30,000	0.52	5.20
Or < $40,000	0.39	3.67
Or < $50,000	0.18	1.77
Or < $60,000	0.08	0.72
Or < $75,000	0.41	4.07
Or < $100,000	0.38	3.94
Or < $125,000	0.38	3.60
Or < $150,000	0.36	3.22
Or < $200,000	0.54	4.52
Or > $200,000	0.06	0.56
Population density (per square mile)	-0.2E-4	-3.13
Employment density (per square mile)	0.1E-4	2.48

Variables with Heteroscedasticity		
Age of respondent (in years): Exponential model	-0.6E-2	-16.62
Fit statistics		
Final log-likelihood	-705.4893	
Pseudo R-square	0.6140	
Likelihood Ratio Chi-Square	2244.21	
Number of observations	2,588	

The changes in responses and marginal expected value of WTP are calculated for this model similarly to the previous hurdle model, as shown in Table 10.14. The percentage deviation of the expected value of WTP helps identify potential policy impacts to privacy and location anonymization decisions. Negative changes on all covariates showed that, although Americans seem to want privacy and may be willing to pay for anonymized trips, it may be unlikely that privacy will of trip locations will be a concern in the future. They also suggest that, moving forward, with the aging population and increasing average wages, there may be a decline in dollar amount that Americans are WTP to anonymize a trip.

Table 10.14 Covariate elasticities for WTP to anonymize location in an SAV

Independent Variables	% Change in WTP
No disability?	Y: -35.14%
	N: -13.14%
Household owns 0 vehicles?	-55.58%
1 vehicle?	-33.06%
2 vehicles?	-30.49%
3 vehicles?	-32.00%
4+ vehicles?	-30.85%
Household size equal to 1?	-36.15%
equal to 2?	-30.34%
equal to 3?	-26.54%
equal to 4+?	-40.16%
Household workers equal to 0?	-29.73%
equal to 1?	-33.98%
equal to 2?	-33.25%
equal to 3?	-46.97%
equal to 4+?	-48.47%
Age of respondent (in years)	+1SD: -55.57%
	-1SD: -15.00%
Is Male?	Y: -40.04%
	N: -27.38%
Household income: < $20,000	-20.95%
Or < $30,000	-32.46%
Or < $40,000	-42.49%
Or < $50,000	-40.00%
Or < $60,000	-42.33%
Or < $75,000	-28.55%
Or < $100,000	-32.34%
Or < $125,000	-35.31%
Or < $150,000	-18.09%
Or < $200,000	-31.93%
Or > $200,000	-26.23%
Have U.S. driver's license?	Y: -32.77%
	N: -39.13%
Caucasian?	Y: -35.15%
	N: -31.29%
Population density (per square mile)	+1SD: -36.51%
	-1SD: -29.98%
Employment density (per square mile)	+1SD: -29.50%
	-1SD: -34.41%

10.4.3. Long-distance Mode Choice with and without AVs and SAVs

Mode choice for LD travel was studied by first estimating a binary logit model when there are no AVs and SAVs available. Then, a multinomial logit model was estimated based on another survey question that included these modes. Correlation is allowed between responses from the same

respondent and an independent identically distributed Gaussian error term was assumed for observations between different respondents.

Table 10.15 shows the estimated coefficients as well as changes in the expected mode share of airplane for the binary logit model between the mode choices of a private car and an airplane. The private car was chosen as the base case and all coefficients can be interpreted with respect to this. The model suggests that business and recreational trip types are typically completed by airplanes. Trips greater than 500 miles, as expected, also use airplane for travel. Households owning one or more vehicles are less likely to prefer flying, provided everything else is constant. Single-person households seem the most interested in preferring to fly. Changes in household occupancies estimates a 32% increase if more single households were to exist. It is interesting to see the gradual change in preference among different income groups towards air travel. This is shown considerably well by their elasticities (the gradual change from negatives to positives for air travel). It is expected that wealthier households are more likely to fly to their destination, irrespective of business or pleasure. Interestingly, older people prefer to travel in their own vehicle as compared to the time-luxury of air travel. This could be because of lowered comfort level in an airplane as compared to that of their own vehicle. Caucasians prefer to drive their own car as compared to flying and this is most likely due to the heritage of driving in America. Households with children are unlikely to travel by air as compared to households without children.

Table 10.15 Model estimation and covariate elasticities for mode choice in LD travel without AVs and SAVs

Binary Logit Model				
	Independent Variables	***Coefficient***	***T-stat***	***Changes in Mode Share***
Alternative: Airplane (Base – Private Car)	Constant	0.76	1.15	
	Trip Type – Personal?	(base)		-20.63%
	– Business?	0.97	5.48	+23.56%
	– Recreation?	0.71	4.85	+10.87%
	Distance: 100 – 500 miles	(base)		-41.70%
	> 500 miles	1.78	13.69	+43.95%
	Household owns 0 vehicles	(base)		+38.09%
	1 vehicle?	-0.69	-1.38	+5.20%
	2 vehicles?	-0.87	-1.69	-2.88%
	3 vehicles?	-0.79	-1.42	+0.91%
	4+ vehicles?	-1.45	-2.25	-26.72%
	Household size equal to 1?	(base)		+32.25%
	equal to 2?	-1.30	-4.84	-25.89%
	equal to 3?	-0.50	-1.42	+8.23%
	equal to 4+?	-0.67	-1.37	+0.82%
	Household workers equal to 0?	(base)		+14.44%
	equal to 1?	-0.62	-2.58	-13.40%
	equal to 2?	-0.16	-0.52	+7.02%
	equal to 3?	-0.59	-1.18	-12.31%
	equal to 4+?	0.52	0.69	+39.84%
	Age of respondent (in years)	-0.01	-2.12	+1SD: -12.39% -1SD: + 8.70%
	Caucasian?	-0.68	-3.37	Y: -11.32% N: +19.60%
	No child in the household	(base)		+12.77%
	Children in the household: 1 child?	-1.57	-4.51	-49.92%
	2 children?	-0.22	-0.43	+2.53%
	3 children?	-0.25	-0.40	+0.99%
	4+ children?	-1.59	-2.12	-50.57%
	Household income: < $20,000	-1.53	-2.46	-64.76%
	Or < $30,000	-1.09	-2.23	-52.29%
	Or < $40,000	-0.58	-1.15	-34.44%
	Or < $50,000	-0.36	-0.77	-25.43%
	Or < $60,000	-0.78	-1.59	-41.89%
	Or < $75,000	0.33	0.64	+5.76%
	Or < $100,000	0.51	1.13	+14.70%
	Or < $125,000	0.94	1.90	+37.07%
	Or < $150,000	1.27	2.39	+54.04%
	Or < $200,000	1.17	2.35	+48.82%
	Or > $200,000	2.20	3.56	+100.16%
	Population density (per square mile)	0.4E-4	1.39	+1SD: +14.08% -1SD: - 6.60%
	Employment density (per square mile)	-0.7E-4	-1.31	1SD: -13.36% -1SD: +4.81%
Fit statistics				
Number of observations (number of respondents)			8,735 (2,039)	
F-test (33, 2006)			10.92	
Prob > F			0.00	

The multinomial logit model estimated under the assumption that AVs and SAVs are available and affordable shed some interesting inferences. Table 10.16 shows the estimated coefficients for this scenario. SAVs seem to be a dominating choice for business travel as compared to the other

modes as well as personal travel. Distance seems to only play a vital part in deciding to choose to fly. Current vehicle ownership does indicate that one may be less interested in AVs and SAVs, however, this is still a competing mode choice when other factors come into play. Older people still seem to prefer the private car as the most preferred alternative with AVs as their next choice, when everything else is constant. Having a current driver's license also deters people from using these automated modes. Regardless of the household's income bracket, there seems to be wide consensus in favoring SAVs as they are expected to turn out to be the most affordable alternative.

Table 10.16 Model estimation for future mode choice in LD Travel with AVs and SAVs

Multinomial Logit Model			
Alternatives (Base Case – Private Car)	**AVs**	**SAVs**	**Airplane**
Independent Variables	***Coefficient (T-stat)***	***Coefficient (T-stat)***	***Coefficient (T-stat)***
Constant	1.79 (1.67)	-0.48 (-0.34)	1.92 (1.49)
Trip Type – Personal?	(base)		
– Business?	-0.03 (-0.15)	1.23 (4.83)	0.56 (3.22)
– Recreation?	-0.06 (0.78)	0.15 (0.86)	0.16 (1.94)
Distance: 100 – 500 miles	(base)		
> 500 miles	0.10 (0.86)	0.05 (0.29)	1.55 (10.45)
Household owns 1 vehicle?	-0.84 (-1.23)	-0.36 (-0.45)	0.20 (0.27)
2 vehicles?	-1.27 (-1.85)	-0.24 (-0.28)	-0.21 (-0.27)
3 vehicles?	-0.65 (-0.88)	0.41 (0.44)	-0.26 (-0.31)
4+ vehicles?	-0.72 (-0.83)	0.26 (0.26)	-0.80 (-0.91)
Household size equal to 2?	0.91 (2.21)	0.42 (0.74)	-0.37 (-0.79)
equal to 3?	0.12 (0.21)	-0.23 (-0.29)	-0.01 (-0.01)
equal to 4+?	-0.25 (-0.33)	-0.51 (-0.48)	-0.21 (-0.31)
Household workers equal to 1?	-0.45 (-1.20)	-0.97 (-1.82)	-0.97 (-2.29)
equal to 2?	-0.30 (-0.69)	-0.32 (-0.49)	-0.12 (-0.25)
equal to 3?	-0.59 (--0.84)	-1.40 (-1.61)	-0.94 (-1.30)
equal to 4+?	0.75 (0.60)	-0.72 (-0.46)	0.07 (0.05)
Age of respondent (in years)	-0.02 (-2.14)	-0.03 (-1.92)	-0.03 (-2.63)
Have U.S. driver's license?	-2.41 (-3.85)	-2.26 (-3.30)	-1.88 (-2.31)
Caucasian?	-0.26 (-0.81)	-1.01 (-2.41)	-0.75 (-2.12)
Children in the household: 1 child?	0.50 (1.05)	0.90 (1.48)	-0.96 (-2.07)
2 children?	1.35 (1.75)	0.89 (0.86)	-0.68 (-1.01)
3 children?	2.30 (2.42)	1.87 (1.59)	0.21 (0.23)
4+ children?	-0.43 (-0.37)	0.19 (0.15)	-1.10 (-1.27)
Household income: < $20,000	0.78 (1.06)	1.75 (1.35)	0.34 (0.29)
Or < $30,000	0.94 (1.27)	3.21 (2.63)	-0.21 (-0.22)
Or < $40,000	0.69 (1.00)	2.98 (2.48)	0.22 (0.25)
Or < $50,000	0.20 (0.32)	2.37 (2.04)	0.79 (0.90)
Or < $60,000	1.76 (2.32)	4.84 (3.83)	0.88 (0.90)
Or < $75,000	1.35 (1.87)	1.75 (1.42)	1.43 (1.53)
Or < $100,000	0.83 (1.17)	3.72 (3.16)	1.50 (1.60)
Or < $125,000	1.51 (2.20)	3.75 (3.27)	2.03 (2.23)
Or < $150,000	1.62 (1.99)	3.10 (2.50)	2.30 (2.29)
Or < $200,000	1.74 (2.22)	2.41 (1.87)	2.29 (2.50)
Or > $200,000	1.41 (1.72)	2.60 (2.04)	2.11 (2.08)
Has attended some college?	0.23 (0.80)	0.89 (2.12)	0.75 (2.61)
Currently working at least part-time?	1.52 (3.07)	1.34 (2.02)	1.36 (2.34)
Single?	0.49 (2.17)	0.12 (0.37)	0.17 (0.65)
Population density (per square mile)	0.2E-4 (0.65)	0.5E-4 (1.24)	0.4E-4 (1.53)
Employment density (per square mile)	-0.5E-4 (-0.96)	-0.1E-4 (-1.06)	-0.8E-4 (-1.59)
Fit statistics			
Number of observations (no. of respondents)		9,257 (2,005)	
F-test (114, 1891)		5.74	
Prob > F		0.00	

In this case, the expected change in mode shares for all the modes discussed above is calculated. This is done by identifying the expected value of the mode share at the new mean value of the covariate. This helps see the practical effect of each covariate on future mode share. Table 10.17 shows the percentage change in mode-shares with respect to the previously determined share and gives an idea of the impact of each of the covariates. As evaluated from the coefficients previously, the absence of children may have a deep impact in choosing to fly compared to the other modes for LD travel. There may be a 67% increase in SAVs' mode-share mainly due to business travel. Absence of vehicle in the household also seems to favor use of AVs for future LD travel. Households with few (up to three) children may and significant number of workers prefer AVs for their LD travel and this could be directly from high total household income. Interest in SAVs is spread out through all income groups while results suggest that some income brackets may not use SAVs for their LD needs.

Table 10.17 Covariate elasticities for future mode choice in LD travel

Independent Variables	Change in Mode Share		
	AVs	SAVs	Airplane
Trip Type – Personal?	+3.84%	-24.96%	-7.22%
– Business?	-22.15%	+67.41%	+11.91%
– Recreation?	-5.01%	-16.40%	+1.43%
Distance: 100 – 500 miles	+19.53%	+24.46%	-38.74%
> 500 miles	-18.62%	-22.56%	+37.34%
Household owns 0 vehicles?	+43.60%	-10.43%	-18.83%
1 vehicle?	+2.08%	-31.03%	+12.20%
2 vehicles?	-15.35%	+1.82%	+4.84%
3 vehicles?	+14.32%	+51.67%	-18.32%
4+ vehicles?	+22.59%	+51.75%	-37.60%
Household size equal to 1?	-8.90%	+8.43%	+11.65%
equal to 2?	+33.38%	+22.20%	-27.20%
equal to 3?	-14.94%	-13.82%	+14.08%
equal to 4+?	-22.72%	-20.15%	+10.60%
Household workers equal to 0?	+0.61%	+33.71%	+8.97%
equal to 1?	+6.23%	-11.91%	-17.88%
equal to 2?	-10.77%	+11.72%	+14.82%
equal to 3?	+1.96%	-37.25%	-12.84%
equal to 4+?	+50.29%	-44.93%	-6.80%
Age of respondent (in years)	+1SD: -10.49%	-11.84%	-8.01%
	-1SD: +9.47%	-7.97%	+4.60%
Have U.S. driver's license?	Y: -5.47%	-3.45%	-0.22%
	N: +57.88%	+50.56%	-7.35%
Caucasian?	Y: +5.92%	-22.51%	-8.83%
	N: -6.31%	+32.34%	+13.96%
No child in the household	-17.73%	-23.61%	+19.81%
Children in the household: 1 child?	+23.66%	+65.69%	-39.38%
2 children?	+64.14%	+23.49%	-43.47%
3 children?	+83.96%	+38.39%	-39.40%
4+ children?	-31.94%	+36.73%	-14.44%
Household income: < $20,000	+14.55%	-53.05%	-29.39%
Or < $30,000	+23.18%	+56.66%	-54.99%
Or < $40,000	-3.97%	+45.38%	-32.73%
Or < $50,000	-32.27%	-32.00%	+6.72%
Or < $60,000	+23.35%	+196.60%	-44.63%
Or < $75,000	+22.17%	-77.55%	+6.73%
Or < $100,000	-23.53%	+44.52%	+17.40%
Or < $125,000	-5.78%	+6.78%	+30.00%
Or < $150,000	-4.64%	-51.50%	+45.19%
Or < $200,000	+5.64%	-76.17%	+43.47%
Or > $200,000	-8.94%	-61.88%	+44.25%
Has attended some college?	Y: -3.13%	+13.49%	+7.77%
	N: +9.88%	-27.15%	-16.67%
Currently working at least part-time?	Y: +54.87%	+13.29%	-8.15%
	N: -8.89%	-8.12%	+0.62%
Single?	Y: -40.26%	-7.50%	+21.73%
	N: +21.98%	-0.54%	-16.15%
Population density (per square mile)	+1SD: -5.41%	+20.62%	+10.06%
	-1SD: +1.27%	-7.28%	-5.04%
Employment density (per square mile)	+1SD: -1.82%	-15.68%	-9.44%
	-1SD: -0.53%	+9.10%	+2.07%

10.5. Practical Significance of CAVs

Americans appear apprehensive about using AVs, with Texans more willing to employ such automation. While Americans anticipate many crash benefits, 67% are concerned about initial safety issues and 76% are concerned about software defects. Trip-making privacy and the ability to leave things behind in a personally owned AV are key factors in preferring to own, rather than share, an AV. Average WTP is $2073 (above the cost of a conventional vehicle to own), plus another $1078 to ensure the AV retains a human-driving option. While the average Texan was more WTP for SAV rides (per mile traveled) than the average American, Texans demonstrated somewhat lower WTP for these privately owned AV technologies (at $1948 plus another $949, respectively).

More than 65% of survey respondents have not yet used a ride-hailing service, and only 25% of users had shared their ride (with an unknown traveler) in such vehicles. Most of these people (i.e., prior ride-hailing users) are not comfortable sending their children in a ride-hailing vehicle by themselves. Of those who are willing, the average minimum recommended age is 16 years old, for ride-hailing use. Their responses are similar for children using their parents' personally owned AVs in the future, with Texans suggesting (on average) a 17-year-old age threshold. If children are traveling in a group, Americans appear comfortable with the idea if at least one child being almost 17 years old; alternatively, all travelers should be at least 16 years of age (which is the standard minimum driving age in the U.S.).

Ride-sharing preferences among adults were assessed in detail here. For example, the WTP to share rides, with a stranger, is rather stable, at $0.75 per mile for the average American (and just $0.65 per mile for the average Texan), even in the face of added travel times up to 1 hour—at least for the small share of respondents willing to share rides under such trip-delayed circumstances. A hurdle model to predict this WTP during the day suggests that added travel time, respondent age and gender, household size and income (between $75k and $125k), disability and driver's license status, and presence of a worker in the home are important predictors of one's being WTP to share one's ride. After clearing these criteria, added travel time was not statistically significant, but variables of household size and vehicle ownership, respondent age, race, and land use variables were valuable predictors.

Few respondents appear willing to use DRS at night, but those who are willing state an average WTP of $0.87 per mile. Most people do not want to share a ride with someone they do not know for more than 30 minutes; but those who are most open to DRS are willing, on average, to ride-share for 45 minutes. More respondents are willing to share rides at night if their location is made constantly available to a family member or friend, adding another WTP of $0.19 per trip-mile (for this security benefit). While Texans were less WTP for adding automation to a privately held vehicle, they are more WTP for SAV service, per mile traveled (83¢ per mile, on average, versus 75¢/mile for the average American) when they have to give up one personal vehicle.

Higher levels of concern emerge when privacy is the focus of a survey question, rather than one among many potential issues to be selected by a respondent. Respondents are WTP, on average, more than $1 per trip to anonymize their location information (presumably wanting to obscure their home address most often). A hurdle model was used to estimate the WTP for anonymizing location while using AVs. Age, number of children in the household, vehicle ownership, and income were major predictors in determining one's WTP. They are against targeted advertising (based on their trip coordinates, for example), but comfortable with their data being used for policing, community surveillance, and/or traffic management decisions.

Crash ethics were also investigated, using three targeted questions based on different crash scenarios. The largest single share of Americans (54%) feel that any AV, when having no choice but to crash into one or more pedestrians (or other vehicles, in a related question [with 31% of respondents]) should not change its trajectory (to select a different pedestrian or vehicle to crash into), even if the current trajectory does not minimize overall harm. Avoiding children was also a popular response, but not the top response. AV manufacturers were dominantly (60.9% of respondents) deemed fully responsible for all such crashes. (Of course, some instances, like an inebriated pedestrian running about between two parked cars, would not be ascribed to manufacturers. Either way, the vehicle owner and specific design engineers are not being deemed responsible in such settings.)

Americans expect much of their LD travel (for trips over 50 miles, one-way) to shift toward AVs and SAVs. For example, nearly 50% of trips between 50 and 500 miles (one-way) are expected to eventually take place in an AV or SAVs, and this is considerably lower than LaMondia et al.'s (2016) prediction of around 55%, on average for these ranges. Airplanes are expected to deliver a major share of business trips (more than currently stated by respondents, perhaps due to some future-optimism bias about affordability). A binary logit model estimated that income played a vital role in determining mode choice in the current scenario without AVs or SAVs. However, a multinomial logit for LD mode choices in the presence of affordable AVs and SAVs, suggests that Americans prefer SAVs, irrespective of their household's income, ceteris paribus. Some business travel under 500 miles is also expected to be completed using SAVs. Older people are estimated to prefer to use their own vehicles, now and in the future.

These results suggest that Americans are not yet very confident about AV use, but expect to develop heavy usage levels. WTP, demand levels, perception, and public opinion are helpful to transportation planners and policymakers, technologists and vehicle manufacturers, fleet managers and system operators, as well as airlines, land developers, attorneys, insurers, and the tourism industry. Privacy in trip-making is a concern, with some respondents WTP to anonymize location data. Perceptions of ethics in crash choices should facilitate design of anti-crash algorithms. The aviation sector may wish to adjust its investments and future marketing strategies in response to changes in market share for LD travel. Regardless of position, preferences will evolve, as designs are rolled out and experience by more and more people, around the world. Regular survey efforts help nations and regions, companies and public agencies, better prepare for the coming paradigm shifts, hopefully with equity, environment, and efficiency in mind. The limitation to keep the

survey relatively brief meant that some other new innovative questions were removed before final dissemination.

Chapter 11. Traffic and Trade Impacts of Automated Trucking

11.1. Background

11.1.1. Motivation

Self-driving, fully automated, or autonomous vehicles (AVs) are an emerging transportation technology that may transform both passenger and freight transport decisions. Semi-automated trucks may enable automated driving under supervision and limited circumstances, such as driving long distances on an interstate. Fully automated self-driving trucks or "Atrucks" (autonomous trucks) are those that can leave the truck terminal and travel to a destination without human intervention or presence in the truck cab (Goodwill, 2017; Hawkins, 2018; Wakabayashi, 2018). Atrucks may be equipped with other automated functions, like drop-offs and pick-ups, but most experts expect an attendant on board, doing other types of work, sleeping as needed, and ensuring thoughtful deliveries and pick-ups. Such multi-tasking of vehicle attendants will allow for extended use of commercial trucks (e.g., every day, almost 24 hours a day) and greater labor productivity, resulting in lower per-mile and per-ton-mile freight delivery costs.

In 2014, trucks carried 1,996 billion ton-miles of freight around the U.S., or 37.7% of the nation's total ton-miles transported that year (BTS, 2017). Investment in and use of Atrucks will affect not only national and regional economies (Clements and Kockelman, 2017), but trade patterns, production levels, and goods pricing. Commercial trucks consume about 20% of the nation's transportation fuel, and self-driving technologies are predicted to reduce those diesel fuel bills by 4 to 7% (Liu and Kockelman, 2017; Barth et al., 2004; Shladover et al., 2006).

Atrucks can reduce some environmental impacts, lower crash rates, and increase efficiency in warehousing operations, line-haul transportation, and last-mile deliveries. Platooned convoys should enable following truck drivers to avoid certain restrictions on service hours, enabling longer driving distances. Uranga (2017) predicts greater use of Atrucks before passenger vehicle automation, thanks to the more obvious economic benefits of self-driving trucks (which start with higher price tags, making the automation investments less of a cost burden). Of course, driver job loss is also a concern, and the International Transport Forum (O'Brien, 2017) predicts that up to 70% of all U.S. truck-driving jobs could be lost by 2030 (due to vehicle automation). But trucks may still require driver presence, due to loading dock restrictions, unusual problems on the road, and more complex operating systems.

While there is active investigative interest on the travel and traffic effects of self-driving cars, research into the travel and traffic impacts of Atrucks is dearly lacking. This chapter anticipates Atrucks' trade pattern and production impacts across the U.S. and begins with a review of relevant works. It then discusses the random-utility-based multi-regional input-output (RUBMRIO) model methodology for tracking trade across zones or regions, describes a sub-nested mode choice model

for Atrucks (versus Htrucks), and the results of various trade-scenario simulations across U.S. regions, highways, railways, and industries.

11.1.2. Review on Autonomous Trucking and Trade Model

Two papers currently investigate U.S. travel shifts for long-distance (LD) passengers due to AV use (LaMondia et al., 2016; Perrine et al., 2017). Related topics include fuel consumption, congestion impacts, shared-fleet operations, dynamic ride-sharing, energy use, emissions, and roadside investments (see, e.g., Fagnant and Kockelman, 2014; Chen et al., 2016; International Transport Forum 2015; Land Transport Authority, 2017; Kockelman et al., 2016). LaMondia et al. (2016) forecasted U.S. mode shares for person-trips over 50 miles (one-way) from the state of Michigan, following the introduction of AVs. They predicted that 25% demand of airline passenger trips under 500 miles will shift to autonomous vehicles. Perrine and Kockelman (2017) anticipated destination and mode-choice shifts in LD U.S. person-travel, including a major loss (47%) of airline revenue, using 4,566 National Use Microdata Area zones (NUMAs). The anticipated long-term effects of AV access on LD personal travel are striking.

Some companies have written about the potential benefits of Atrucks. A DHL report (Kückelhaus, 2014) noted that Atrucks could lower their freight costs by 40% per vehicle- or ton-mile. Convoy systems would allow LD drives with large quantities of goods, through which Atrucks could reduce fuel use by 10 to 15% (Clements and Kockelman, 2017). Crash counts may fall by 50 percent or more (Kockelman and Li, 2016), along with various insurance costs. Atrucks cost-savings impacts on freight movement and industry siting and sizing decisions have been neglected. This new topic area of Atrucks is explored here.

Input-Output (IO) analysis, originally proposed by Leontief (1941), uses matrix algebra to characterize inter-industry interactions within a single region, as households and government agencies spend money on goods, which are produced by mixing inputs from other industries, and so on. Demand is met by production adjustments, based on expenditure linkages across industries. Isard's (1960) spatial IO model allows for spatial disaggregation using fixed shares. More recent extensions exploit random utility theory and entropy-maximization properties, as evident in the MEPLAN (Echenique et al., 1990), DELTA (Simmonds and Still, 1998), TRANUS (De la Barra et al., 1984), PECAS (Hunt and Abraham, 2003) and KIM models (Kim et al., 2002). These models also allow a land-use transportation feedback cycle, with freight and person (labor and consumer) flows responding to changes in network routes and travel costs.

The open-source RUBMRIO model is a similar extension, with applications to the state of Texas and U.S. counties. Kockelman et al. (2005) described the RUBMRIO's application to Texas's 254 counties, across 18 social-economic sectors and two modes of transport, meeting foreign export demands at 31 key ports. Huang and Kockelman (2010) developed a dynamic RUBMRIO model to equilibrate production and trade, labor markets and transportation networks simultaneously for Texas' counties over time (better recognizing starting distributions of labor and employment). Kim et al. (2002) used such a model for estimating interregional commodity flows and transportation

network flows to evaluate the indirect impacts of an unexpected event (an earthquake) on nine U.S. states, represented by 36 zones.

Guzman and Vassallo (2013) used a RUBMRIO-style approach to evaluate the application of a distance-based charge to heavy-goods vehicles across Spain's motorways. Maoh et al. (2008) used the RUBMRIO model to simulate weather impacts on Canada's transportation system and economy. Du and Kockelman (2012) calibrated the RUBMRIO model to simulate U.S. trade patterns of 13 commodities among 3,109 counties, with its nested-logit model for input origin and truck-versus-rail mode choices. They noted how transportation cost changes (from generically more efficient or less efficient travel technologies, for example) were important, especially for central U.S. counties.

This study builds off of the Du and Kockelman's (2012) work by adding the Atruck option into a sub-nest for mode choice, allowing for strong correlation in the Atruck vs. Htruck choice (since these are two very similar modes). Thirteen aggregate "industries" or socio-economic sectors are used here, since all nested logit model parameters are calibrated from FAF[4] data, which rely on SCTG commodity classes. Corresponding NAICS and IMPLAN codes are shown in Table 11.1, which is adapted from Du and Kockelman's (2012) work. The application's 13 sectors, technology costs, and other assumptions are described below.

Table 11.1 Description of economic sectors in RUBMRIO model

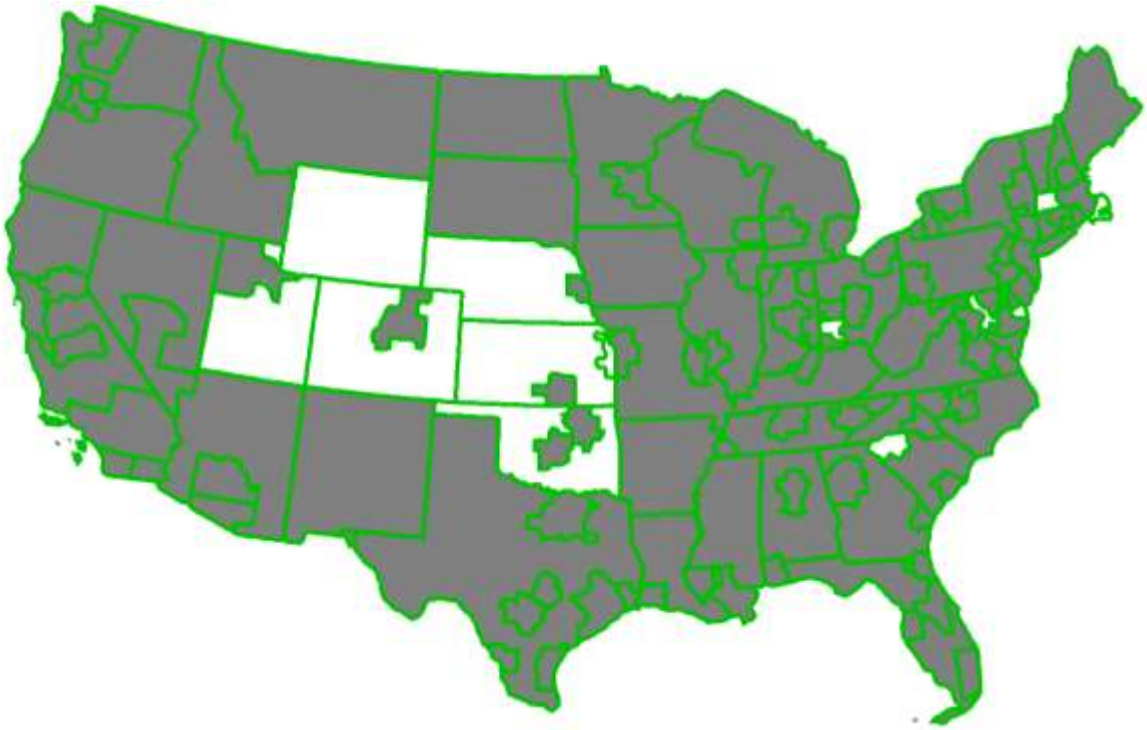
Sector	Description	IMPLAN Code	NAICS Code	SCTG Code
1	Agriculture, Forestry, Fishing and Hunting	1~19	11	1
2	Mining	20~30	21	10~15
3	Construction	34~40	23	--
4	Food, Beverage and Tobacco Product Manufacturing	41~74	311, 312	2~9
5	Petroleum and Coal Product Manufacturing	115~119	324	16~19
6	Chemicals, Plastics and Rubber Product Manufacturing	120~152	325, 326	20~24
7	Primary Metal Manufacturing	170~180	331	32
8	Fabricated Metal Manufacturing	181~202	332	33
9	Machinery Manufacturing	203~233	333	34
10	Computer, Electronic Product and Electrical Equipment Manufacturing	234~275	334, 335	35, 38
11	Transportation Equipment Manufacturing	276~294	336	36, 37
12	Other Durable & Non-Durable Manufacturing	75~114, 153~169, 295~304	313~316, 321~323, 327, 337	25~31, 39
13	Miscellaneous Manufacturing	305~318	339	40, 41, 43

11.2. Data Set Description

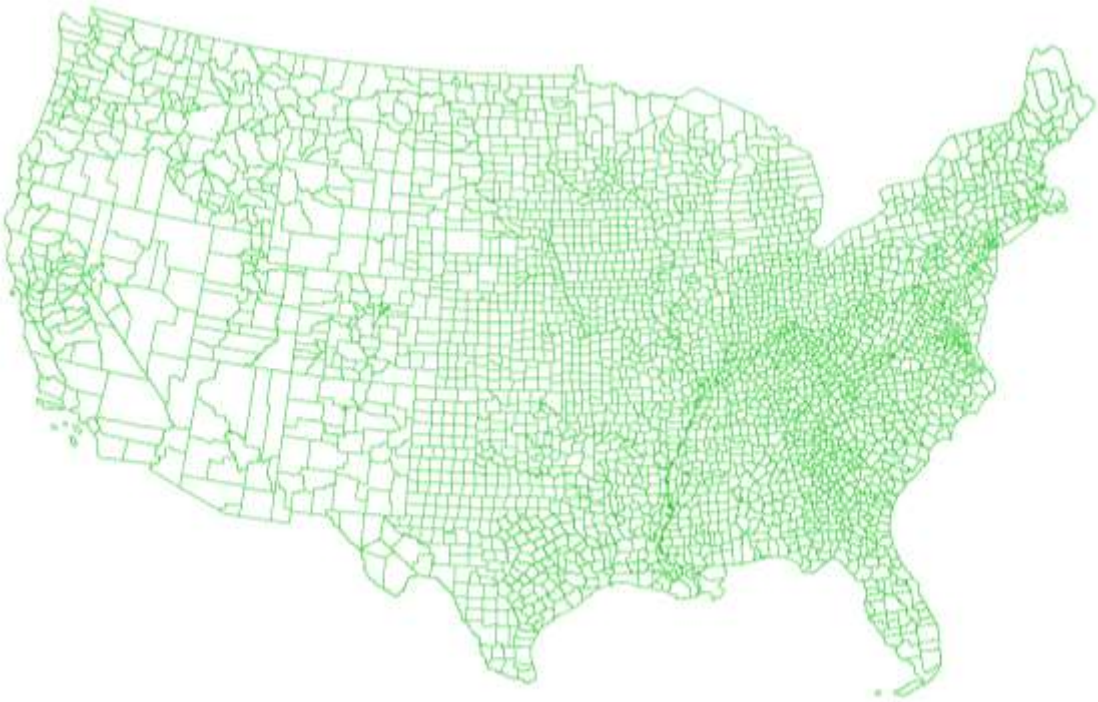
Data sets used here include the disaggregated freight zonal data from the U.S. Commodity Flow Survey (CFS), trade flow data from the U.S. DOT's Freight Analysis Framework (FAF) version 4, industry-by-industry transaction tables and regional purchase coefficients (in year 2008) from IMPLAN, and railway and highway network data from Caliper's TransCAD 7.0.

11.2.1. Freight Data

FAF^4 integrates trade data from a variety of industry sources, with emphasis on the Census Bureau's 2012 CFS and international trade data (Fullenbaum and Grillo, 2016). It provides estimates of U.S. trade flows (in tons, ton-miles, and dollar value) by industry, across 7 modes (truck, rail, water, air, pipeline, and others), and between FAF^4's 132 aggregate zones. FAF^4's origin-destination-commodity-mode annual freight flows matrices were used to predict domestic and export trade flows by zone. FAF^4 data show foreign export flows exiting the U.S. from 117 of these 132 zones, as shown in gray in Figure 11.1(a). So these same 117 zones serve as both production and export zones in this paper's trade modeling system.FAF^4 zones were then disaggregated into county-level matrices using the 2012 CFS boundary data (which identify the counties belonging to each FAF^4 zone). Ten metro areas were also added to the CFS data in year 2012, and 3109 contiguous counties [as shown in Figure 11.1(b)] remain, after excluding the distant states of Hawaii and Alaska. Interzonal travel times and costs by rail, Atruck and Htruck were all computed using TransCAD software, for the 3109×3109 county matrix based using shortest highway and railway paths in terms of free flow travel time. All intra-county travel distances were assumed to be the radii of circles having that county's same area.



(a) Continental United States' FAF[4] 132 Zones, with 117 Export Zones (shown in grey)



(b) Continental United States' 3109 Domestic Freight Counties

Figure 11.1 U.S. domestic and export zones for trade modeling

11.2.2. Economic Interaction Data

The model's embedded IO matrices' technical coefficients and regional purchase coefficients (RPCs) were obtained through IMPLAN's transaction tables, as derived from U.S. inter-industry accounts. Technical coefficients reflect production technology or opportunities (i.e., how dollars of input in one industry sector are used to create dollars of product in another sector) and are core parameters in any IO model. RPCs represent the share of local demand that is supplied by domestic producers. RPC values across U.S. counties are assumed constant here, since variations are unknown. However, counties closer to international borders are more likely to "leak" sales (as exports) than those located centrally, everything else constant. And production processes or technologies can vary across counties (and within industries, across specific manufacturers and product types, of course). This application assumes that all U.S. counties have access to the same production technologies, or technical coefficients table.

IMPLAN's 440-sector transaction table was collapsed into 13 industry sectors, plus Household and Government sectors to represent the U.S. economy in this trade-modeling exercise. Since FAF[4] uses the same 43 two-digit Standard Classification of Transported Goods (SCTG) classes (BTS, 2017) as the 2007 U.S. Commodity Flow Survey (CFS), IMPLAN's 440 sectors were bridged to a corresponding SCTG code based on the 2007 North American Industry Classification System or NAICS (Census Bureau, 2017). SCTG code 99 (for other good types) is not tracked here. See economic sectors for RUBMRIO model application table from Du and Kockelman (2012).

11.3. Specification of the RUBMRIO Model

In random utility choice theory, error terms enable unobserved heterogeneity in the decision-making process. Here, the RUBMRIO multinomial logit model has three branches, for origin choice, rail versus truck mode choice, and autonomous vs human-driven truck choice, as shown in Figure 11.2.

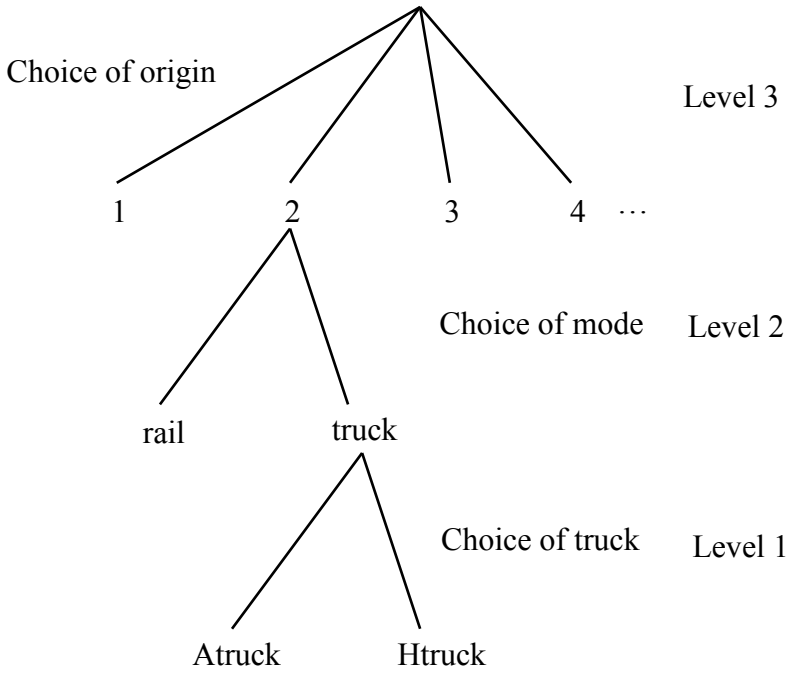


Figure 11.2 Random utility structure for shipment origin, mode, and truck-type choices

Equation (11.1) provides the three mode-choice utilities, conditioned on knowing a shipment's origin (i), destination (j), and industry or commodity type (m):

$$\begin{aligned} U^m_{ij,\,rail} &= \tilde{V}^m_{ij,\,rail} + \tilde{V}^m_{ij} + \varepsilon^m_{ij,\,rail} + \varepsilon^m_{ij} \\ U^m_{ij,\,truck,\,Atruck} &= \tilde{V}^m_{ij,\,truck,\,Atruck} + \tilde{V}^m_{ij,\,truck} + \tilde{V}^m_{ij} + \varepsilon^m_{ij,\,truck,Atruck} + \varepsilon^m_{ij,\,truck} + \varepsilon^m_{ij} \\ U^m_{ij,\,truck,\,Htruck} &= \tilde{V}^m_{ij,\,truck,\,Htruck} + \tilde{V}^m_{ij,\,truck} + \tilde{V}^m_{ij} + \varepsilon^m_{ij,\,truck,Htruck} + \varepsilon^m_{ij,\,truck} + \varepsilon^m_{ij} \end{aligned} \quad (11.1)$$

where

$\tilde{V}^m_{ij}$ = systematic utility of selecting origin *i* for acquisition of commodity *m*,

$\tilde{V}^m_{ij,\,rail}$, $\tilde{V}^m_{ij,\,truck}$ = systematic utilities associated with selecting origin *i* and rail mode/any truck type for movement of commodity *m*,

$\tilde{V}^m_{ij,\,truck,\,Atruck}$,$\tilde{V}^m_{ij,\,truck,\,Htruck}$ = systematic utilities associated with selecting origin *i* and Atruck/Htruck for movement of commodity *m*, and

ε^m_{ij}, $\varepsilon^m_{ij,\,rail}$, $\varepsilon^m_{ij,\,truck}$, $\varepsilon^m_{ij,\,truck,Htruck}$, $\varepsilon^m_{ij,\,truck,Atruck}$ = random error terms associated with shipment origin, rail mode, truck mode, human-driven truck and self-driving truck choice, respectively.

11.3.1. Nested Logit Model

11.3.1.1. Origin Choice (Level 3)

Relying on nested logit formulae provided in Ben-Akiva and Lerman (1978), the probability of commodity-type m inputs coming to zone *j* from zone *i* (i.e., the choice likelihood [or input share] of zone *i* as an origin for this good's demand in zone *j*) is given by:

$$P_{ij}^{m} = \frac{\exp(V_{ij}^{m})}{\sum_{i} \exp(V_{ij}^{m})} \tag{11.2}$$

where

$$V_{ij}^{m} = -p_{i}^{m} + \gamma^{m} \ln(pop_{i}) + \lambda^{m} \theta_{ij,mode}^{m} \Gamma_{ij,mode}^{m} \tag{11.3}$$

is the system utility using origin *i* for commodity *m*, and

$$\Gamma_{ij,mode}^{m} = \ln\left(\exp\left(\frac{V_{ij,\,rail}^{m}}{\theta_{ij,mode}^{m}} \right) + \exp\left(\frac{V_{ij,truck}^{m}}{\theta_{ij,mode}^{m}} \right) \right) \tag{11.4}$$

is the logsum of mode choice, with scale parameter $\theta_{ij,mode}^{m}$=1.2.

11.3.1.2. Mode Choice (Level 2)

Since the mode choice nested logit's random error terms are assumed to follow an independent and identically distributed Gumbel distribution, and setting the initial dispersion to scaling factor to 1, the probability of commodity *m* being transported by each of the two major modes (rail and truck), between any given *ij* pair, are as follows:

$$P_{rail|ij}^{m} = \frac{\exp\left(\frac{V_{ij,\,rail}^{m}}{\theta_{ij,mode}^{m}} \right)}{\exp\left(\frac{V_{ij,\,rail}^{m}}{\theta_{ij,mode}^{m}} \right) + \exp\left(\frac{V_{ij,truck}^{m}}{\theta_{ij,mode}^{m}} \right)}$$

$$P_{truck|ij}^{m} = \frac{\exp\left(\frac{V_{ij,truck}^{m}}{\theta_{ij,mode}^{m}} \right)}{\exp\left(\frac{V_{ij,\,rail}^{m}}{\theta_{ij,mode}^{m}} \right) + \exp\left(\frac{V_{ij,truck}^{m}}{\theta_{ij,mode}^{m}} \right)} \tag{11.5}$$

where

$$V_{ij,\,rail}^{m} = \beta_{0,\,rail}^{m} + \beta_{r,time}^{m} \times time_{ij,\,rail} + \beta_{r,cost}^{m} \times cost_{ij,\,rail}$$

$$\text{and } V_{ij,\,truck}^{m} = 0 + \theta_{ij,truck}^{m} \Gamma_{ij,truck}^{m} \tag{11.6}$$

are the general modes' systematic utilities and

$$\Gamma_{truck}^{m} = \ln\left(\exp\left(\frac{V_{ij,\,truck,Atruck}^{m}}{\theta_{ij,truck}^{m}} \right) + \exp\left(\frac{V_{ij,\,truck,Htruck}^{m}}{\theta_{ij,truck}^{m}} \right) \right) \tag{11.7}$$

is the logsum for the truck-mode choice, with scale parameter $\theta_{ij,truck}^{m}$ = 1.4 for base case. Travel time is a common component for the Atruck and Htruck utilities, since this work does not assume one is faster. In fact, Atrucks may complete long trips faster than Htrucks, since Atruck operators can sleep while the vehicle is en route. Here, the truck mode serves as the base mode, so only the rail mode has an alternative specific constant (ASC).

11.3.1.3. Truck Choice (Level 1)

The probability of freight flow commodity m from zone i to zone j using mode Atruck and Htruck respectively in nest truck is given by:

$$P^m_{Atruck|ij,truck}=P^m_{truck|ij}\times P^m_{Atruck|truck}=\frac{\exp\left(\frac{V^m_{ij,truck}}{\theta^m_{ij,mode}}\right)}{\exp\left(\frac{V^m_{ij,\ rail}}{\theta^m_{ij,mode}}\right)+\exp\left(\frac{V^m_{ij,truck}}{\theta^m_{ij,mode}}\right)}\times\frac{\exp\left(\frac{V^m_{ij,\ truck,Atruck}}{\theta^m_{ij,truck}}\right)}{\exp\left(\frac{V^m_{ij,\ truck,Atruck}}{\theta^m_{ij,truck}}\right)+\exp\left(\frac{V^m_{ij,\ truck,Htruck}}{\theta^m_{ij,truck}}\right)}$$

$$P^m_{Htruck|ij,truck}=P^m_{truck|ij}\times P^m_{Atruck|truck}=\frac{\exp\left(\frac{V^m_{ij,truck}}{\theta^m_{ij,mode}}\right)}{\exp\left(\frac{V^m_{ij,\ rail}}{\theta^m_{ij,mode}}\right)+\exp\left(\frac{V^m_{ij,truck}}{\theta^m_{ij,mode}}\right)}\times\frac{\exp\left(\frac{V^m_{ij,\ truck,Htruck}}{\theta^m_{ij,truck}}\right)}{\exp\left(\frac{V^m_{ij,\ truck,Atruck}}{\theta^m_{ij,truck}}\right)+\exp\left(\frac{V^m_{ij,\ truck,Htruck}}{\theta^m_{ij,truck}}\right)} \quad (11.8)$$

where

$$V^m_{ij,\ truck,Atruck}=\beta^m_{0,\ Atruck}+\beta^m_{t,time}\times time_{ij,truck}+\beta^m_{t,cost}\times cost_{ij,Atruck}$$
$$V^m_{ij,\ truck,\ Htruck}=0+\beta^m_{t,time}\times time_{ij,truck}+\beta^m_{t,cost}\times cost_{ij,Htruck} \quad (11.9)$$

are the system utilities of moving commodity *m* from zone *i* to zone *j* using Atruck and/or Htruck modes (in the truck nest).

11.3.2. RUBMRIO Model Specification

An equilibrium trade-flow solution (where all producers obtain the inputs they need, and all export demands are met) can be achieved in RUBMRIO via Figure 11.3's iterative equation sequence. Zhao and Kockelman (2004) proved this solution's uniqueness. Flow-weighted averages of shipments' travel costs create input costs, which merge together to create output costs, as commodities (and labor) flow through the production and trade system. Once the solutions have stabilities (with domestic flow value changing by less than 1% between iterations), final disutilities of travel and trade provide mode shares by origin-destination (OD) pair and commodity or industry sector.

This iterative process' calculations required about 2.25 hours using an Atruck-modified version of Kockelman et al.'s C++ open-source program.[27]

[27] This program is available at http://www.caee.utexas.edu/prof/kockelman/RUBMRIO_Website/homepage.htm.

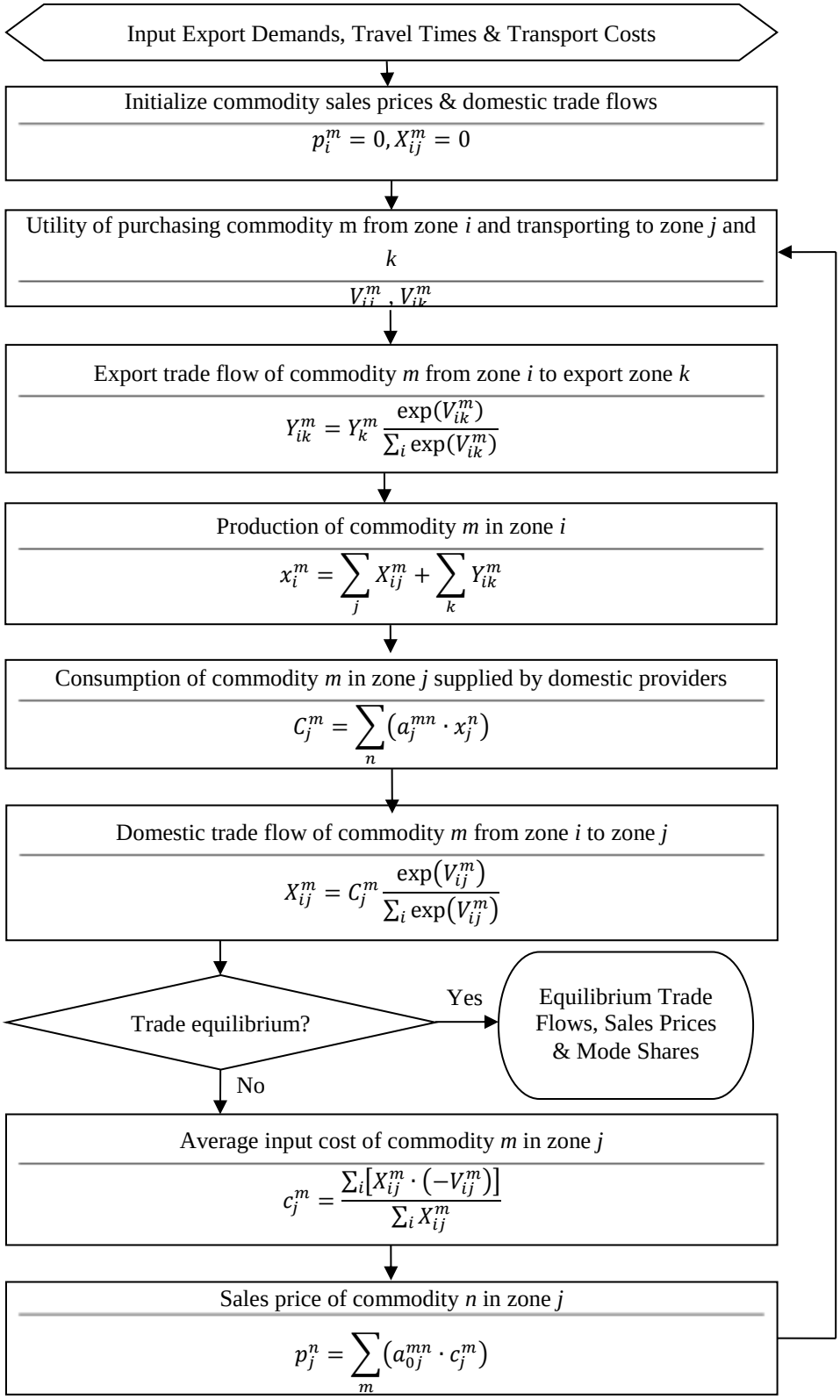


Figure 11.3 RUBMRIO solution algorithm (Adapted from Du and Kockelman [2012], Figure 2)

RUBMRIO's utility functions for domestic and export trade-flow splits (across shipment origin alternatives) depend on the cost of acquiring input type m from zone *i*, as well as zone *i*'s "size" (measured as population here). Since there are three mode alternatives for these shipments, with the two truck modes sub-nested, the competing travel costs can be shown as logsums (which reflect the expected maximum utility or minimum cost of acquiring that input from different origin zones). After substituting those logsums into Figure 11.3's trade-flow equations, one has Equations (11.10) and (11.11), where V_{ij}^m and V_{ik}^m are the utilities of purchasing one unit of industrial *m*'s goods from region *i* for use as inputs to zone *j*'s production process, or for export via zone *k*, respectively.

$$V_{ij}^m = -p_i^m + \gamma^m \ln(pop_i) + \lambda^m \times \theta_{ij,mode}^m \times \ln\left(\exp\left(\frac{\beta_{0,\,rail}^m + \beta_{r,time}^m \times time_{ij,\,rail} + \beta_{r,cost}^m \times cost_{ij,\,rail}}{\theta_{ij,mode}^m}\right) + \exp\left(\frac{\theta_{ij,truck}^m}{\theta_{ij,mode}^m} \times \ln\left(\exp\left(\frac{\beta_{0,\,Atruck}^m + \beta_{t,time}^m \times time_{ij,truck} + \beta_{t,cost}^m \times cost_{ij,Atruck}}{\theta_{ij,truck}^m}\right) + \exp\left(\frac{\beta_{t,time}^m \times time_{ij,truck} + \beta_{t,cost}^m \times cost_{ij,Htruck}}{\theta_{ij,truck}^m}\right)\right)\right)\right) \quad (11.10)$$

$$V_{ik}^m = -p_i^m + \gamma^m \ln(pop_i) + \lambda^m \times \theta_{ik,mode}^m \times \ln\left(\exp\left(\frac{\beta_{0,\,rail}^m + \beta_{r,time}^m \times time_{ik,\,rail} + \beta_{r,cost}^m \times cost_{ik,\,rail}}{\theta_{ik,mode}^m}\right) + \exp\left(\frac{\theta_{ik,truck}^m}{\theta_{ik,mode}^m} \times \ln\left(\exp\left(\frac{\beta_{0,\,Atruck}^m + \beta_{t,time}^m \times time_{ik,truck} + \beta_{t,cost}^m \times cost_{ik,Atruck}}{\theta_{ik,truck}^m}\right) + \exp\left(\frac{\beta_{t,time}^m \times time_{ik,truck} + \beta_{t,cost}^m \times cost_{ik,Htruck}}{\theta_{ik,truck}^m}\right)\right)\right)\right) \quad (11.11)$$

Parameter assumptions for γ^m, λ^m and β^m are based on Du and Kockelman's (2012) work, which has two levels of random utility structure: for origin and mode choices. Here, the rail's ASCs were set equal to the negative of the ASCs used for truck in their research, since a second type of truck mode was added as Atrucks. Moreover, the Atruck ASCs were assumed to be -0.1, because Atrucks should be somewhat preferred, after travel-cost and time considerations, thanks to safety and communications benefits. After assembling all these inputs, shown in Table 11.2, a series of different network and Atruck cost scenarios can be examined, using the RUBMRIO solution algorithms.

Table 11.2 Parameter estimates for origin, mode, and truck choice equations

Sector	Origin Choice Parameters		Mode Choice Parameters			Truck Choice Parameter			VOTT ($/hr)
	θ_{ij}^{m}=1		$\theta_{ij,mode}^{m}$=1.2			$\theta_{ij,truck}^{m}$=1.4			
	γ^{m}	λ^{m}	$\beta_{0,\ rail}^{m}$	$\beta_{r,time}^{m}$	$\beta_{r,cost}^{m}$	$\beta_{0,\ Atruck}^{m}$	$\beta_{t,time}^{m}$	$\beta_{t,cost}^{m}$	
1	0.05	0.90	-3.38	-4.81	-4.85	-5.61	-5.66	-0.10	24.18
2	0.41	7.66	-1.11	-1.03	-2.01	-1.20	-2.34	-0.10	2.12
4	0.86	-2.86	-3.36	2.17	0.56	2.53	0.65	-0.10	6.15
5	0.10	2.02	-1.00	-1.87	-4.09	-2.18	-4.77	-0.10	52.46
6	0.79	1.60	-0.85	-1.21	-1.34	-1.41	-1.57	-0.10	26.61
7	0.75	3.38	-0.86	-0.99	-1.54	-1.15	-1.79	-0.10	37.31
8	0.90	0.35	-1.91	-0.57	-0.89	-0.67	-1.04	-0.10	37.17
9	0.78	0.68	2.17	-10.20	-8.38	-11.90	-9.77	-0.10	19.71
10	1.00	0.19	0.95	-7.20	-4.99	-8.40	-5.82	-0.10	16.64
11	1.02	-1.68	2.08	-7.31	-6.32	-8.53	-7.38	-0.10	20.77
12	0.89	2.18	-3.32	1.85	0.69	2.16	0.81	-0.10	8.96
13	0.92	1.61	-1.70	-2.28	-2.35	-2.66	-2.74	-0.10	24.76

11.4. Impact of Automated Trucking on Trade Flow across U.S. and Texas

Figure 11.3's RUBMRIO equations were used to estimate U.S. trade flows between the nation's 3109 contiguous counties, as well as to 117 FAF[4] export zones, across 13 industries and 3 travel modes. $8.3 trillion in trade flows were generated to meet the year 2015 export demand of $1.04 trillion, as obtained from FAF[4] (with 24%, 18%, 17%, and 16% of those exports headed to Canada, Mexico, Europe and East Asia, respectively). The model's total flow predictions account for 91.3% of FAF[4]'s total $15.0 trillion trade flow. It is not 100% because the nation has another $2.5 trillion in import flows (according to FAF[4], coming from other countries), which are not tracked here.

The base-case scenario assumes travel costs of $1.85 per Htruck-mile based on $1.38 in 2013 (Truck report, 2013) and railcar costs of $0.6 per container-mile (with different commodities filling containers differently, in terms of dollars per container). Table 11.3 compares RUBMRIO trade flow results to those in the FAF[4] database, after aggregating the model's 3109 trade zones into the nation's 129 FAF zones and counting the number of OD pairs that deliver the first 10 percent of trade flows (in dollar terms, rather than ton-miles or dollar-miles, for example), then the next set of OD pairs, and so forth (summing to 129 x 129 [domestic flows] zones pairs or 129 x 117 [export flows] zone pairs each). For example, the model's smallest-value domestic shipments come from 13,896 FAF-zone pairs, for $0.85 trillion, or the first 10% of the total ($8.5 trillion) in domestic flows. FAF[4]-based values (for highly aggregate regions/zones) suggest something similar: over 12,000 FAF-zone pairs are involved in that first 10% (smallest-shipment-size) set of flows.

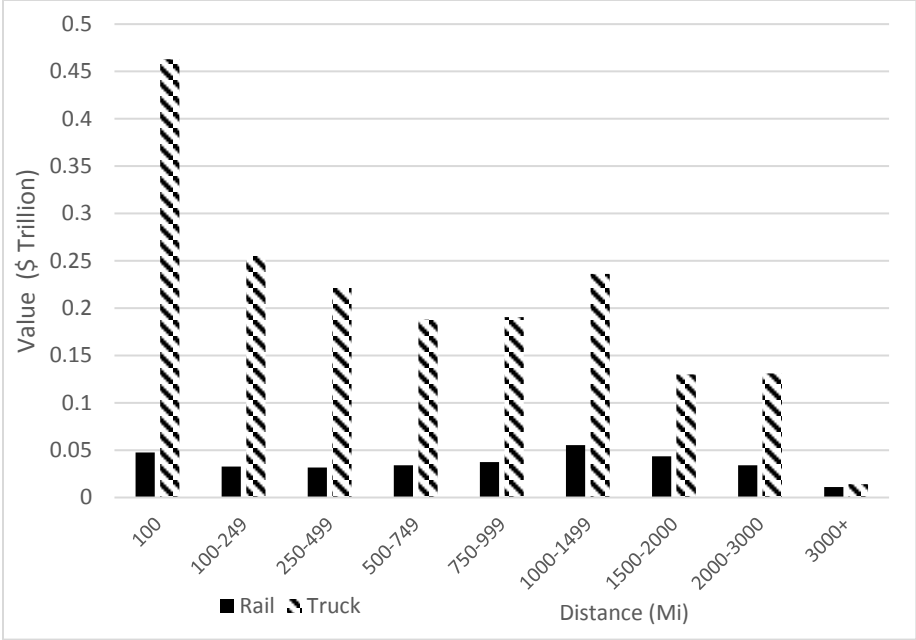
Table 11.3's comparison suggests that the base case RUBMRIO model equations and assumptions deliver reasonable trade-flow estimates, of FAF[4] volumes. However, RUBMRIO tends to "spread out" the trades across more OD pairs (with fewer small-size shipments) than FAF[4] data suggest.

In other words, RUBMRIO predictions suggest less concentration of trade dollars or shipment sizes in the biggest OD trading patterns, for both domestic and export flows. There is obviously much more to U.S. trade than an origin's population and its relative location on railways and highways, versus competing shipment sources. It is interesting how close RUBMRIO can come to replicating many trade patterns with a concise and transparent set of equations (Figure 11.3 plus Equations 11.10 and 11.11).

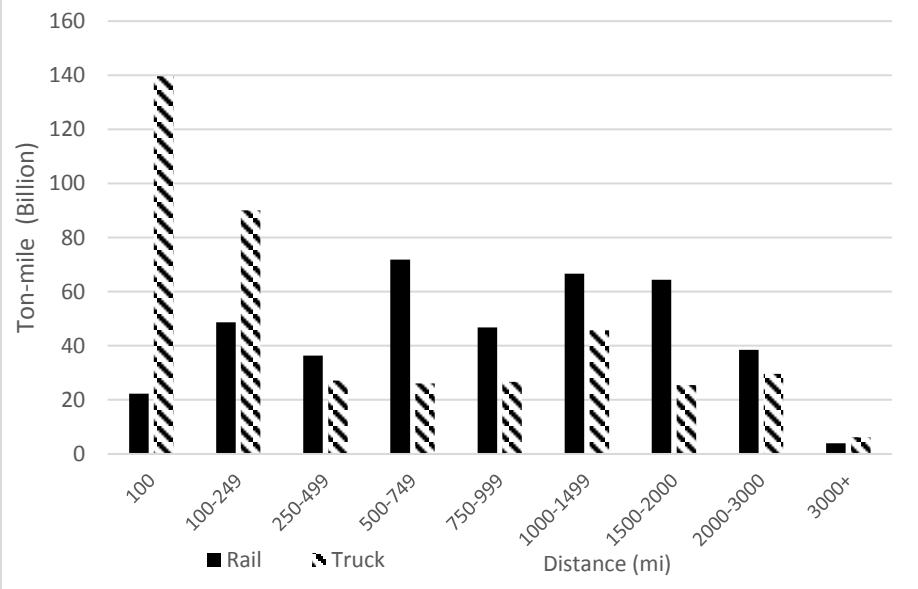
Table 11.3 Cumulative distribution of RUBMRIO and FAF4 trade flows

	Domestic Flows		Export Flows	
	RUBMRIO	FAF[4]	RUBMRIO	FAF[4]
0%-10%	13,896	12,646	14,217	13,971
10%-20%	1,354	2064	617	552
20%-30%	621	935	267	257
30%-40%	324	479	149	146
40%-50%	183	262	97	81
50%-60%	118	134	65	40
60%-70%	82	64	37	26
70%-80%	49	36	19	14
80%-90%	12	16	9	4
90%-100%	2	5	3	2

Figure 11.4 shows RUBMBRIO's base case trip distribution by trade values and ton-miles and appears reasonable compared to FAF statistics (Strocko et al., 2014). However, truck trade-value flows are much greater than rail's values across all distances. In ton-mile trading, truck dominates among lower-distance flows, while rail dominates at longer distances.



(a) Trade flow distribution in value before Atrucks Implementation

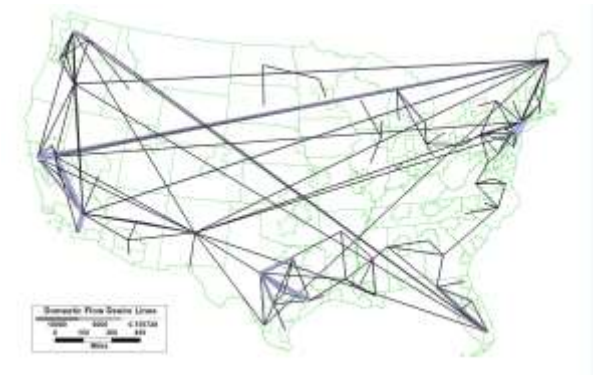


(b) Trade flow distribution in ton-mile before Atrucks Implementation

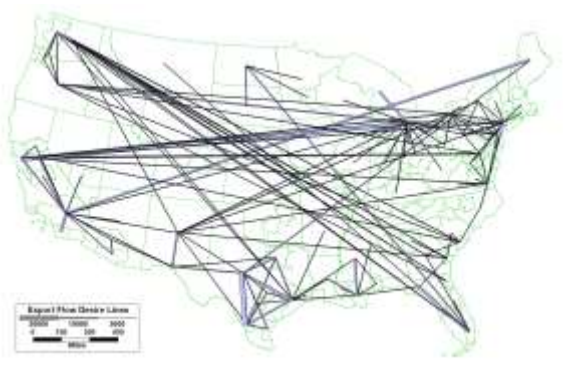
Figure 11.4 Trade distributions (by $ value and ton-miles) for base case (business as usual) scenario

For a spatial perspective of these results, Figure 11.5 shows domestic trade flows and export trade flows pattern, without showing lines for value less than 5%. Many major domestic flows exist between western states, like California and Washington, to various eastern regions/FAF zones. In some contrast, major export flows (within the continental U.S., to access a port) also exist between coastal cities and their adjacent regions (often adjacent states). Moreover, exports from California ports appear to come largely from the Great Lakes region instead of from the Eastern Seaboard, thanks to a heavy export of Michigan-manufactured automobiles and trucks. Truck flows show

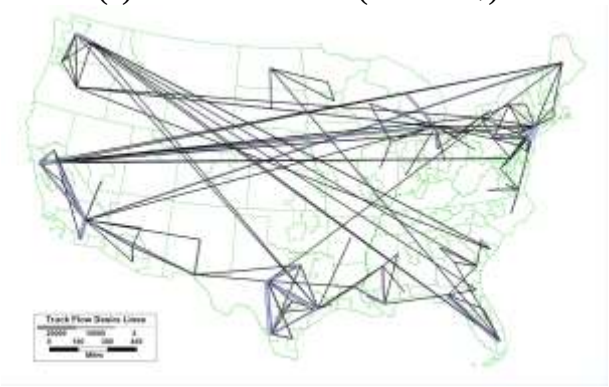
more intra-state trips with shortest distances, like trips within Texas, Florida and New York, while more longer rail flows tend to cross the nation.



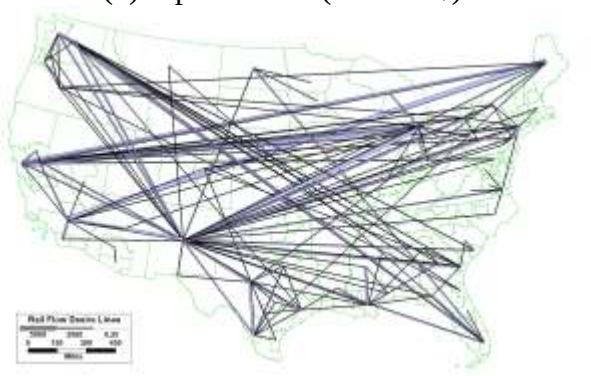
(a) Domestic Flows (Million $)



(b)Export Flows (Million $)



(c) Truck Flows (Million $)



(d) Rail Flows (Million $)

Figure 11.5 Base case domestic and export trade flows (per year), between FAF4 zones

11.4.1. Sensitivity Analysis

Since great uncertainty still exists about the relative costs of acquiring and deploying Atrucks, multiple scenarios were tested here, with different parameter assumptions. Atruck operating costs are expected to be much lower than Htruck costs, overall, thanks to a reduction in operator/attendant burden from the driving task and Atrucks' greater utilization, as their attendants rest/sleep or perform other duties (and are not subject to strict hours of service regulations, since they cannot cause a fatal crash, for example). Wages and benefits may fall, or simply shift from administrative and service workers that used to be officed (e.g., those managing carrier logistics, customer service calls, or shipper billing) to workers that now travel between states on-board a moving office (and help with pick-ups and deliveries, as those arise).

Scenario 1 serves as a reference, high-technology (Atrucks in operation) case for the following discussion of nine different Atruck scenarios. Base case is the mode share before Atrucks implementation. After the introduction of Atrucks, the mode share of trucks increases compared to rail, but the total ton-mile and dollar mile decreases. Compared to Scenarios 1 through 3, the cost of Htruck use is assumed to be 20% higher (in Scenarios 4 through 6) or lower (Scenarios 7

through 9), while Atruck costs are assumed to be 75%, 50%, and 25% of Htruck costs (per ton-mile, container-mile or commodity-mile), respectively, resulting in 9 (3 x 3) separate scenarios. Table 11.4 presents basic mode split results for FAF[4] and these 9 scenarios. Interestingly, Atruck splits (either by dollar-miles carried or ton-miles transported) are very stable across the 9 scenarios, regardless of the relative price variation.

Sensitivity analysis is also applied for Atruck ASCs and scaling parameters for the nested logit model. With slight changes, the more attractive that one makes Atrucks, relative to Htrucks, the more dollar-miles and ton-miles will be carried by trucks. For the test of scaling parameter, if increased substitution is assumed between alternatives in the truck nest or the mode nest, the truck split will increase.

Table 11.4 Sensitivity analysis

(a) Operation Cost Test Results

Scenario	Cost of Htruck	Cost of Atruck	$ Trillion				Billion dollar-miles				Billion ton-miles			
			Rail	%	Truck	%	Rail	%	Truck	%	Rail	%	Truck	%
Base	-	-	0.33	15.3	1.83	84.7	631	43.5	820	56.5	399	49.0	416	51.0
1*	100%	75%	0.21	9.6	1.95	90.4	417	28.4	1,051	71.6	371	44.9	455	55.1
2	100%	50%	0.24	11.2	1.91	88.8	505	33.7	995	66.3	380	45.2	461	54.8
3	100%	25%	0.22	10.4	1.91	89.6	432	27.9	1,114	72.1	374	43.1	494	56.9
4	80%	75%	0.24	10.9	1.92	89.1	494	33.0	1,003	67.0	383	43.8	493	56.2
5	80%	50%	0.25	11.5	1.90	88.5	518	33.6	1,022	66.4	387	43.2	509	56.8
6	80%	25%	0.22	10.1	1.92	89.9	425	26.9	1,154	73.1	379	41.1	543	58.9
7	120%	75%	0.26	11.9	1.90	88.1	595	41.2	848	58.8	384	48.8	402	51.2
8	120%	50%	0.23	10.9	1.91	89.1	459	30.2	1,059	69.8	373	45.0	455	55.0
9	120%	25%	0.23	10.9	1.91	89.1	489	29.7	1,159	70.3	393	44.7	485	55.3

(b) Atruck ASCs Test

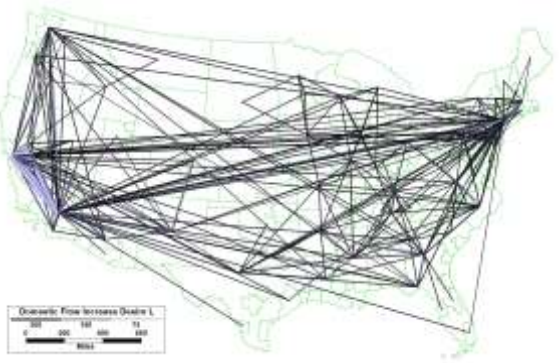
Scenario	ASC for Atruck	$ Trillion				Billion Dollar-miles				Billion Ton- miles			
		Rail	%	Truck	%	Rail	%	Truck	%	Rail	%	Truck	%
1*	-0.1	0.24	11.2	1.91	88.8	505	33.7	995	66.3	380	45.2	461	54.8
2	-0.3	0.24	11.4	1.91	88.6	505	33.7	994	66.3	380	45.2	461	54.8
3	0.1	0.24	11.3	1.91	88.7	505	33.7	995	66.3	380	45.1	462	54.9

(c) Scaling Parameters Test

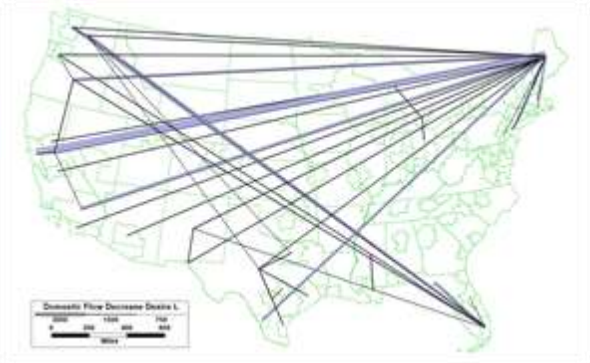
Scenario	$\theta^m_{ij,mode}$	$\theta^m_{ij,truck}$	$ Trillion				Billion Dollar-miles				Billion Ton-miles			
			Rail	%	Truck	%	Rail	%	Truck	%	Rail	%	Truck	%
1*	1.2	1.4	0.24	11.2	1.91	88.8	505	33.7	995	66.3	380	45.2	461	54.8
2	1.2	1.3	0.21	9.9	1.92	90.1	426	26.4	1,187	73.6	385	39.0	603	61.0
3	1.1	1.4	0.22	10.3	1.92	89.7	459	29.8	1,081	70.2	379	41.5	535	58.5

Figure 11.6 illustrates estimated changes in flow patterns for trucks and railroads before and after the introduction of Atrucks (where truck flows are the sum of Atruck and Htruck flows), with spider maps of rising versus falling flows shown separately. The measurement scale is adjusted to reflect only major flow values (million dollars between OD pairs greater than 5% of total flow value) since much more value is carried by truck [than by rail] in the U.S. and for domestic [rather than export] purposes). Results suggest that increases in domestic flow types occur most heavily along the nation's western coast (through California) and between California and New York. Export flows have their greatest increases between the Great Lakes region (including Michigan and Illinois) and California. Both domestic and export flows are estimated to fall from trucking automation options along the nation's northeastern areas and between Florida and Washington.

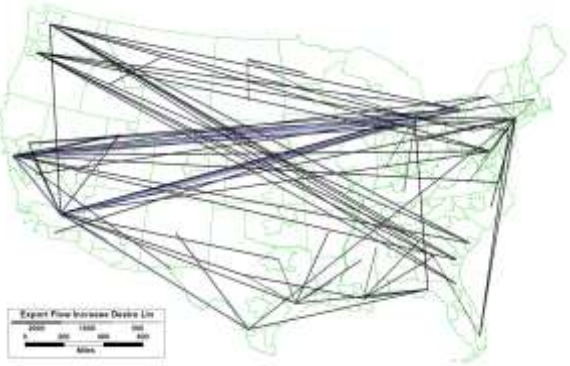
As shown in Figure 11.6, truck flows are also predicted to lose many interactions between the western U.S. and Florida and northeastern states, while experiencing greater interactions between Northwestern (Washington and Oregon) and Eastern (Georgia and South Carolina), and also between the Great Lakes region (including Michigan and Illinois) and California. This is probably due to Atrucks being better able to meet freight demand in Florida and northeastern areas by obtaining more inputs from the nation's northwestern areas. Rail flows are estimated to rise only in and around New Mexico, while noticeably elsewhere (e.g., in Texas and from San Francisco and Arizona to the Great Lakes and northeastern areas, respectively).



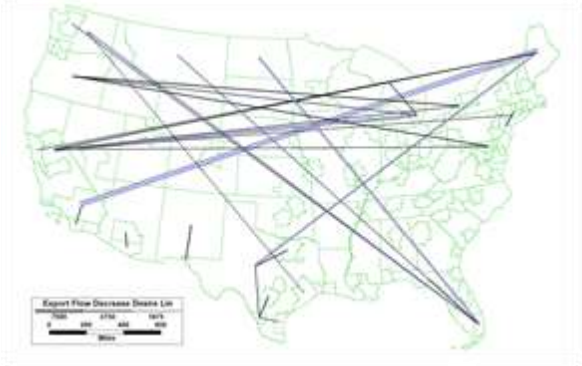
(a) Increase in Domestic Flow (Million $)



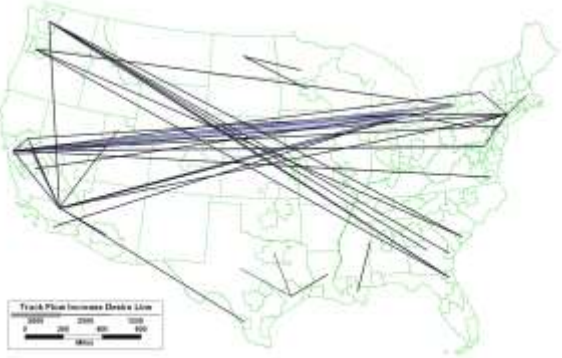
(b) Decrease in Domestic Flow (Million $)



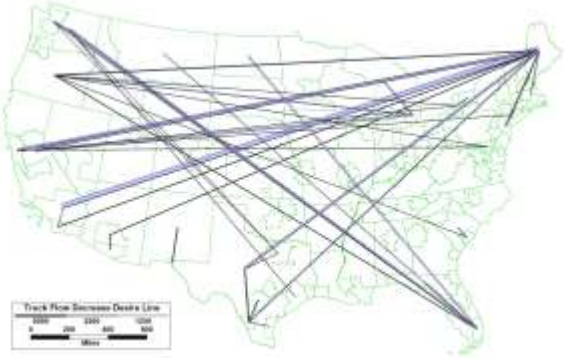
(c) Increase in Export Flow (Million $)



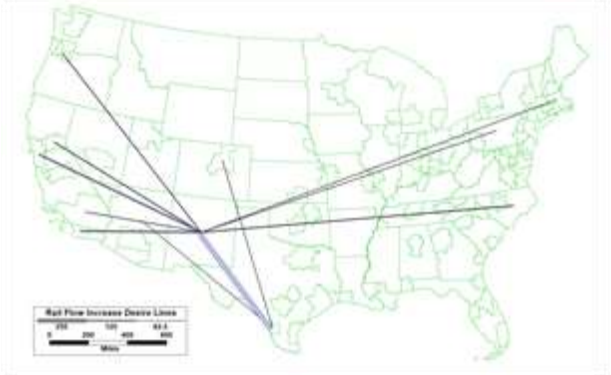
(d) Decrease in Export Flow (Million $)



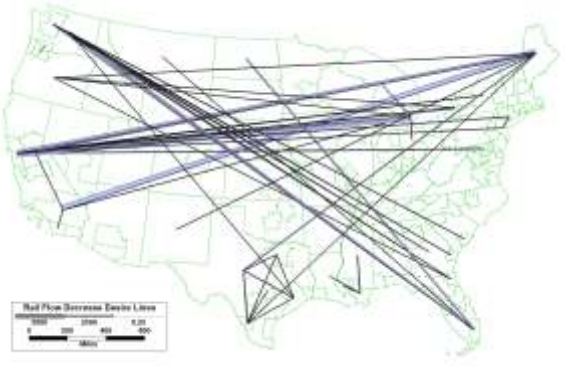
(e) Increase in Truck Flow (Million $)



(f) Decrease in Truck Flow (Million $)



(g) Increase in Rail Flow (Million $)



(f) Decrease in Rail Flow (Million $)

Figure 11.6 Principal U.S. trade flow patterns before and after Atrucks ($ million per year)

11.4.2. Analysis of Major Cities Trade Flow

Table 11.5 shows estimates of flow changes across major U.S. cities. Most (like Sacramento, Washington DC, Indianapolis, and Nashville) experience increases in trucking flows, both into and out of the city. However, Miami, Detroit, Salt Lake City, and Houston are estimated to experience roughly a 10% decrease in their current outbound truck, alongside increases in their pass-through trucking volumes (due to the travel-cost benefits that automation brings the trucking mode). All major cities are predicted to see lower rail flows (inbound and outbound), with San Jose CA and Washington DC experiencing more than 70% reductions in outbound rail flows, and a similar situation happens for rail flows into Jacksonville FL and Washington DC.

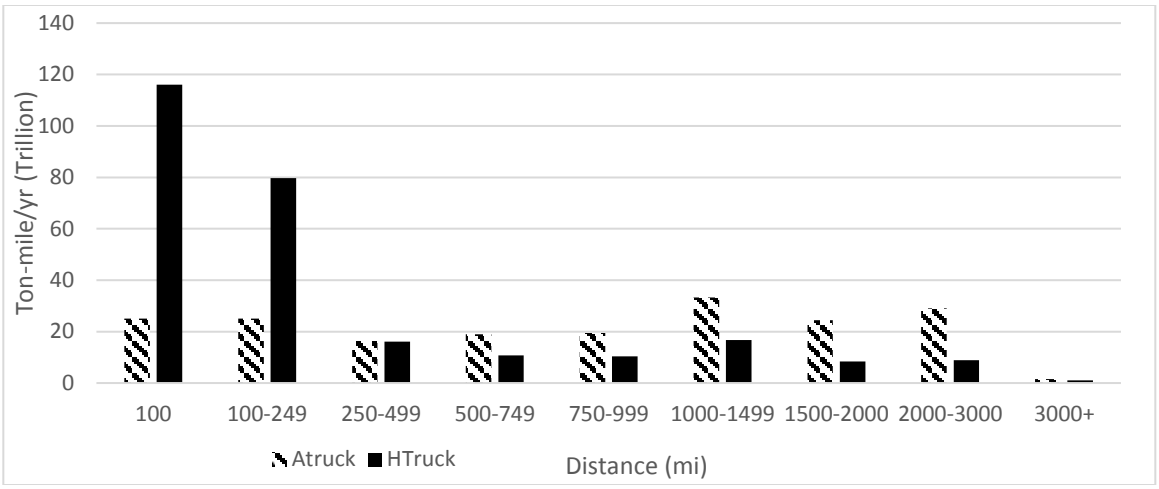
Table 11.5 Automated trucking's impact on trade flows originating from or destined for major U.S. cities

State	City	Truck Flow (change in $)		Rail Flow (change in $)	
		Out	In	Out	In
AZ	Phoenix	0%	-3%	-35%	-42%
CA	Los Angeles	4%	-1%	-37%	-45%
CA	Sacramento	22%	15%	-40%	-35%
CA	San Diego	10%	5%	-25%	-26%
CA	San Jose	19%	2%	-72%	-42%
CO	Denver	14%	9%	-6%	-15%
DC	Washington	38%	34%	-77%	-74%
FL	Miami	-21%	-3%	-67%	-53%
FL	Orlando	5%	5%	-43%	-39%
FL	Jacksonville	5%	19%	-44%	-73%
GA	Atlanta	11%	10%	-40%	-44%
IL	Chicago	7%	5%	-46%	-41%
IN	Indianapolis	18%	16%	-42%	-34%
KY	Louisville	15%	9%	-40%	-49%
MA	Boston	5%	10%	-48%	-38%
MD	Baltimore	8%	9%	-41%	-52%
MI	Detroit	-12%	6%	-43%	-50%
MN	Minneapolis	17%	13%	-44%	-36%
MO	Kansas City	17%	17%	-50%	-42%
NC	Charlotte	14%	13%	-42%	-36%
NJ	New York	1%	4%	-39%	-37%
NJ	Philadelphia	8%	9%	-40%	-34%
NV	Las Vegas	8%	4%	-34%	-39%
OH	Columbus	14%	13%	-41%	-34%
OK	Oklahoma City	12%	9%	-43%	-39%
OR	Portland	17%	4%	-53%	-39%
TN	Memphis	16%	7%	-45%	-50%
TN	Nashville	22%	19%	-41%	-34%
TX	Austin	0%	-7%	-39%	-38%
TX	Dallas	-2%	-3%	-41%	-41%
TX	Houston	-11%	-1%	-42%	-44%
TX	San Antonio	-6%	-8%	-40%	-41%
TX	El Paso	9%	5%	-44%	-41%
UT	Salt Lake City	-11%	-1%	-46%	-50%
WA	Seattle	3%	-4%	-52%	-39%

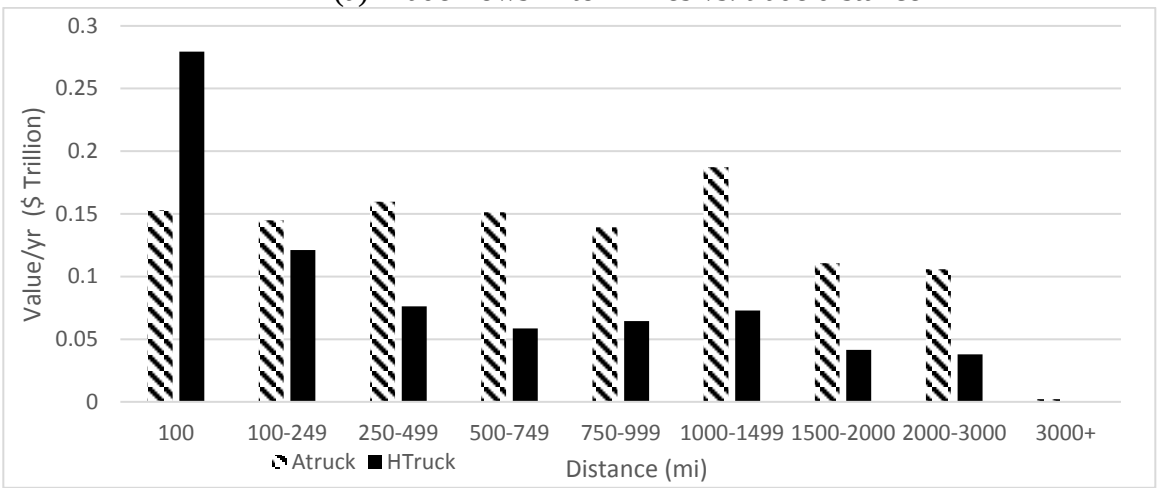
11.4.3. Trip length Distribution

Trip-length distributions are another meaningful way to view Atrucks' effects on travel patterns. Figure 11.7 shows such distributions for total rail shipments, total truck shipments, and Atruck versus Htruck shipments. Figures 11.7(a) and 11.7(b) illustrate mode splits between Atrucks and Htrucks, across domestic trade-flow distances. Htrucks appear to still dominate up to about 250 miles of distance, while Atrucks appear to clearly dominate after about 500 miles of travel distance. Htruck flows fall as distance increases, while Atruck flows are quite robust across all distances. Atruck trade volumes appear to peak at 1000 to 1500 miles, which is approximately the distance from Seattle, Washington to Los Angeles, California, or from Dallas, Texas to San Francisco, or from New York to Miami. These are major OD pairs for many commodities (like finance, insurance, and service goods).

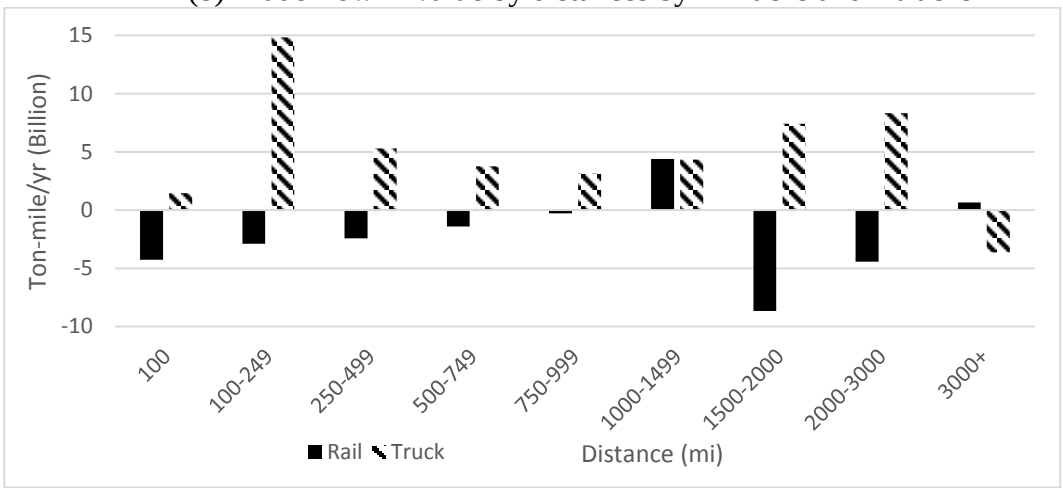
Figures 11.7(c) and 11.7(d) show how ton-mile truck flows are predicted to rise for all trip distances, excepting those over 3,000 miles. Trade increments by truck peaks at 100-249 miles, indicating that trade flows are also predicted to transport more within counties. It is interesting to see that the trade value decreases for both truck and rail at smaller distance, showing that trade flows are moving towards longer distances. Rail flow values appear to drop at distances up to 3,000 mi, with a slight increase for very long rail distances—over 3,000 miles. This is likely because Atrucks are quite competitive for mid- and long-distance trade. However, when input access distances exceed 3000 miles, railway's lower costs prove very competitive, for many commodities (e.g., those that are less time-sensitive, low value per ton, and/or perishable). There is also a 6.6% increase of rail flow of ton-mile at 1,000 to 1,499 miles. This is probably due to the specific demand of a certain commodity for some interstate OD pairs.



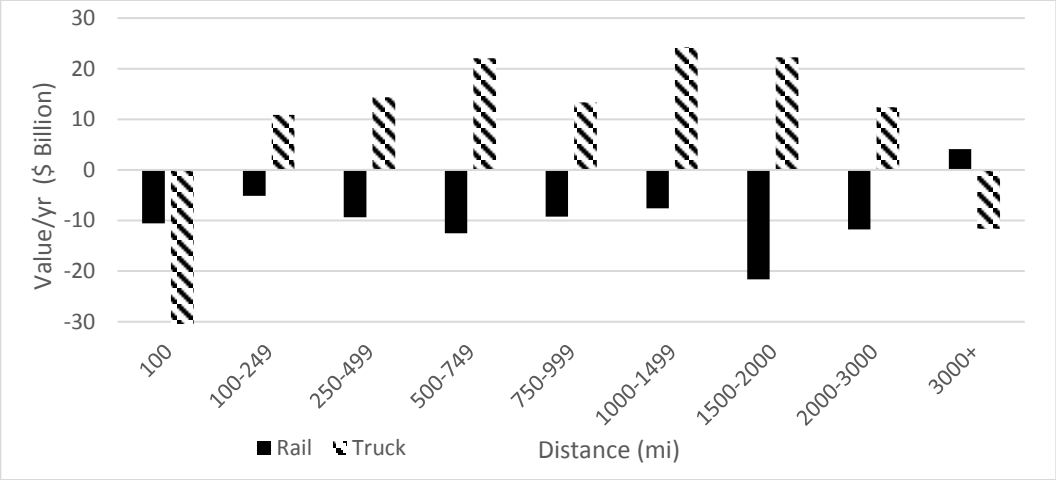
(a) Trade flows in ton-miles vs. trade distance



(b) Trade flow in value by distances by HTrucks and Atrucks



(c) Trade flow change in ton-mile by distances before and after Atrucks



(d) Trade flow change in value by distances before & after Atrucks

Figure 11.7 Trip length distributions for U.S. rail and trucks flows, before and after Atrucks

11.4.4. Commodity Analysis

Table 11.6 shows commodity flow changes by mode, following the introduction of Atrucks, under the Base Case vs. reference Scenario 2. Introduction of automated trucking or "Atrucks" is expected to increase both total domestic flows and total export ton-mile and value flows, by 2% to 4% respectively. Domestic truck flows (in ton-miles) are forecast to rise 11% (versus a BAU/no-new-technology scenario) and rail flow values fall by 24%. Transportation equipment manufacturing and durable and non-durable manufacturing trade flows (between U.S. counties) are predicted to fall, while construction, food, beverage, tobacco products, primary and fabricated metal manufacturing are all predicted to see a small increase in their trade flows, as a result of automated trucking. Agriculture, forestry, fishing, hunting, chemicals, plastics, petroleum and coal products show some of the biggest relative increases (greater than 10%), presumably because Atrucks making trucking relatively more useful in these domains due to its' advantages of safety and time efficiency. As expected, railway becomes a *relatively* less effective or efficient way to transport such commodities. Ten sectors see a decrease in total (domestic) value shipped by rail while only three sectors are predicted to rise. Although machinery manufacturing, computers, other electronic products and electrical equipment manufacturing transported by rail rise by more than 500% following automated trucking's introduction, this increment is still much less than the increases transported by truck.

Finally, export truck flows are estimated to rise, from range of 5% to 47%, excepting only durable and non-durable manufacturing trades, which are forecast to shift almost all to rail. Total rail flows of 328 billion ton-miles/year headed for U.S. export zones remains stable, while total truck flows are expected to rise by 11%. Total ton-miles (sum of Truck and Rail or sum of Domestic and Export) increase by 3.1%. As readers can see, RUBMRIO's system of trading equations (Figure 11.3) deliver a wide array of meaningful predictions, the complexity of which would not be quantifiable without such programs.

Table 11.6 Change in U.S. trade flow ton-miles before and after Atrucks

Million ton-miles	Domestic Truck			Domestic Rail			Truck			Domestic		
Sector	Before	After	%	Before	After	%	Before	After	%	Before	After	%
1	4,103	5,004	22	7	3	-54	4,203	5,126	22	4,110	5,007	22
2	64,544	76,257	18	14,530	10,442	-28	71,482	84,572	18	79,075	86,699	10
3	149,723	155,453	4	32,655	30,037	-8	156,662	162,741	4	182,379	185,490	2
4	3,382	3,956	17	1,944	1,518	-22	35,715	42,644	19	5,326	5,474	3
5	3,273	4,243	30	554	330	-40	9,170	11,937	30	3,827	4,573	19
6	6,423	8,013	25	1,583	987	-38	18,189	23,070	27	8,006	9,000	12
7	5,511	6,228	13	1,618	1,298	-20	8,157	9,255	13	7,129	7,526	6
8	39,130	50,775	30	10,716	1,006	-91	47,617	61,961	30	49,846	51,781	4
9	2,980	3,825	28	7	47	582	5,403	7,103	31	2,986	3,872	30
10	2,372	2,855	20	15	91	512	6,770	8,454	25	2,387	2,946	23
11	7,581	3,457	-54	3,392	5,630	66	30,145	36,587	21	10,973	9,087	-17
12	203	0.01	-100	425	183	-57	16,701	0.02	-100	628	183	-71
13	1,926	2,346	22	94	75	-19	6,470	8,088	25	2,019	2,422	20
SUM	291,150	322,412	11	67,540	51,647	-24	416,683	461,539	11	358,691	374,059	4
Million ton-miles	Export Truck			Export Rail			Rail			Export		
Sector	Before	After	%	Before	After	%	Before	After	%	Before	After	%
1	100	122	22	0.18	0.08	-55	7	3	-54	100	122	22
2	6,937	8,316	20	1,739	1,257	-28	16,269	11,700	-28	8,676	9,573	10
3	6,939	7,288	5	1,745	1,619	-7	34,400	31,656	-8	8,684	8,907	3
4	32,333	38,688	20	18,153	14,542	-20	20,097	16,060	-20	50,486	53,230	5
5	5,897	7,695	30	1,013	607	-40	1,567	937	-40	6,910	8,302	20
6	11,766	15,058	28	3,029	1,769	-42	4,613	2,757	-40	14,796	16,827	14
7	2,645	3,027	14	807	646	-20	2,425	1,943	-20	3,453	3,672	6
8	8,488	11,186	32	2,396	163	-93	13,113	1,170	-91	10,884	11,350	4
9	2,424	3,278	35	4.72	0.61	-87	12	47	309	2,429	3,279	35
10	4,398	5,599	27	29	0.46	-98	44	92	110	4,427	5,599	26
11	22,563	33,129	47	17,816	6,256	-65	21,208	11,886	-44	40,379	39,385	-2
12	16,498	0.01	-100	284,834	301,447	6	285,259	301,629	6	301,332	301,447	0
13	4,544	5,742	26	226	96	-58	319	171	-46	4,769	5,838	22
SUM	125,533	139,127	11	331,793	328,404	-1	399,333	380,051	-5	457,326	467,531	2

11.5. Summary

In this chapter, the RUBMIO trade model is used to anticipate the shifts in U.S. trade patterns due to the introduction of Atrucks. Lower-cost trucking operations will impact choice of mode and input origins, affecting production and flow decisions for domestic and export trades across states, nations, and continents. Here, 13 commodity types were tracked using the 2012 CFS and FAF[4]

data sets. Sensitivity analysis allows for variations in predictions, given the great uncertainty that accompanies shippers' future cost-assessments, adoption rates, and use of Atrucks. Such predictions should prove helpful to counties and regions, buyers and suppliers, investors and carriers, as they prepare for advanced automation in our transportation systems.

This early attempt to reflect self-driving trucks in LD freight systems relies on U.S. highway and railway networks as well as FAF[4] trade data. Extensions of this work may wish to reflect other modes, like airlines, waterways, and pipelines, as well as multi-modal and inter-modal flows, local supply-chains, urban logistics, and local production capabilities and port capacities. In terms of the RUBMRIO model's specification, reflecting the dynamic evolution of population and employment patterns (as in Huang and Kockelman [2010]), commuting and shopping trips, with intra-regional and inter-regional congestion, as well as seasonal variations in certain shipments (like agriculture and coal) may prove very helpful. Further extensions on random utility models employed here can come through different nesting structures, as well as operator awake hours, routing, and delivery scheduling.

Chapter 12. Agent-Based Population from Traditional Four-Step Data

12.1. Need for Person-Level Data

Observation of travel patterns is evolving in many aspects including the new operational strategies using existing transportation tools (e.g., vehicle/ride sharing) and upcoming transportation innovations (e.g., connected and autonomous vehicles). The emerging travel patterns may require advanced modeling techniques for traffic forecasting and evaluations of transportation policies and projects.

Car-sharing is transforming the way people travel, live, and socialize (Cohen et al., 2016). Advanced communication technologies including the internet and smartphones provide a platform that allows individuals to be part of car-sharing, such as booking a car online at home or requesting a ride on the roadside. Including Uber, Car2go, Lyft, Zipcar, Hertz, and Enterprise, there were more than 35 major car-sharing industrial participants/competitors in North America that managed or operated more than 25 thousand shared vehicles in July 2015 (Martin and Shaheen, 2016). As reported, until 2015 the global car-sharing market size was over $126.1 million. A 34.8% compound annual growth rate is expected between 2016 and 2024; the car-sharing market size is projected to exceed $16.5 billion by 2024 (Global Market Insights, 2017). Car-sharing offers mobility to travelers without the burden of owning a vehicle and the car-sharing services are more flexible than transit (Liu and Kockelman, 2017). In addition, shared vehicles require fewer parking spaces (both on-street and off-street), as these vehicles only need a spot to pick up or drop off customers. Car-sharing may help lower traffic congestion and improve air quality, as car-sharing users are generally unlikely to own or buy a car (Martin and Shaheen, 2016).

Emerging transportation tools such as connected and autonomous vehicles (CAVs) will further facilitate the growth of the car-sharing market. Existing car-sharing services either require a driver in the vehicle to pick up/drop off customers (e.g., Uber) or need the customer to make a trip to access the service at car-sharing stations (e.g., Car2go). CAVs can drive themselves to pick-up/drop-off locations requested by customers.

CAVs are expected to significantly improve surface transportation systems from three aspects: safety (Kockelman and Li, 2016), mobility (Fagnant and Kockelman, 2015a; Chen and Kockelman, 2016a; Fagnant and Kockelman, 2016), and sustainability (Bansal et al., 2015; Reiter and Kockelman, 2017). CAVs will improve conditions for safe travelling and decrease crash frequency and severity (Rau et al., 2015; Schoettle and Sivak, 2015; Kockelman and Li, 2016). CAVs are expected to reduce travel times and costs, mitigate traffic congestion, and offer a more convenient long-distance transport option for disabled travelers or those who are too young or old to drive (Anderson et al., 2014; Fagnant and Kockelman, 2015a; Chen and Kockelman, 2016a; Fagnant and Kockelman, 2016; Chen et al., 2016b; LaMondia et al., 2016). Regarding the sustainability benefits, CAV technologies are anticipated to help reduce energy consumption and emissions. These technologies include vehicle weight-lighting (Greene, 2008; Ford, 2012; Chapin

et al., 2013), fuel efficiencies, alternative fuels (Chapin et al., 2013; Liu et al., 2015; Reiter and Kockelman, 2017), and engine technologies (Paul et al., 2011; Folsom, 2012; Bansal et al., 2015; Reiter and Kockelman, 2017).

Currently, most in-use state and regional travel models are "four-step" trip-based (NCHRP, 2012) and the information captured in these models is often aggregated at the level of traffic analysis zones (TAZs). New travel patterns require the modeling of individual trips (rather than aggregated trips between TAZs) at great spatial and temporal details. For example, the car-sharing system needs a model to capture how a service may connect two individual trips, such as modeling the shared car's travel between the present customer's drop-off location and next one's pick-up location. If two trips are connected in the same TAZ, the four-step travel model is unable to capture such car-sharing patterns. Therefore, people are seeking advanced travel modeling approaches; activity-based modeling (ABM) is considered one of the most promising approaches. As compared with the widely used trip-based travel modeling, the activity-based approach is more sensitive to person-specific behavioral attributes (e.g., age, gender, value of time, and willingness-to-pay), capturing how individuals allocate their time for activities and travel though the day (Castiglione et al., 2015). The ABM approach is tour-based, capturing trips made by the same person during the course of a day and within the same tour. A tour is a chain of trips made by the same person to conduct activities throughout the day and typically a tour starts and ends at the same place. Trip-based models replicate the TAZ-aggregated decisions, only considering trip characteristics (e.g., trip distance, speed, duration and cost, and mode availability), while the activity-based approach simulates individual decisions that account for characteristic of both trips and activities (activity duration, and value of conducting an activity). Therefore, ABM appears to be able to capture car-sharing behaviors and answering questions regarding car-sharing operational strategies (e.g., evaluating car-sharing services or estimating the demand given one proposed car-sharing policies).

The properties of ABM present a challenge to transportation planning practitioners, since the modeling input information must also be at the desegregated personal-level. ABM is a data-hungry approach that requires detailed input information about individuals instead of TAZs in trip-based model. For example, in a trip-based model, the origin-destination (OD) matrix is the key travel demand input in the procedure of traffic assignment; the OD matrix contains the number of trips between TAZs. In ABM, the travel demand is derived from the motivation of performing activities. Travel demand becomes a tour for conducting activities. Every individual has a unique tour (travel demand input in ABM) made up of chained trips and activities. In order to prepare the ABM travel-demand input data, one may think of conducting a comprehensive travel survey that asks every person in a modeling region about their activity diary (key information should include the times, locations and types of activities). However, it sounds financially infeasible.

Previous practices offer great insights in preparing data for ABM. For example, ARC's (Atlanta Regional Commission, 2012) Activity-Based Travel Model created synthetic population and households based on the samples of persons and households in Public Use Microdata Areas (PUMA). The synthetic population and households are balanced to match the PUMA controls at both PUMA level (a collection of Census tracts within counties) and county level. The activity

patterns and trips in ARC model were generated based on the statistical analyses with travel survey data from Columbus, Atlanta, and the San Francisco Bay Area. The activity patterns and trip attributes are associated with the person types and household characteristics. Regarding the locations, ARC model used small TAZs to represent the locations of activities (trip origins and destinations). Therefore, in ARC model, the activities are embedded in zones, and are not assigned to specific locations. In 2015, the Transportation Research Board released a report that synthesizes well-agreed concepts and practices on activity-based travel demand models (Castiglione et al., 2015). Generally, in existing practices, the method of preparing data for ABM may be regarded as the method of "start-from-scratch." This method is to prepare data from the raw data that are related to travel demand, including PUMA, LEHD (Longitudinal Employer-Household Dynamics), land-use data, travel surveys, etc. As a matter of factor, existing trip-based travel models are also built upon such data through a rigorous process of data processing. Many metropolitan planning organizations or transportation planning organizations have developed such trip-based models for their jurisdictions. Trip-based models also have information about population/households and travel trips (by purpose) aggregated at TAZ level. Compared with the raw data, the information in trip-based models is more structured. Further, the data (including both the current- and future-year data) in trip-based models must be approved by officials before transportation practitioners use them for travel demand forecasting. In sum, the existing trip-based travel models use the familiar data sources for model input data; the information in trip-based models is more structured and cleaned; and the information in trip-based models is accepted and approved by local officials who have a good sense of the local situations and future developments.

To this end, the objective of this study is to develop a methodology utilizing existing trip-based models to prepare the disaggregated travel demand data for ABM. Using existing trip-based models rather than "start-from-scratch" presents two advantages: 1) the information in trip-based models is structured, 2) trip-based models often contain data for future years that are accepted and approved by officials. Since trip-based models use the same raw data for inputs as the existing ABM practices, either the method proposed in this study or the "start-from-scratch" would result in the similar outcomes, as along as the data contained in trip-based models are valid.

This study is particularly useful for transportation practitioners who develop and apply trip-based travel models in their jurisdiction since the input data used in this study are commonly available for them. The methodology offers insights in preparing the data for ABM that help simulate and understand the individuals' travel patterns and evaluate the transportation policies/strategies under the environment of shared economy and new travel modes, e.g., shared connected and autonomous vehicles. This study presents an example of using data that are easily accessible by the public. Other data sources, such as transportation's big data platforms like Streetlights (www.streetlightdata.com) and AirSage (www.airsage.com), which (may be private but provide great travel data) can also be helpful in preparing ABM input data.

12.2. Methodological Framework

ABM works at the disaggregated person/household level and provides great spatial and temporal details about the individuals' travel patterns. The activity-based approach is tour-based, modeling individuals' travel tours with a chain of trips and activities. ABM's input data is required to be disaggregated at the person/household level, containing detailed spatial information and a schedule for chained travels and activities throughout a day. This study proposes a methodology of preparing the disaggregated input data for ABM. The input data may be summarized as "4Ws" for each traveler's choices, as shown in Figure 12.1. The core of the framework consists of a series of algorithms that output "4Ws" by inputting the aggregated data at zone level. The framework starts with generating synthetic population and households based on land use and socioeconomic data. The output at this step provides information of "Who," defining travelers individually based on age, gender, employment, car ownership, and household characteristics. The next step is locating of households and employments, the information of "Where", taking advantage of the OpenStreetMap data that contains the layout of buildings in a region. These locations are designated areas for conducting activities. This study assumes that all activities are either household or employment-related. Home activities occur at household locations, while other activities are generally employment-related, though not all other activities are for work. For example, shopping activities are associated with the employment of salespersons, and school activities are linked with the work of teachers. The following two steps together output the information of "What," a chain of daily activities that form a travel tour. Zone level travel demand is converted to person-level travel demand by chaining the trips between zones and assigning locations for trips' origins and destinations (that are also the activity locations). The last step is to prepare the information of "When," a tentative schedule for traveling or performing activities. This schedule is only a tentative timeline for an individual to travel and perform the planned activities. The travel plan may change during the ABM process in order to make the most optimized use of a person's time (e.g., leaving the office early to avoid afternoon traffic congestions).

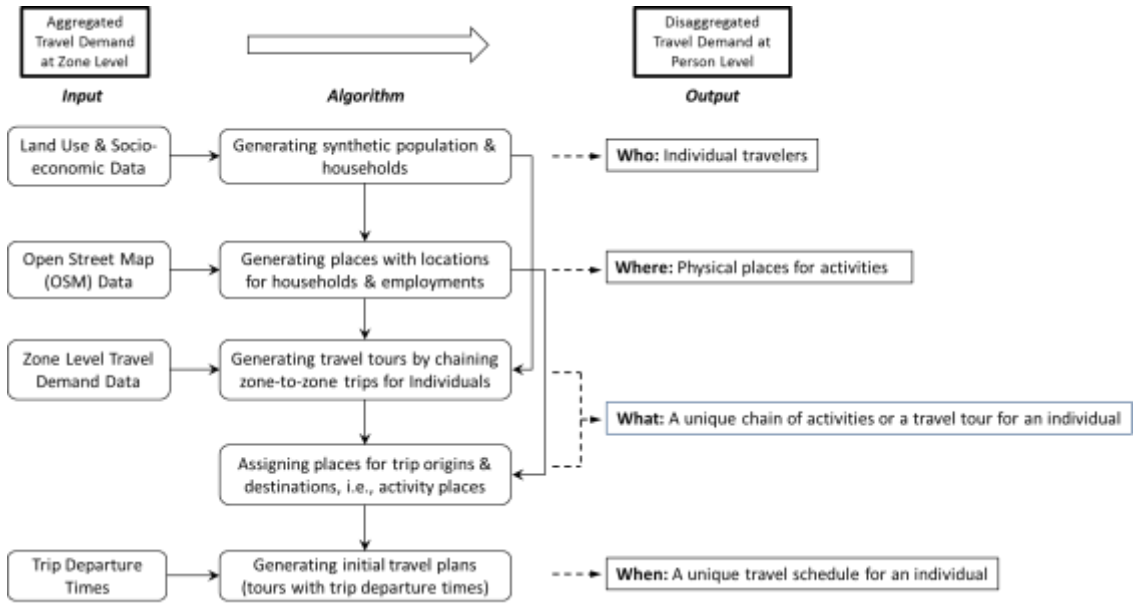


Figure 12.1 Methodological framework of outputting personal level travel demand at person level from zone-level travel demand

12.3. Data Preparation

Three data types were suggested for synthesizing a region's population and generating their travel tours or itineraries: 1) travel demand data from trip-based or four-step travel models, 2) model equations' parameter values, and 3) open-source map data. Table 12.1 lists the specific data sets used here, for method illustration.

12.3.1. Travel Model Data

Travel model data are extracted from Austin's regional travel demand model, created by the Capital Area Metropolitan Planning Organization (CAMPO). The region covers over 5,000 square miles, including Bastrop, Burnet, Caldwell, Hays, Travis, and Williamson Counties in Texas. CAMPO's 2010 Planning Model is a largely traditional four-step macroscopic travel demand model (CAMPO, 2015). This study used data in the model's 2020 scenario, including TAZ land use data and trip tables. The TAZ land use data is important for population synthetization. In synthetic population, every person has an individual profile with their socio-economic information including age, gender, employment, car ownership, household members, household size, and household income. The synthetic population is the basis for generating tour data for individuals. Census data also provide land-use or socio-economic data, as an alternative source. This study used CAMPO model's estimates for 2020. The trip table is also called the origin-destination matrix (OD matrix), offering a big picture of possible trips between/within TAZs (trips are not specified to a specific person in four-step models). Six types of trip purposes (implying a destination's activity type) were considered in the tour generation process: home-based work (HBW), home-based school (HBSc), home-based retail (HBR), home-based other (HBO), non-home-based work (NHBW), and non-home-based other (NHBO) trips. There are five associated activities including

home, work, school, shopping, and other activities. Time skims from CAMPO model represent the average travel time between TAZs. The data is critical for generating initial travel plans which include the duration a traveler may spend in a trip.

Table 12.1 Data sources for preparing AMB inputs

Source	Data	Key information	Data source
Travel model data	TAZ land use data and its shape file	- Population - Household size - Employment - Car ownership - Income level	Regional travel demand models: https://www.campotexas.org/ Alternative sources: Longitudinal Employer-Household Dynamics https://lehd.ces.census.gov/data/ Census Demographic and Economic Data https://www.census.gov/geo/maps-data/data/tiger-data.html
	Trip table (i.e., OD matrix)	- Trip purpose - Number of trips between TAZs	Regional travel demand models: https://www.campotexas.org/
	Time skims	- Travel time between TAZs	Regional travel demand models: https://www.campotexas.org/
Parameter data	Population age distribution	- Age - Percent	Census: https://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml
	Trip departure time distribution	- Trip purpose - Time of day - Percent	Regional travel demand models: https://www.campotexas.org/ Alternative source: NCHRP Report 716 http://www.trb.org/Publications/Blurbs/167055.aspx
	Trip patterns	- Number of trips in a daily tour - Percent	NHTS datasets: http://nhts.ornl.gov/download.shtml
Map data	OpenStreetMap data	- Road network - Building/housing footprint	OpenStreetMap data: http://www.openstreetmap.org/

12.3.2. Parameter Data

Parameters are used to shape the attributes of generated data (e.g., synthetic population and tours). The age distribution parameter is used to control population age structure in a model region. A person's age is related to many travel characteristics, including the number of trips in a daily tour, trip purposes, travel mode (and car ownership), etc. Some assumptions in the tour generation process are related to the traveler's age. For example, a person younger than 5 years old or older than 85 years old is likely to make zero trips in a day; and a person whose age is between 5 and 24 is likely to have a trip to school on a weekday basis. Further, it may be impossible that all members in a household are minors (< 16 years old), and minors are rarely permitted to own a car or drive (alone). The trip pattern parameter informs that how many trips a person may make in one day. Such information is not available in the four-step travel models. Therefore, this study used the data of the 2009 National Household Travel Survey (NHTS). According to NHTS, the average number of daily trips for Texans is 3.76 trips (or 3.78 trips-per-day nationally). Figure 12.2(a) presents the distribution of daily trips per person, with 15.7% of Texans making zero trips on any given day, and 22.6% making exactly two trips in one day.

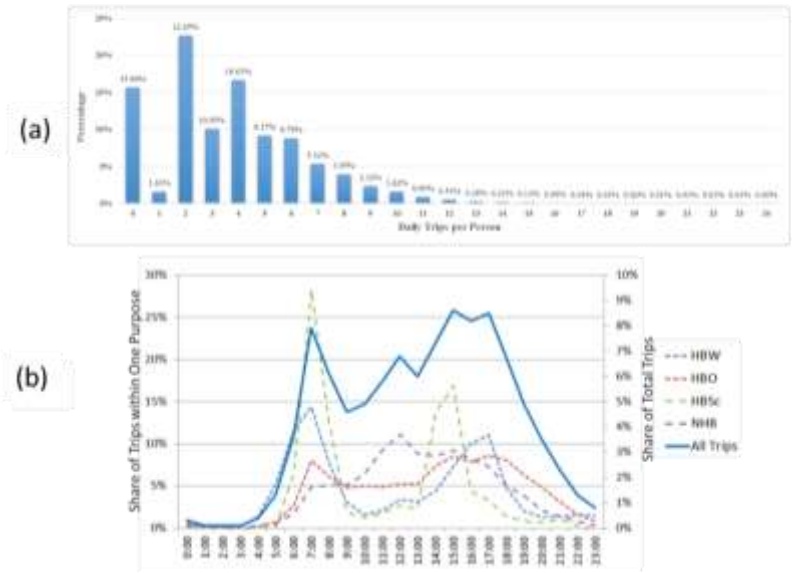


Figure 12.2 Parameter data: (a) trip count in daily travel tours and (b) time-of-day distributions

The trip departure time shows how many trips (in percent) may start at certain times. This parameter is important for observing the time-of-day (TOD) variation of travel demand. Four-step models often take into account four TOD periods including morning peak, afternoon peak, mid-day, and night. The CAMPO model has the hourly TOD factors to simulate temporal variations of travel demand. Figure 12.2(b) shows TOD factors for trip departure times used in CAMPO's model. Four trip purposes are considered in this study: HBW, HBSc, HBO (including HBR), and non-home-based (NHB, including NHBW and NHBO). NCHRP Report 716 is an alternative source for this parameter data (NCHRP, 2012).

12.3.3. Map Data

In trip-based models, location-related information is aggregated at the TAZ centroids. For example, trip generators and attractors are at TAZ centaurs, and a trip starts from or ends at a TAZ centroid. ABM requires the information for specific locations for activities, i.e., origins and destinations. This study used the OpenStreetMap data from www.openstreetmap.org to generate specific locations for individuals and their activities. The data contain the road networks and the house/building footprints. The road networks are composed of nodes and links. The nodes are identified by their IDs, longitudes, and latitudes. The link attributes are identified by link IDs, from and to node IDs. In addition, the links have attributes such as link length, link capacity, free flow speed, number of lanes, and travel mode. Link length can be calculated based on the geo-coordinates of two nodes. Link capacity and free flow speed are determined according to the roadway types indicated in OpenStreetMap. The number of lanes is also available in the data. All public drivable roadways are included in the modeling network. The house/building footprint data provide information about possible locations for performing activities and receiving or starting a trip.

12.4. Program Outputs

12.4.1. Synthetic Population

The program was designed to use the surveyed data and projected demographics used in travel models (summarized at TAZ level) to generate a synthetic population, though the randomness is included in the generation process. The data outputted from the program is supposed to match the statistics of input data at a large extent. Minor differences (<1%) are found between the outputted synthetic population and the inputs (socio-demographic data of CAMPO travel model). The differences are mainly due to the randomness and number rounding. Using the CMAPO's 2020 model inputs, the program generated a synthetic population of 2,325,116 individuals of 895,082 households in the model region. Each individual is generated with age, gender, employment, and car ownership. In addition, individuals are also linked with their household characteristics including household size, household income level, number of employed members, number of vehicles and household locations (longitude and latitude). All these factors are important in activity-based travel modeling process. Figure 12.3(a) and (b) presents the example data of synthetic population at household and person level. From the spatial perspective, the synthetic population is also expected to mirror the aggregated input data. Figure 12.3 also presents (c) the input data of population and households aggregated at TAZ level from the CAMPO's 2020 Travel Model, (d) the spatial distribution of synthetic households, and (e) the density map of synthetic population.

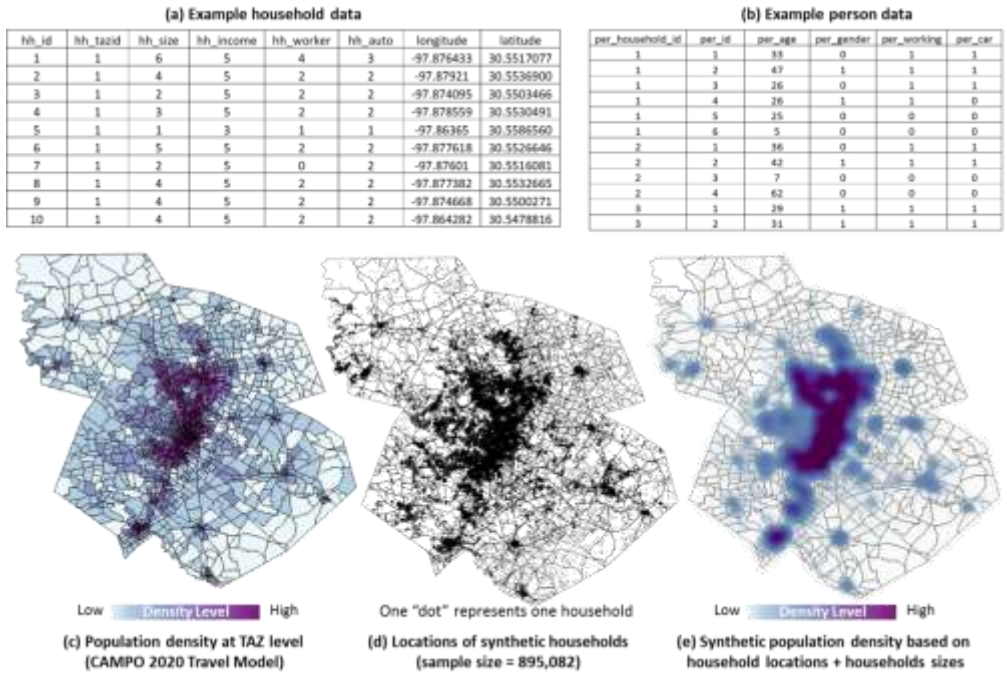


Figure 12.3 Synthetic population and households

12.4.2. Activities and Synthetic Locations

The trip-based travel models offer information about trip purposes which are associated with the activity types at the trip origins and destinations. Five major activity types were generated in the program proposed in this study, including home, work, school, shopping, and other activities. Besides home activities, the program generated about 1.5 million work activities, 0.46 million school activities, 2.5 million shopping activities, and 2.4 million other activities. Synthetic locations are needed to house these activities in the model region. The household locations are for home activities. For the other types of activities, the program generated employment-based locations to house them, though people may not go there for work but for other purposes such as shopping or taking classes. Figure 12.4 presents the example data of generated facilities for activities and the locations for four types of activities. Compared with the household locations (as shown in Figure 12.4), the school and shopping locations are more likely to concentrate to the urban centers; locations for other activities are close to how households are spatially distributed in space.

Example Data of Synthetic Facilities for Employments and Activities

facility_id	Total_ employment	basic_work	shop	other	school	TAZ_id	longitude	latitude
1	94	0	18	6	70	1	-97.86418	30.55943
2	65	0	14	2	49	1	-97.865997	30.55937
3	34	0	26	0	8	1	-97.864968	30.56016
4	11	0	8	0	3	1	-97.86087	30.55812
5	14	0	11	0	3	1	-97.858756	30.54917
6	14	0	11	1	2	1	-97.863279	30.55795
7	20	0	20	0	0	1	-97.859803	30.55556
8	29	0	29	0	0	1	-97.863641	30.54751
9	19	0	19	0	0	1	-97.861061	30.55083

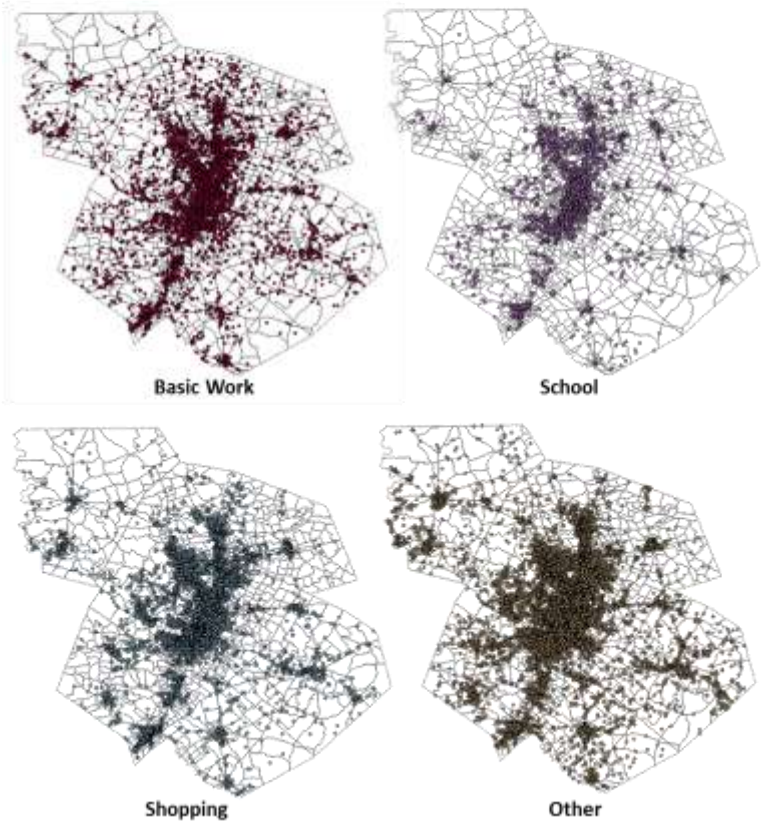


Figure 12.4 Example data of synthetic facilities and spatial distributions of facilities for different types of activities (except home activity)

12.4.3. Trip Chains

Travelers make trips to perform activities. An activity can be regarded as a chain or linkage between two sequential trips, and therefore travelers make a tour to perform a series of activities planned for the day. The program proposed in this study took advantage of the existing CAMPO's travel model (which is trip-based) to generate the daily travel tours for each individual in the model region. The core procedures of tour generation involved chaining the trips between TAZs (estimated in CAMPO's model) to form a tour for an individual, according to this traveler's demographics and NHTS's survey about the daily tour-making patterns (i.e., the number of trips made by a person, as shown in Figure 12.5). The program generated in total 1.96M tours that chain 8.7M trips for 1.96M individuals who actually travel on a daily basis (which leaves 0.36M persons

who do not travel during 24 hours and are assumed staying at home for the whole day). The output resulted in about 3.9 trips per traveler in model region. Figure 12.5 presents the example data of synthetic trip chains, and two example tours in space: a four-trip tour with HBO ⇨ NHBO ⇨ NHBO ⇨ HBO trips, and a five-trip tour with HBW ⇨ NHBO ⇨ NHBW ⇨ NHBO ⇨ HBR trips.

Example Data of Synthetic Tours or Trip Chains

PerId	TripId	activityType	OriginFacilityTAZ	DestinFacilityTAZ	OriginFacilityID	DestinFacilityID
1	0	other	1	1866	1000001	56690
1	1	school	1866	1057	56690	35852
1	2	shop	1057	1595	35852	47825
1	3	home	1595	1	47825	1000001
2	0	work	1	162	1000002	6880
2	1	other	162	929	6880	34081
2	2	work	929	162	34081	6880
2	3	shop	162	118	6880	3937
2	4	home	118	1	3937	1000002

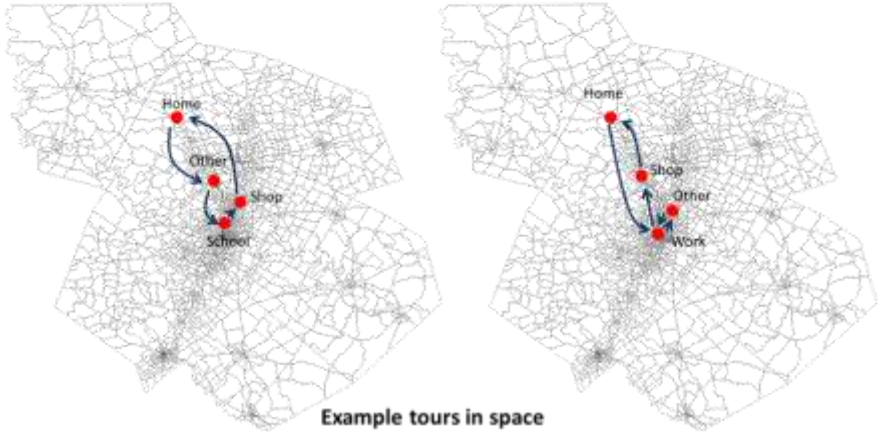


Figure 12.5 Example of synthetic tours or trip chains

12.4.4. Travel Plans

The travel plans provide critical information about when a trip may depart from its origination. The outputted travel plan contains information about the person's age, employment status, and a chain of activities with a tentative schedule. Figure 12.6 shows two example travel plans. The scheduled times were determined by considering the three pieces of time information: 1) activity durations, 2) trip duration, and 3) distributions of trip departure times. The travel plan is the core input of ABM. The travel plan reveals a typical schedule for travel and activities. During the modeling process, the travel plan may be modified given constraints of one-day time and space in roadway network. Late arrival, early departure, or cancelling an activity will cause loss of utility, while being stuck in traffic will also negate the production of values. Therefore, travelers will tend to stick with the schedule but may also adjust the schedule to avoid excessive waste of time on road owing to the traffic congestions.

Figure 12.6 Example travel plans

12.4.5. Spatial Details

The program proposed in this study generates specific physical locations for individuals to perform activities and these locations are the origins and destinations of trips (rather than TAZ centroids in four-step travel models). These locations are scattered in TAZs, as shown in Figure 12.7 (a). There are two types of scatter patterns. One type has quite clear patterns, shown in Figure 12.7(b), along the road links, as these locations are known places for households and employments according to the open-source data. The other type seems to be irregular patterns, shown in Figure 12.7(c). These locations were generated according to the road link/node locations and the number of households and employments in a TAZ. The irregularity is due to the limitations in open source data (e.g., incomplete records) and the need for understanding future travel patterns.

Figure 12.7 Spatial details for activity locations

12.5. Limitations and Summary

The accuracy of synthetic data generated in this study is heavily dependent upon the accuracy of inputs, including the travel demand data, parameter data, and map data. The travel demand model data in future years may contain inaccurate predictions about regional population growth and economic development. The parameter data include the age distributions, tour patterns, and trip departure times. The age distribution parameter may cause inaccuracy in the vehicle ownership assignment and trip-making characteristics (as the kids cannot own a vehicle, and seniors are expected to make fewer trips than young people do). The tour pattern parameter affects the number of trips in a daily travel tour. The inaccurate time parameter in the trip-departure model may not reflect Austinites' actual schedules. In addition, the program presented in this study generates synthetic activity and travel data according to limited data sources with a number of assumptions. The validity of these assumptions remains unknown, and surveys are needed to validate these assumptions in the future. If using a desktop level computer or laptop, the generation of synthetic data using the current program may be a computational burden for large-scale travel model regions (population > 1 M), due to the massive searching cases (e.g., assigning a location for an activity), and matching requirements (the disaggregated synthetic data are required to match the aggregated data at TAZ level from various prospects, e.g., the total population, household, vehicle ownership, employments, etc.). The use of workstation level computers may facilitate the run of the program.

New travel patterns, e.g., car-sharing behavior, present an opportunity and also a challenge for transportation planners and researchers to explore the disaggregated travel demand at person level, in addition to the aggregated demand at zone level which has been well modeled using trip-based

approach. This opportunity allows planners and researchers to confront the new questions regarding the new travel patterns and emerging transportation modes (e.g., autonomous vehicles), while the challenge may hold them back due to difficulty of obtaining disaggregated input data for advanced travel demand modeling at person level. This study offers a methodological framework to prepare input data for ABM, one of the most promising modeling approaches for person-level travel demand. The core of this framework is composed of a series of algorithms that take advantage of publicly available data sources (that are often aggregated at zone level) and produce the disaggregated data at person-level for ABM. The data sources used in this study include land use and socio-economic data, household travel surveys, OpenStreetMap, and regional trip-based models.

This study summarized ABM data into "4Ws" regarding an individual's daily travel: who this person is, where this person lives and works, what daily activities this person does, and when this person plans to perform activities. A program, consisting of a series of algorithms, was designed to generate the data that provide information about the "4Ws". First, the program generated synthetic population based on the zone-level land use and socioeconomic data. Every individual in the modeling region is included in synthetic population; generated attributes include age, gender, employment, car ownership, and household characteristics. Second, places for households and employments were generated to answer where a person lives and works. OpenStreetMap data provide the information about possible locations/places for households and employments. Then the program converted the zone-level travel demand (i.e., trips between zones) to person-level demand (i.e., a unique chain of activities, forming a travel tour which connects specific physical locations instead of zone centroids in trip-based models). The program gave answers to what activities a person does. Last but not least, a schedule for traveling or performing activities was generated by the program to tentatively answer when a person plans to perform activities. Example outputs are shown. The outputs show great temporal and spatial details about individuals' travel patterns.

This study offers both methodological and practical contributions. The framework proposed in this study offers theoretical insights about the "4Ws" as the input components for constructing activity-based travel models and from what public data sources can be used to prepare the "4W" information. This study delivers a practical tool that can help transportation planners and researchers to prepare the "4W" information for ABM. The tool is a computer-based program developed in R environment, composed of a series of algorithms that take advantages of the publicly available data sources and produce person-level information for ABM. This study is particularly useful for transportation planners who already have developed trip-based regional travel models which contain most of the key inputs of the program. Continuing efforts are needed for integrating other data sources, such as transportation's big data, such as Streetlights and AirSage, into this program.

Chapter 13. Potential for Dynamic Ride-Sharing with SAVs: Study with Cellphone Data

13.1. Background and Motivation

Traffic safety and congestion are key transportation issues for many regions around the world. Driver error remains the predominant reason for vehicle crashes (NHTSA, 2015), and rising vehicle-miles traveled (VMT) is worsening traffic congestion (FHWA, 2017). The introduction of autonomous vehicles (AVs) for personal use may dramatically reduce vehicle collisions by eliminating driver error. AVs will also improve mobility options for many travelers, especially those without driver's licenses.

Several transportation network companies (TNCs) offer a dynamic ride-sharing (DRS) option, like uberPOOL and Lyft Line. These services offered by TNCs attempt to match riders with similar trip plans so that the overall cost of travel is minimized for the rider and the operator/driver makes a greater profit. However, a small delay is introduced to the riders' trips as they will have to wait to accommodate other riders needing to be dropped off or picked up. This is often referred to as *ridesplitting* as noted by Shaheen et al. (2016b), but this report will resort to DRS as it is more widely used in the literature. Ride-sharing is not a new concept (Chan and Shaheen, 2012), with carpooling often being feasible for those with common origins and destinations, and stable, similar departure times on both ends of a round-trip (e.g., for many school trips within a neighborhood and for certain work trips). In practice, only casual carpooling or 'slugging' tends to serve real-time demands of flexible departure times (Ma and Wolfson, 2013; Dai, 2016), and is limited to very special corridors (where high toll and time savings induce many drivers to open their doors to different, unknown passengers every day).

Smartphone technology is fundamental to more widespread use of DRS, since it enables real-time access to traveler (and vehicle) locations (Amey et al., 2014). Shaheen et al.'s (2016a) FHWA report notes how important smartphone technology has been in improving travel information access for transit (Transit App), providing shortest paths in real time for many modes (Waze and Google Maps), and increasing carpool-use (Carma). Exploiting this feature, TNCs have designed user-friendly ride-sourcing platforms that interface passengers and drivers, at any time of day and in any region the TNCs serve. By selecting the DRS option, travelers' costs (but not travel times) are lowered, thanks to TNCs working to match two or more travelers with overlapping real-time routes. Such matches add some travel time but deliver significant trip-cost savings and often good conversations among those sharing the ride, who had been strangers (alongside a TNC driver also on board).

AVs will be expensive, at least initially, and not be available for personal ownership for many years (Bansal and Kockelman, 2017). Fleet operators may profitably invest in a fleet of AVs and manage them as TNCs currently manage their (driver-supplied) fleets, but with lower labor costs and complete control of plans and routes. Safer technologies should eventually bring down insurance costs, making shared AVs, or SAVs, more economically viable. In terms of congestion,

SAVs offering DRS can increase average vehicle occupancy (AVO) and reduce regional VMT (Fagnant and Kockelman, 2016; Rodier et al., 2016). It is useful to quantify the level of opportunity for such services, across a range of settings.

This chapter studies the DRS potential for trip-making across the Orlando metropolitan area in Florida, as serviced by a fleet of SAVs. It relies on trip tables derived from cellphone data, as provided by AirSage across a period of 30 consecutive days, to provide a sense of day-to-day trip-making variations. The remaining chapter summarizes related work, describes the AirSage dataset, and then explains the methodology used to match distinct vehicle trips or traveling parties and simulate a fleet of SAVs.

Over the past 10 years, several contributions have been made to optimize and/or implement DRS, with various researchers suggesting that DRS is a key method for reducing future roadway congestion (Levofsky and Greenberg, 2001; Berbeglia et al., 2010; Ma et al., 2013; Farhan and Chen, 2017; Levin et al., 2017). More recently, DRS has been successfully demonstrated using agent-based models (see, e.g., Fagnant and Kockelman, 2016; Bischoff et al., 2016; Loeb et al., 2017; and Hörl, 2017), such as MATsim (Horni et al., 2016) and a synthetically generated dataset of people and journeys to simulate dynamic traffic conditions.

When it comes to actual trip-making, mode choices, and traffic patterns, DRS has been investigated for cities like Atlanta, Georgia, Taipei, Taiwan, and New York City. DRS applications include the entire U.S. state of New Jersey and the nation of Singapore, using travel demand model trip-making predictions, publicly available taxi datasets, and/or synthetically generated itineraries. Investigations demonstrate system feasibility and/or assess the computational efficiency of different methods for assigning vehicles and/or matching travelers in shared rides. (See Agatz et al., 2011; Santi et al., 2014; Alonso-Moro et al., 2016; Brownell and Kornhauser, 2014; Bhat, 2016; Tao, 2007; and Spieser et al., 2014.)

Agatz et al. (2011) developed a sophisticated algorithm to match riders to their drivers and conducted a simulation using person-trip data obtained from Atlanta's travel demand model. Their results suggest that DRS works well not only in high-density, high-use settings, but also in sprawling suburbs and at low rates of utilization. However, they focused on driver (and thus TNC vehicle) unavailability, which can hamper sharing and dilute DRS opportunities. Brownell and Kornhauser (2014) focused on SAV system performance for the state of New Jersey. Employing a gridded-network for the entire state, along with synthetic trip-making data, valuable precision, accuracy, and applicability may have been lost in assessing optimal fleet requirements.

Santi et al. (2014) and Alonso-Moro et al. (2016) overcame both these issues by using publicly available taxi datasets for New York City and real networks (via OpenStreetMaps, an open-source platform for map data). Alonso-Moro et al. observed that 98% of the City's 3 million taxi trips could be served with just 2,000 vehicles and low waiting times (averaging just 2.8 minutes), backing DRS capabilities. Bhat (2016) confirmed those New York City taxi results, and added a vehicle repositioning algorithm. Tao (2007) also used a taxi data set, but for the city of Taipei. He developed a heuristic DRS algorithm using real-time taxi movements (not just trip calls by

travelers) to test its efficiency in a realistic network setting. Tao (2007) achieved 60% ride matches and concluded that a higher matching rate could be obtained across larger networks with greater density of trip-making.

Of course, taxis do not represent all person-trips in any region. Such trips tend to be shorter than household-vehicle trips (due to their cost), more often for business reasons or those without parking access (again due to their cost), and for visitors (due to their unfamiliarity with the region). DRS investigations of more representative trip-making are desired. By using a population-weighted cellphone dataset, as done here, one overcomes the drawbacks of faked or taxi-based trip patterns. However, certain details are lost (such as trip-to-trip connections throughout the day), in order to protect travelers' privacy, over space and time. Thus, cell-phone-based trips or other forms of extensive diary data tend to be aggregated by traffic analysis zones (TAZs) or neighborhoods, to obscure home and work addresses. To keep data size manageable (for dataset sharing), trips are often aggregated into hourly or multi-hour time-of-day bins as well. More detailed trip ends and trip schedules can be simulated/faked and disaggregated, while preserving the population's basic trip patterns. This process ensures that matches are less obvious (with trips coming from all over a zone and hour, rather than from its centroid or mid-point, for example), and so was used here. But it comes at the expense of some accuracy and precision (versus the reality of actual trip locations and times, which are rarely available to anyone, for any large population).

13.2. Cellphone Dataset

The cellphone-based dataset employed here was generated by AirSage for the month of April 2014 and for travel across the Orlando metropolitan area in Florida. AirSage uses the regular location pings of cell phones that are turned on and carried by customers of its partner companies (like Verizon and Sprint). Cellphone trips observed were aggregated based on six factors: each trip's inferred origin and destination TAZs, the hour and day in which *most* of the trip was made (e.g., 0100-0200 on April 4 or 1600-1700 on April 20), inferred trip purpose, and cell-phone subscriber class. All trips (and basic demographics) inferred from phone pings (of the carriers' cell towers) were then expanded to reflect all trip-making in the region using population-weighted trip counts (including travel by persons who do not own cell phones or carry theirs with them, turned off). This type of cellphone data has been proven to represent origin-destination (OD) flows to a reasonably high-degree of accuracy by capturing activity-based individualistic data (Calabrese et al., 2011; Alexander et al., 2015) but it is acknowledged that there are limitations based on market share of the phone-operator whose data is used for OD flow determination and the size of the TAZs considered.

The Orlando region's metropolitan planning agency models travel across 1,267 TAZs (with 1,261 of them representing metropolitan area and the remaining 6 representing external TAZs). External-zone trips can be very long, with ambiguity in their true destination or origin, so all external trips were removed from the dataset before seeking matches. The remaining 1,261 TAZs have a mean area of 2.22 sq. mi., a standard deviation of 9.92 sq. mi., and a median of 0.53 sq. mi. Traveler type based on work-type (such as, someone who works from home, works within the study area,

commutes to the study area for work, or commutes away from the study area for work) also is not relevant, so it is not used here, in making matches. The population-weighted dataset obtained from AirSage lacks mode-specific classification, but since this study attempts to prove the viability of DRS considering all trips, this information can be neglected for the purposes of this study.

MetroPlan Orlando, the region's metropolitan planning organization, provided a detailed network, with nearly 24,000 nodes and around 61,000 links. Shortest-path travel times between each TAZ were used while disaggregating the trips, as discussed in the next section.

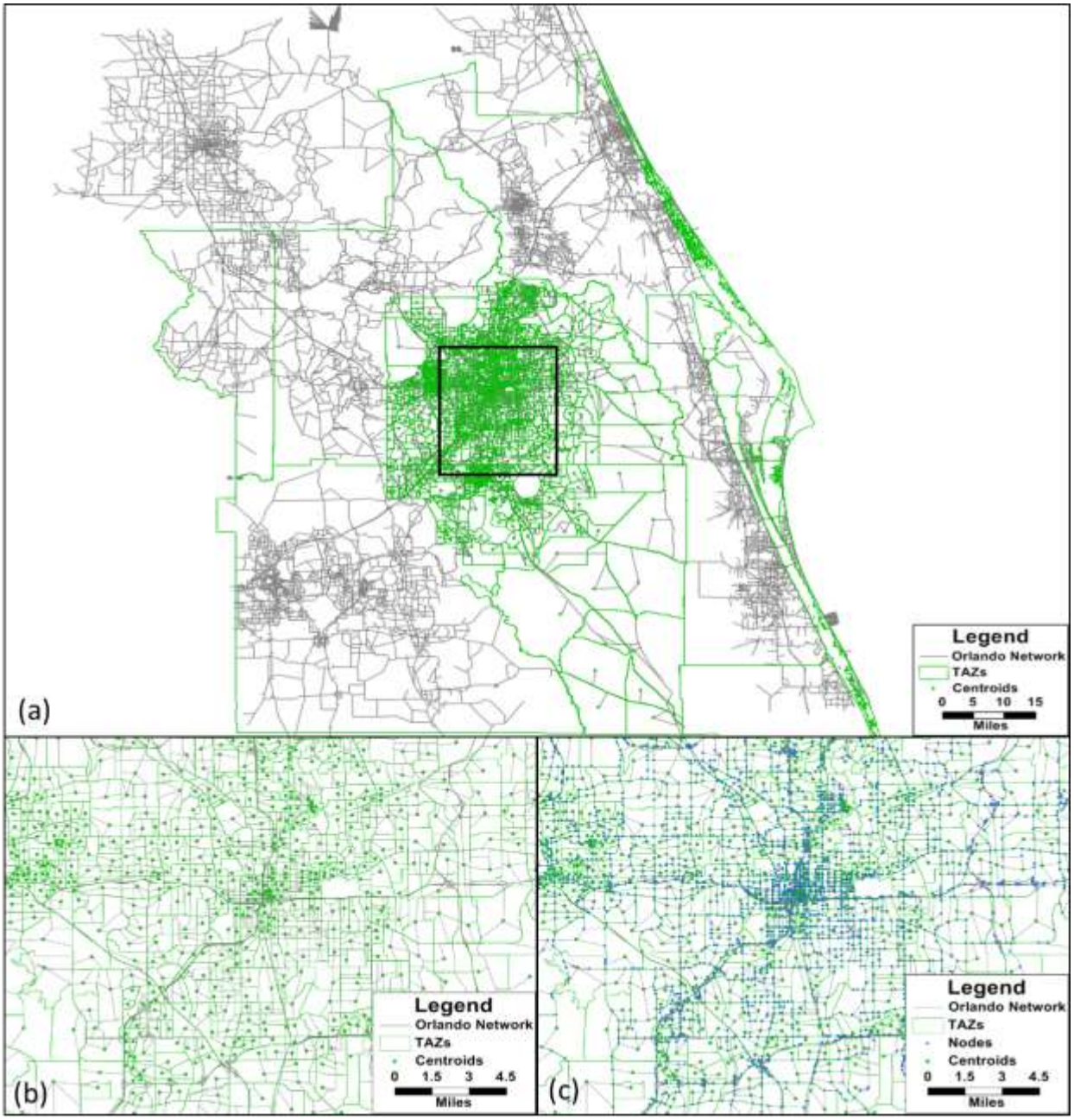
13.3. Methodological Framework

13.3.1. Data Disaggregation

Since AirSage provided an anonymized, spatially and temporally aggregate dataset (with trips classified into hourly bins and their origins and destinations by TAZs), smaller time steps, and more detailed locations (instead of centroids) were needed for a DRS application of intra-regional trips. Also, the departure times of these trips need not always be in the hourly bin that AirSage indicated for each trip, because trips (within this region) can begin many minutes earlier (or can end many minutes later). This is because only the majority of the trip's duration had to have occurred in the hour bin to which the trip was assigned by AirSage. Keeping these in mind, the data was disaggregated as explained below.

A time-step of one minute was used here, to facilitate computation while preserving dataset integrity, and origins and destinations were randomly sampled from within the origin and destination TAZs. To simplify the process, the trips occurring within an hourly bin were uniformly distributed within the bin. Then, to account for the variability in departure time as mentioned above, 30 minutes of overflow was permitted into the previous and next hour bin, obtained by randomizing the minute-level departure time. The origins and destinations for these trips, with varying departure times, were then sampled with equal probability from within their respective TAZs. Once a start time was assigned for these spatially disaggregated trips, the shortest-path travel times for that time of day, as obtained via Caliper Corporation's TransCAD software, a travel-demand modeling tool, were used to sample individual trip travel times from a normal distribution, whose mean equaled this shortest-path travel time and had a standard deviation of ± 2 minutes.

Thus, the original 30-day 24-hour dataset was disaggregated resulting in smooth, minute-by-minute trip-request files for each of the 30 days, with higher spatial detail and natural looking departure and arrival time patterns throughout each of the 30 days. The uniform disaggregation in time and space employed here would serve as conservative estimates of the actual DRS capabilities. One day in this disaggregated dataset contains nearly 6.2 million person-trips. Figure 14.1 illustrates the Orlando network and nodes.



a) Orlando network separated by TAZ gridlines b) Centroids used in aggregated data c) Nodes available for spatial disaggregation.

Figure 13.1 The Orlando network and nodes used for spatial disaggregation

13.3.2. Day to Day Variability in Travel Patterns

The cumulative trip distribution for each of the 30 days was obtained by time of day, as shown in Figure 13.2. It is evident that trip patterns are similar between weekdays and weekends. Variability, and consequently correlation, between each day was assessed using R software's statistical tool. Figure 14.2 shows correlation coefficients for trip counts across all OD pairs and across all 30 days of the month, with shading to highlight correlation patterns. Figure 14.2 indicates

that high correlation exists for trip patterns on Saturdays and Sundays, and for those made on weekdays, as one would expect (since weekdays have high shares of work and school trips, starting early in the day, while weekends have more flexible departure times and more recreational trip-making). Given these similarities, the following results are presented for a single weekday and a single weekend day. Results are very similar for other days of the 30-day dataset.

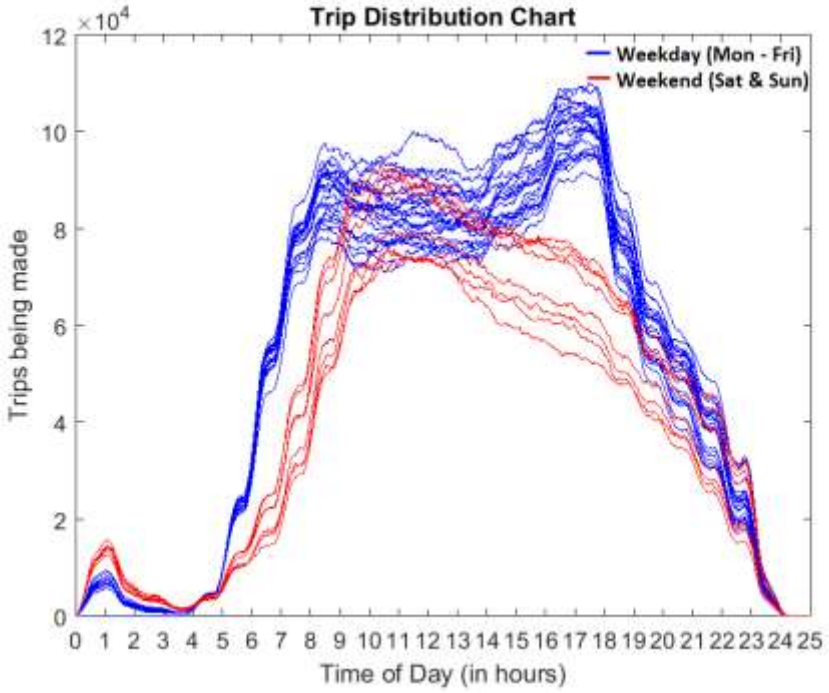


Figure 13.2 Orlando trip distribution differences, by time of day, between weekdays and weekends

13.3.3. Trip Matching

An analysis of these trip patterns suggests how many single-person trips can be matched with other trips, enabling ride-sharing, under different trip-delay and re-routing assumptions. A MATLAB code was developed to identify trips whose rides (in an SAV, for example) can be shared. An assumption of 4-person maximum vehicle occupancy was made, along with various travel delay thresholds, before running the code, for various maximum-delay scenarios (ranging from 5 minutes of extra travel time, to a maximum of 30 minutes).

Figure 13.3 illustrates how travel times under DRS conditions is calculated for this exploratory analysis, with ride-sharing en route, as compared to those sharing an origin zone and a destination zone and having similar departure times. As noted above, the OD DRS program matches individual trip-makers so that the earliest departing traveler (in a group of matched travelers, all having the same O and D zone pair) does not experience a wait time greater than the pre-determined limit. The en route DRS is more complex in matching travelers, in that it anticipates travelers arriving from different origins and destinations such that they have an intersecting path where each of their wait times between pick-up and drop-off are within the same predetermined limits. This is more in line with services available currently (with a human driver, of course).

Including the entire dataset of trips would mean that trips that are already shared/performed together, like family members travelling together for dinner, inflate the trip-sharing percentages. The Florida DOT (2013) estimates that over 50% of all automobile trips in that state are driven alone and 90% of all person-trips are driven in an automobile. Thus, only the person-trips in the AirSage dataset that may have been single-occupancy were used here, to perform matching (of solo travelers with one another, rather matching those already in traveling parties). This resulting data subset was found to contain nearly 2.8 million single-occupancy vehicle trips.

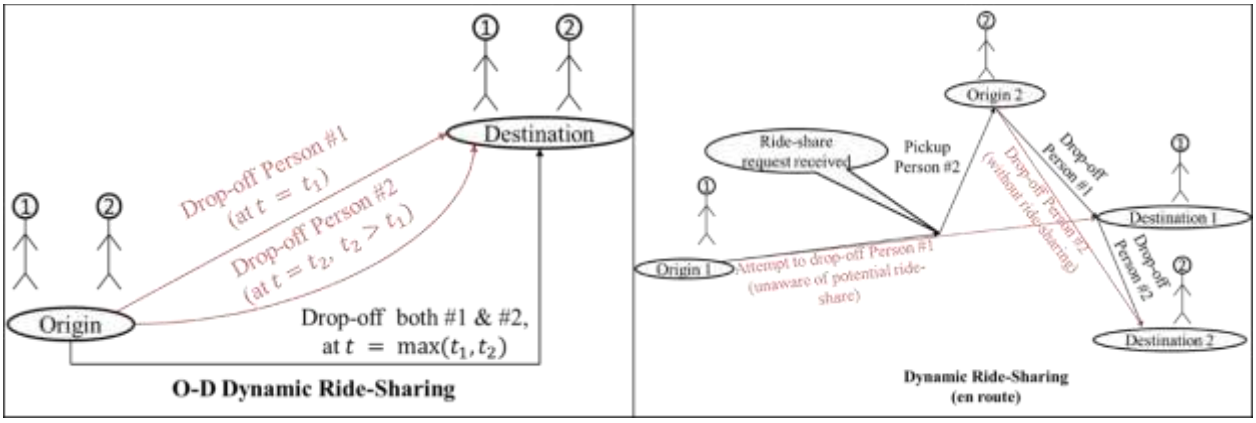


Figure 13.3 Illustrations of fleet-sharing of OD DRS and DRS en route

Table 13.1 Correlation between hourly trip-count vectors between all days for the month of April

	Tue	Wed	Thu	Fri	Sat	Sun	Mon	Tue	Wed	Thu	Fri	Sat	Sun	Mon	Tue	Wed	Thu	Fri	Sat	Sun	Mon	Tue	Wed	Thu	Fri	Sat	Sun	Mon	Tue	Wed
Tue	1.0000	0.9979	0.9984	0.9955	0.9346	0.9143	0.9979	0.9942	0.9986	0.9988	0.9954	0.9410	0.9248	0.9984	0.9972	0.9976	0.9974	0.9832	0.9351	0.9166	0.9979	0.9987	0.9988	0.9986	0.9953	0.9450	0.9235	0.9972	0.9961	0.9977
Wed	0.9979	1.0000	0.9991	0.9980	0.9476	0.9305	0.9984	0.9956	0.9989	0.9989	0.9978	0.9526	0.9394	0.9982	0.9984	0.9985	0.9988	0.9889	0.9482	0.9305	0.9982	0.9977	0.9989	0.9990	0.9976	0.9563	0.9382	0.9972	0.9943	0.9971
Thu	0.9984	0.9991	1.0000	0.9984	0.9451	0.9262	0.9979	0.9955	0.9987	0.9995	0.9982	0.9500	0.9357	0.9982	0.9980	0.9980	0.9990	0.9879	0.9452	0.9258	0.9982	0.9978	0.9986	0.9996	0.9982	0.9536	0.9342	0.9966	0.9940	0.9967
Fri	0.9955	0.9980	0.9984	1.0000	0.9534	0.9356	0.9963	0.9958	0.9963	0.9971	0.9997	0.9567	0.9436	0.9956	0.9975	0.9958	0.9978	0.9918	0.9537	0.9319	0.9962	0.9943	0.9958	0.9979	0.9989	0.9597	0.9424	0.9949	0.9908	0.9938
Sat	0.9346	0.9476	0.9451	0.9534	1.0000	0.9941	0.9411	0.9415	0.9423	0.9419	0.9531	0.9986	0.9966	0.9432	0.9467	0.9526	0.9541	0.9768	0.9990	0.9917	0.9414	0.9287	0.9384	0.9445	0.9531	0.9981	0.9955	0.9352	0.9218	0.9296
Sun	0.9143	0.9305	0.9262	0.9356	0.9941	1.0000	0.9258	0.9292	0.9227	0.9220	0.9341	0.9896	0.9989	0.9244	0.9311	0.9353	0.9370	0.9683	0.9958	0.9952	0.9261	0.9096	0.9195	0.9254	0.9332	0.9910	0.9990	0.9201	0.9061	0.9128
Mon	0.9979	0.9984	0.9979	0.9963	0.9411	0.9258	1.0000	0.9975	0.9970	0.9976	0.9961	0.9453	0.9344	0.9984	0.9992	0.9975	0.9973	0.9886	0.9422	0.9266	0.9996	0.9980	0.9977	0.9980	0.9946	0.9505	0.9342	0.9995	0.9979	0.9979
Tue	0.9942	0.9956	0.9955	0.9958	0.9415	0.9292	0.9975	1.0000	0.9933	0.9944	0.9950	0.9436	0.9361	0.9934	0.9977	0.9935	0.9949	0.9908	0.9441	0.9253	0.9970	0.9943	0.9945	0.9959	0.9935	0.9487	0.9367	0.9976	0.9956	0.9965
Wed	0.9986	0.9989	0.9987	0.9963	0.9423	0.9227	0.9970	0.9933	1.0000	0.9993	0.9963	0.9489	0.9330	0.9978	0.9968	0.9986	0.9982	0.9849	0.9427	0.9252	0.9970	0.9983	0.9995	0.9989	0.9970	0.9524	0.9313	0.9959	0.9935	0.9974
Thu	0.9988	0.9989	0.9995	0.9971	0.9419	0.9220	0.9976	0.9944	0.9993	1.0000	0.9970	0.9477	0.9321	0.9983	0.9974	0.9983	0.9986	0.9857	0.9420	0.9234	0.9980	0.9986	0.9992	0.9995	0.9976	0.9513	0.9307	0.9966	0.9944	0.9972
Fri	0.9954	0.9978	0.9982	0.9997	0.9531	0.9341	0.9961	0.9950	0.9963	0.9970	1.0000	0.9568	0.9424	0.9958	0.9973	0.9957	0.9976	0.9910	0.9529	0.9310	0.9957	0.9942	0.9957	0.9978	0.9988	0.9595	0.9411	0.9945	0.9905	0.9934
Sat	0.9410	0.9526	0.9500	0.9567	0.9986	0.9896	0.9453	0.9436	0.9489	0.9477	0.9568	1.0000	0.9939	0.9494	0.9508	0.9586	0.9586	0.9768	0.9974	0.9907	0.9456	0.9349	0.9448	0.9498	0.9574	0.9988	0.9921	0.9394	0.9269	0.9354
Sun	0.9248	0.9394	0.9357	0.9436	0.9966	0.9989	0.9344	0.9361	0.9330	0.9321	0.9424	0.9939	1.0000	0.9345	0.9395	0.9451	0.9457	0.9730	0.9977	0.9970	0.9349	0.9202	0.9296	0.9351	0.9421	0.9952	0.9993	0.9290	0.9155	0.9224
Mon	0.9984	0.9982	0.9982	0.9956	0.9432	0.9244	0.9984	0.9934	0.9978	0.9983	0.9958	0.9494	0.9345	1.0000	0.9981	0.9986	0.9977	0.9861	0.9430	0.9281	0.9984	0.9977	0.9980	0.9985	0.9951	0.9535	0.9334	0.9973	0.9956	0.9963
Tue	0.9972	0.9984	0.9980	0.9975	0.9467	0.9311	0.9992	0.9977	0.9968	0.9974	0.9973	0.9508	0.9395	0.9981	1.0000	0.9974	0.9974	0.9907	0.9477	0.9309	0.9988	0.9967	0.9973	0.9983	0.9960	0.9550	0.9393	0.9984	0.9962	0.9972
Wed	0.9976	0.9985	0.9980	0.9958	0.9526	0.9353	0.9975	0.9935	0.9986	0.9983	0.9957	0.9586	0.9451	0.9986	0.9974	1.0000	0.9986	0.9894	0.9531	0.9390	0.9974	0.9966	0.9983	0.9984	0.9961	0.9621	0.9437	0.9961	0.9931	0.9962
Thu	0.9974	0.9988	0.9990	0.9978	0.9541	0.9370	0.9973	0.9949	0.9982	0.9986	0.9976	0.9586	0.9457	0.9977	0.9974	0.9986	1.0000	0.9914	0.9543	0.9374	0.9976	0.9962	0.9978	0.9988	0.9974	0.9621	0.9445	0.9956	0.9925	0.9953
Fri	0.9832	0.9889	0.9879	0.9918	0.9768	0.9683	0.9886	0.9908	0.9849	0.9857	0.9910	0.9768	0.9730	0.9861	0.9907	0.9894	0.9914	1.0000	0.9787	0.9647	0.9886	0.9809	0.9841	0.9877	0.9892	0.9807	0.9736	0.9866	0.9804	0.9825
Sat	0.9351	0.9482	0.9452	0.9537	0.9990	0.9958	0.9422	0.9441	0.9427	0.9420	0.9529	0.9974	0.9977	0.9430	0.9477	0.9531	0.9543	0.9787	1.0000	0.9930	0.9425	0.9297	0.9392	0.9447	0.9531	0.9976	0.9972	0.9368	0.9237	0.9313
Sun	0.9166	0.9305	0.9258	0.9319	0.9917	0.9952	0.9266	0.9253	0.9252	0.9234	0.9310	0.9907	0.9970	0.9281	0.9309	0.9390	0.9374	0.9647	0.9930	1.0000	0.9273	0.9125	0.9215	0.9259	0.9306	0.9931	0.9969	0.9212	0.9084	0.9141
Mon	0.9979	0.9982	0.9982	0.9962	0.9414	0.9261	0.9996	0.9970	0.9970	0.9980	0.9957	0.9456	0.9349	0.9984	0.9988	0.9974	0.9976	0.9886	0.9425	0.9273	1.0000	0.9982	0.9975	0.9981	0.9948	0.9511	0.9346	0.9990	0.9973	0.9974
Tue	0.9987	0.9977	0.9978	0.9943	0.9287	0.9096	0.9980	0.9943	0.9983	0.9986	0.9942	0.9349	0.9202	0.9977	0.9967	0.9966	0.9962	0.9809	0.9297	0.9125	0.9982	1.0000	0.9988	0.9980	0.9942	0.9402	0.9191	0.9981	0.9972	0.9983
Wed	0.9988	0.9989	0.9986	0.9958	0.9384	0.9195	0.9977	0.9945	0.9995	0.9992	0.9957	0.9448	0.9296	0.9980	0.9973	0.9983	0.9978	0.9841	0.9392	0.9215	0.9975	0.9988	1.0000	0.9990	0.9963	0.9485	0.9283	0.9969	0.9952	0.9985
Thu	0.9986	0.9990	0.9996	0.9979	0.9445	0.9254	0.9980	0.9959	0.9989	0.9995	0.9978	0.9498	0.9351	0.9985	0.9983	0.9984	0.9988	0.9877	0.9447	0.9259	0.9981	0.9980	0.9990	1.0000	0.9980	0.9533	0.9339	0.9970	0.9946	0.9976
Fri	0.9953	0.9976	0.9982	0.9989	0.9531	0.9332	0.9946	0.9935	0.9970	0.9976	0.9988	0.9574	0.9421	0.9951	0.9960	0.9961	0.9974	0.9892	0.9531	0.9306	0.9948	0.9942	0.9963	0.9980	1.0000	0.9595	0.9403	0.9929	0.9884	0.9928
Sat	0.9450	0.9563	0.9536	0.9597	0.9981	0.9910	0.9505	0.9487	0.9524	0.9513	0.9595	0.9988	0.9952	0.9535	0.9550	0.9621	0.9621	0.9807	0.9976	0.9931	0.9511	0.9402	0.9485	0.9533	0.9595	1.0000	0.9941	0.9454	0.9333	0.9401
Sun	0.9235	0.9382	0.9342	0.9424	0.9955	0.9990	0.9342	0.9367	0.9313	0.9307	0.9411	0.9921	0.9993	0.9334	0.9393	0.9437	0.9445	0.9736	0.9972	0.9969	0.9346	0.9191	0.9283	0.9339	0.9403	0.9941	1.0000	0.9291	0.9161	0.9222
Mon	0.9972	0.9972	0.9966	0.9949	0.9352	0.9201	0.9995	0.9976	0.9959	0.9966	0.9945	0.9394	0.9290	0.9973	0.9984	0.9961	0.9956	0.9866	0.9368	0.9212	0.9990	0.9981	0.9969	0.9970	0.9929	0.9454	0.9291	1.0000	0.9988	0.9981
Tue	0.9961	0.9943	0.9940	0.9908	0.9218	0.9061	0.9979	0.9956	0.9935	0.9944	0.9905	0.9269	0.9155	0.9956	0.9962	0.9931	0.9925	0.9804	0.9237	0.9084	0.9973	0.9972	0.9952	0.9946	0.9884	0.9333	0.9161	0.9988	1.0000	0.9976
Wed	0.9977	0.9971	0.9967	0.9938	0.9296	0.9128	0.9979	0.9965	0.9974	0.9972	0.9934	0.9354	0.9224	0.9963	0.9972	0.9962	0.9953	0.9825	0.9313	0.9141	0.9974	0.9983	0.9985	0.9976	0.9928	0.9401	0.9222	0.9981	0.9976	1.0000

13.3.4. Fleet Simulation

A fleet simulation was carried out to assess the optimal SAV fleet requirement for the metropolitan region of Orlando to cater to all the trips with pre-specified service characteristics (such as, maximum waiting time or maximum additional in-vehicle travel time). Here, practicality is defined from an operator's perspective: a practical fleet is one with fewest variables able to serve the most (single-person) trips possible while adhering to these pre-specific characteristics. A framework was developed in MATLAB to simulate a fleet of SAVs for a typical day. The trip request file generated from data disaggregation served as an input to the framework, along with the characteristics that are expected of the fleet. This included fleet size, maximum allowable waiting time before an SAV is assigned to a passenger, maximum allowable time an SAV can take to reach the passenger, maximum additional time that is imposed on passengers who will be detoured for a new pick-up, and maximum additional time that a newly picked-up passenger has to wait while the previous occupants are dropped off. Table 14.2 states all these variables along with their abbreviations and this will stay consistent in definition for the remaining sections of the chapter. In addition to this, Orlando's network was converted into a MATLAB directional graph (digraph) and used to analyze shortest-path routes and times taken by SAVs.

Table 13.2 List of abbreviations used in reference to the simulation framework

Abbreviation	Description	Values Considered
noOfSAVs	Total number of SAVs used in the fleet	{5k, 10k, … 30k, 60k, 120k}
maxExtraTripTime	Minimum time imposed on travelers sharing their trips	{5 minutes, 10 minutes, … 30 minutes}
maxWaitingTime	Maximum time that a passenger had to wait before an SAV reached them	5 minutes
maxSearchTime	Maximum time that a trip was stored on the waitlist before being rejected	{0 minutes, 1 minute, 3 minutes, 5 minutes}
unserviced	Total trips that could not be serviced under the above restrictions	*Internally calculated*
ETA	Estimated time of arrival for an SAV to either pick up or drop off a passenger	*Internally calculated*

The framework was composed of three distinct blocks: SAV allocation, SAV update, and waitlist management. The SAV allocation block allocates the nearest SAV to a trip request based on the *maxWaitingTime* criterion. If no SAV was found satisfying this criterion, the trip request is stored in the waitlist. If an SAV with an existing occupant is located, the *maxExtraTripTime* criterion is checked prior to allocation, to minimize delays imposed on the travelers. After all the trips in a particular time step are either allocated to an SAV or stored in the waitlist, the SAV update block for the next time step is executed. In the SAV update block, the current location, destination and *ETA* of an SAV is monitored. If the SAV has not reached its destination for either a pick-up or a drop-off operation, then its current location and *ETA* are updated. If the SAV has reached its destination for pick-up, the drop-off operation is initiated. If a drop-off was executed, the second

destination for drop off of shared rides is processed, or the SAV stays idle, waiting for the next request. Once the update block has executed, all previously waitlisted trip requests are checked for SAV allocation before moving on to the next time step of trip requests. If the trip requests have been on the waitlist for more than *maxSearchTime*, they are removed from the waitlist and *unserviced* is updated to reflect the same. The flowchart for the process described is shown in Figure 13.4. Fleet sizes varying from 5,000 to 25,000 SAVs, in intervals of 5,000, were used for these simulations and the results are discussed in the next section.

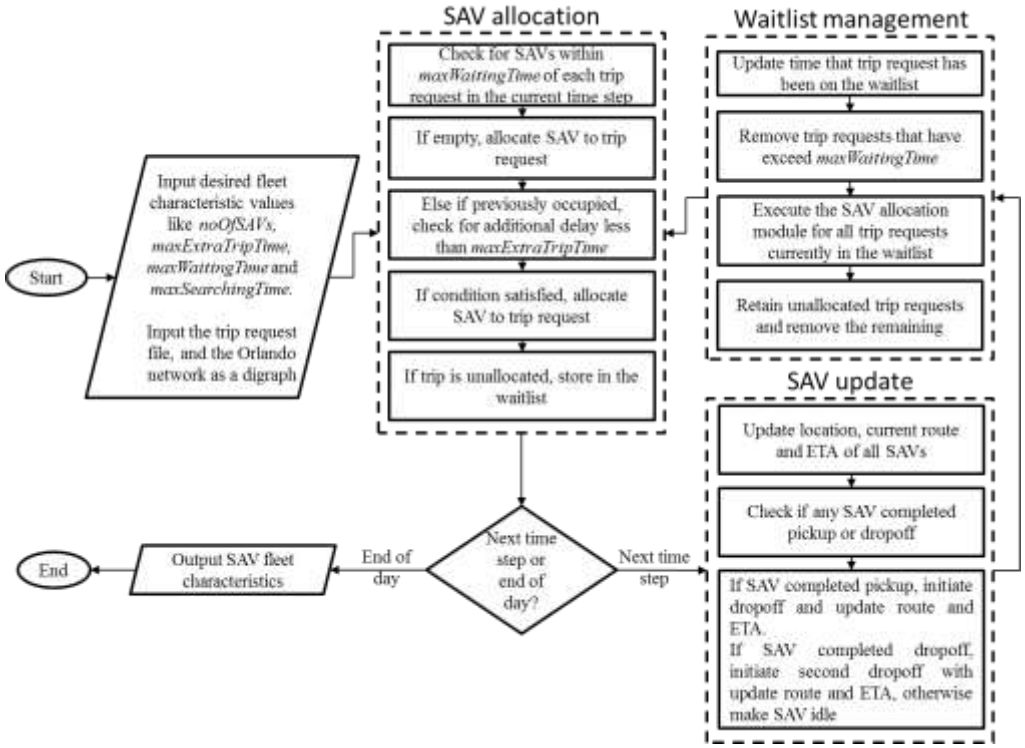


Figure 13.4 The flowchart describing the main modules of the simulation framework

13.4. Results

13.4.1. Infinite-fleet-based Trip Matching

Trips matched assuming availability of an infinite fleet provided optimistic results. As shown in Table 14.3, even after removing a large share of trips that reflect traveling parties (and thus focusing only on Orlando trips undertaken by a single person), nearly 60% of all such single-person trips can be shared with less than 5 minutes of added total travel (for each of the ride-sharing travelers, including any wait time added). This percentage reaches 86% matching or shared when travelers are willing to wait (or delay their destination arrivals, for example) up to 30 minutes. Of course, not all travelers need to be willing to wait that long; most of the matches are made with added delays of under 5 minutes. It is interesting to note that OD DRS remains almost a constant for trips with maximum allowed travel time greater than 10 minutes. This is due to the

spatial constraint on these trips which restricts scope for matches after a point in temporal flexibility.

Table 13.3 Percentage of Orlando trips that can be shared with OD DRS and DRS en route for a 4-passenger SAV under different maximum-delay assumptions

Maximum added travel time (including wait time)	Percentage of trips that can be shared (OD DRS)	Percentage of trips that can be shared (DRS en route)
5 min	18.48%	56.82%
10	20.56	74.15
15	20.55	80.56
20	20.57	83.57
25	20.65	85.29
30	20.65	86.23

13.4.2. Fixed-fleet-based DRS Simulation

A fixed fleet assumption offers reliable results in terms of ready applicability. A simulation based on a fixed fleet size and given service characteristics were simulated to obtain optimal fleet sizes for each permutation and combination that was found to be practically valid. Table 14.4 shows the different fleet sizes assumed in different scenarios, as well as the different service characteristics. The percentage demand served, percentage VMT reduction observed, percentage empty VMT, and the average number of trips served by an SAV has been shown as metrics to assess the best fleet. A conventional vehicle replacement ratio is also calculated, just as done by Loeb and Kockelman (2017) and Fagnant and Kockelman (2016). The average number of trips made by a conventional vehicle in one day is 3.05 according to the NHTS. Since the average SAV focused on solo travelers in the Orlando region serves 17.99 person-trips/day, it appears that nearly 6 conventional vehicles can be replaced by 1 SAV. The change in VMT was calculated relative to the VMT observed by the trips on the network without the fleet. Naturally, larger fleets had lower reductions in VMT.

Table 13.4 DRS potential based on fleet size and service characteristic

Maximum search time	Maximum waiting time	Maximum added trip time	No. of SAVs	% trips unserviced (by SAVs)	% demand captured (by SAVs)
0 min	5 min	5 min	5,000	54.15%	4.07%
			10,000	53.55	5.14
			15,000	51.84	8.17
			20,000	46.20	18.65
			25,000	38.45	32.38
			30,000	32.59	42.57
0 min	5 min	10 min	5,000	54.92%	2.71%
			10,000	53.07	5.99
			15,000	51.06	9.61
			20,000	49.35	12.58
			25,000	43.63	23.12
			30,000	39.20	31.02
0 min	5 min	15 min	5,000	54.78%	2.96%
			10,000	52.81	6.45
			15,000	50.88	9.87
			20,000	48.97	13.25
			25,000	45.76	19.14
			30,000	40.29	29.05
1 min	5 min	5 min	5,000	55.16%	2.29%
			10,000	53.57	5.10
			15,000	51.86	8.13
			20,000	46.16	18.70
			25,000	38.73	31.91
			30,000	33.24	41.45

13.5. Conclusions

This study anticipates the fraction of single-person trips that appear easily matched with one another, making them excellent candidates for DRS across the Orlando metropolitan area. Several studies have simulated the operations of SAV fleets but without the comprehensive nature of this cellphone-based dataset (e.g., taxi datasets do not reflect other modes of travel) and/or without other key data (e.g., actual travel times). With such data in hand, and a new setting for simulation (a Florida city and major destination for many vacationers), the results obtained here may be relevant for many interested in encouraging SAV use and DRS, to keep travel costs, VMT, emissions, and congestion down, as self-driving vehicles start making travel easier.

The trip-matching algorithm employed here suggests that nearly 60% of all single-person trips occurring each weekday in Orlando appear matchable to other trips taking place (for those traveling solo), with less than 5 minutes of added total travel time (including any wait time). Any added willingness to wait (up to 10 minutes or 15 minutes, maximum, for example) brings this

percentage up (to 74.2% and 80.6%, respectively), suggesting substantial opportunities for VMT reduction and shared-fleet activities in many (and probably all) cities around the U.S. and presumably around the world. The second part of the chapter used a fleet simulation algorithm to gauge the fleet size requirements to achieve the above predicted levels of ride-sharing. Results indicated that a fleet size of around 30,000 SAVs were sufficient to cater to less than 45% of Orlando's 2.8 million single-traveler trip demands (i.e., not counting existing carpools by family, friends, and colleagues). This means that one SAV can replace nearly 13 conventional vehicles. Under slightly relaxed trip delays and search times, such as a maximum of 15 and 2 minutes, the optimal fleet size required can be significantly reduced, thus increasing the replacement ratio.

One important limitation arising here is the assumed disaggregation of trips, over space and time. Uniform temporal and spatial disaggregation was used to spread AirSage cellphone trip ends over time and space. In reality, many trips may be more concentrated, increasing the likelihood of trip-matching, especially during peak times of day. Real-world implementations may be even more successful.

In addition, average vehicle occupancies form an integral part of determining how effective the fleet is at matching and sharing trips. To do this, vehicle occupancies need to be computed at each leg and averaged over distance or time. The complexity involved in tracking the fleet with limited information from the network has reduced the scope of the study to understanding only the number of trips served. However, the framework can be modified as a next step to compute this AVO. In addition to this, although it is true that shared trips can significantly reduce the number of vehicles using the road infrastructure, it cannot be said with 100% certainty that congestion will fall. This is because of the induced demand from lower-cost transportation and the intensity with which a shared fleet with a limited size will have to operate. The effect of fleet operations can be captured by computing the VMT of this fleet and comparing it with the literature. Again, this can be included in the simulation framework to achieve more comprehensive results. Regardless of such changes to this work, the results deliver very strong evidence of DRS as a highly viable way to reduce vehicle use and VMT in a metro region. All it requires is travelers' willingness to share rides with people they do not yet know. Hopefully, that will not pose a challenge long-term, so that our cities and nations can reduce fossil fuel reliance, emissions, congestion, and travel costs.

Chapter 14. Pricing Strategies with CAVs in the Mix

14.1. Background

The rate of development of autonomous vehicles (AVs) made in the last few years is generating a great deal of discussion both inside and outside the scientific community. Considerable progress has been in made in AV-technologies thanks to the investment of auto manufacturers (Muoio, 2017) and the support of public institutions (Kang, 2016).

Since the introduction of driverless cars represents an additional option for travelers, some of the trips made with "traditional modes" like car, public transit, and bike will be replaced by trips with autonomous and shared-autonomous vehicles (SAVs). While the benefits of AV implementation are evident from a road safety and energy-consumption perspective (Fagnant and Kockelman, 2015), it is very difficult to predict their effects on mobility and traffic (Litman, 2017; Wadud et al., 2016).

On one hand, automated technologies are likely to improve the traffic performance of road networks, as they will increase the traffic throughput thanks to tighter headways and they will reduce traffic incidents. On the other hand, AVs and SAVs might increase the number and the distance of trips because of the increased comfort and lower value of travel time (VOTT). These changes might ultimately yield to higher levels of congestion. Since infrastructure solutions (such as increasing capacity) are usually inadequate and traffic management strategies can only partially solve the problem (Litman, 2016), it is important to consider demand management options such as congestion pricing (CP).

Charging drivers for the congestion they cause is a well-known concept among economists, traffic engineers, and transport professionals. Many studies have been conducted to explore and better address this issue and in the last twenty years the first CP schemes have finally been implemented in some cities (Singapore, London, Stockholm, Milan, and Gothenburg). Although several models and strategies have been proposed from a theoretical perspective, in reality, CP practices in cities have been limited to cordon and area-based schemes. AVs and SAVs offer the opportunity to implement more efficient and effective strategies, thanks to advanced communication capabilities and fast information sharing.

In this report, we investigate two different CP strategies in possible future scenarios characterized by a large presence of AVs and SAVs.

The topic of CP in scenarios with conventional and autonomous vehicles (shared and private) is relatively unexplored, with the exception of a few theoretical studies (described below). Experiments within this study are performed using the multi-agent transport simulation MATSim. Such typology of model is particularly suitable for large simulations of CP strategies (involving thousands of agents), as it considers important behavioral aspects of traveling such as travel time departure, route and mode choice, and trip chains. Furthermore, MATSim allows for a realistic

representation of innovative travel modes such as AVs and SAVs from both the demand and supply perspective of the traffic congestion problem.

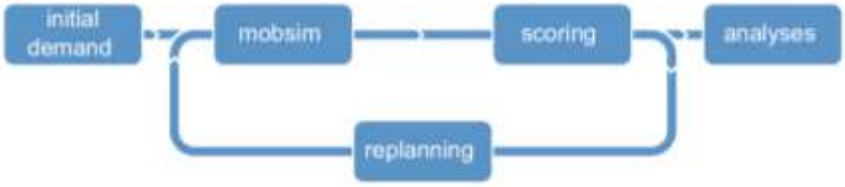
14.2. Modeling AVs, SAVs, and Traditional Modes in an Agent-based Model

In this section after providing a brief overview of the agent-based model MATSim, we present a description of our modeling framework. We then focus on the modeling of AVs and SAVs.

14.2.1. General Framework of MATSim

MATSim simulates an entire daily plan of every single user and it considers endogenous mode choice, departure time choice and route choice into a fully dynamic model. As opposed to models that use single trips, this model allows for predictions on reactions to demand management strategies, such as tolls during the span of a day, accounting for a higher level of realism. In fact, trips are typically linked to each other as a part of a daily plan and not that meaningful just as stand-alone trips (Balmer et al., 2006). Activities often have higher importance in the daily schedule than trips that simply represent connections among them. Since MATSim represents traffic behavior at a highly disaggregated level by modeling individual agents (with different socio-demographic characteristics), it is possible to investigate the effects of transport policies on travel behavior and traffic more in depth than in traditional four-step models (Kickhöfer et al., 2011). The overall process (Figure 13.1) can be summarized in the following stages:

- Each agent independently develops a plan that expresses its preferences in terms of activities, trips and their schedules during the day (Initial demand).
- The agents simultaneously perform all the plans in the physical system in the mobility simulation (Mobsim). Congestion phenomena are modeled using a queue model, which takes both the physical storage capacity and the actual throughput (flow capacity) of a link into account.
- To compare the performance of different plans, each one is associated to a score given by a utility function (Scoring).
- Agents are able to memorize their plans and improve them during the simulation by means of a learning algorithm (Replanning). During the implementation the system iterates between plan generation and traffic flow simulation.
- The cycle continues until the system has reached an equilibrium where no agent can improve anymore his score (Analyses).



Source: Horni et al., 2016

Figure 14.1 MATSim cycle

The choice model generally adopted in MATSim is equivalent to the standard multinomial logit model. Since the number of plans in the memory of agents is limited, the worst performing one is replaced by a new one at each iteration. Thanks to this feedback mechanism agents are able to improve their plans over several iterations until the system reaches the "relaxed" state when agents cannot significantly improve their plans and the outcome of the system becomes stable. This state is also referred as agent-based stochastic user equilibrium (Nagel and Flotterod, 2009). For further information about the simulation framework MATSim, see Horni et al. (2016).

14.2.2. Choice Dimensions and Parameters

Plans can be improved by changing the time of departure, varying the route, and choosing different a transport mode through modules. Agents' travel choices are modeled in MATSim through an iterative learning mechanism based on a quantitative score, referred to as utility. For each iteration agents choose from an existing set of daily plans according to a multinomial logit model.

Every daily plan is associated with a utility, accounting for a trip-related disutility and a performing activity utility:

$$V_{plan} = \sum_{i=1}^{n} (V_{act,i} + V_{trip,i}) \tag{13.1}$$

where V_{plan} is the total utility of a daily plan; n is the total number of activities or trips; $V_{act,i}$ is the utility for performing activity i; and $V_{trip,i}$ is the utility of the trip to activity i. The first and the last activity are wrapped around the day and handled as one activity. Thus, the number of activities and trips is the same. The trip-related utility for each mode is calculated as follows:

$$V_{q,i} = \beta_{0,q} + \beta_{t,q} \cdot t_{i,q} + \beta_c \cdot c_{i,q} \tag{13.2}$$

where $\beta_{0,q}$ corresponds to the alternative specific constant of mode q; $t_{i,q}$ corresponds to the travel time of leg i traveled with mode q; $\beta_{t,q}$ corresponds to the marginal utility of traveling by mode q; $c_{i,q}$ corresponds to the monetary cost of leg i traveled by mode q; and β_c corresponds to the marginal utility of monetary cost.

To calculate the positive utility gained by performing an activity, a logarithmic form is applied (Charypar and Nagel, 2005; Kickhofer et al., 2011):

$$V_{act,i}(t_{act,i}) = \beta_{act} \cdot t_i^* \cdot \ln\left(\frac{t_{act,i}}{t_{0,i}}\right) \tag{13.3}$$

where t_{act} is the actual duration of performing an activity (when the activity is open), t_i^* is an activity's 'typical' duration, and β_{act} is the marginal utility of performing an activity at its typical duration. In the equilibrium, all activities at their typical duration are required to have the same marginal utility; therefore, β_{act} applies to all activities. $t_{0,i}$ is a scaling parameter linked to an activity's priority and minimum duration. In this study, $t_{0,i}$ is not relevant, since activities cannot be dropped from daily plans.

The value of travel time saving (VTTS) is derived as follows:

$$VTTS = \frac{\beta_{act} - \beta_{t,q}}{\beta_m} \tag{13.4}$$

where β_m corresponds to the marginal utility of money.

The travel options modeled in this study include car, public transit, bike and walk (modeled jointly), AV, and SAV. The behavioral parameters for car and public transit used in this study are based on the work of Tirachini et al. (2014) and Kaddoura et al. (2015) and have been adjusted to reflect more realistically current travel costs in the U.S. (2017). The parameters used for the simulation are summarized in Table 13.1. In order to account for aspects such as parking and walking times of car users we have derived an alternative specific constant $\beta_{0,car} = -0.1$. In addition to that, car users pay a monetary cost proportional to the distance traveled corresponding to \$0.20 per mile. Since, waiting, egress and access times are not modeled in these experiments, public transit (PT) has been recalibrated yielding an alternative specific constant $\beta_{0,PT} = -1.5$. This value also accounts for the average ticket cost and for the particular reluctance of American society in using public transit. In similar fashion, the alternative specific constant for walking/biking has been set to $\beta_{0,active} = -0.2$. Similar to Kaddoura et al. (2015), the marginal utility of traveling by car is set to zero. Even if this value is set to zero, traveling by car will be implicitly punished by the opportunity cost of time (Horni et al., 2016). In this study, the marginal utility of money β_m is equal to 0.79 such that the VTTS for car users corresponds to about \$18 per hour. This value has been obtained according to the recommendations from the USDOT (2011).

The parameters for AVs have been mainly derived based on (Kockelman et al., 2017). The monetary costs are estimated to be around \$0.30 per mile. The operating costs might be higher than conventional cars because of the initial purchase cost, but would be partly compensated by increased efficiency and better insurance premiums. We assume AVs to have a null alternative specific constant in order to account for parking and walking time reductions. The marginal disutility of traveling equal to +0.48 to reflect a marginal cost of traveling equal to 50% of those

of car users (corresponding to a VTTS of about $9 per hour), in line with Gucwa (2014) and Kim et al. (2015)[28].

As for SAVs, we assume the same alternative specific constant and marginal cost of traveling of AVs. Unlike AVs, SAVs are characterized by waiting times depending on the availability of vehicles. We assume the monetary costs to be composed of a flat fee, a distance fare that change depending on the scenario.

Table 14.1 Adjusted mode parameters used in this study

Travel Mode	β_0	β_t
Car	-0.1	0
Public Transit	-1.5	-0.36
Walk/Bike	-0.2	0
AV	0	+0.48
SAV	0	+0.48

In addition to travel choices, agents can modify their activities' scheduling decisions by shifting, extending, or shortening activities considering aspects like the optimal duration, and opening/closing times of the facility (Table 13.2). Activities performed outside opening times do not yield any positive gain of utility. Furthermore, agents are subject to schedule penalty costs for being early or late accordingly to the well-known Vickrey's parameters α, β, and γ (Arnott et al., 1990).

Table 14.2 Activity attributes

Activity Type	Optimal duration	Opening time	Closing time
Home	14	undefined	undefined
Education	5	08:00	22:00
Work	7	07:00	undefined
Shopping	1	09:00	01:00
Leisure	2	09:00	01:00

14.2.3. Simulation Scenarios

The impacts of different pricing schemes are investigated for three different scenarios. The first one, to which we will refer as "**Base Scenario**" corresponds to a realistic simulation of the city of Austin and surroundings (Figure 13.2), comprising a considerable portion of the Austin metropolitan area (Greater Austin). The studied region includes a series of satellite cities such as Round Rock, Cedar Park, and Pflugerville. The road network used in the simulation consists of a high-resolution navigation network including about 211,000 road segments (links). The population and its plans have been obtained by adjusting those from Liu et al. (2017) who used CAMPO's households' data for Austin 2020. Although the plans have not been formally validated, resulting trip distances and durations are reasonably realistic. Normally, each agent needs to travel at least once per day to execute his plans. Instead of simulating the full population, a sample of 5%

[28] Note that, in MATSim, setting a positive marginal disutility of traveling does not imply a gain of score from the trip.

(equivalent to 45,000 agents) is used for the experiments of this study. Link capacities are downsized to match these with the sample size. The available transportation modes are car, public transit, and walk/bike (modeled jointly). In order to reflect current trends in availability of car as a travel option, we assume 90 percent of agents to have access to car (either as driver or passenger).

Source: Google Maps

Figure 14.2 Simulation network

The two additional scenarios correspond to possible future scenarios characterized by the presence of AVs and SAVs. Currently, it is not clear whether AVs will mainly replace privately owned vehicles or if they are going to be adopted as shared taxis. On one hand, the auto industry is moving quickly to provide the first "partially autonomous" models (Level 3) by 2020 and full autonomous models by 2030 (Level 4 and Level 5) (Kockelman et al., 2017). Conversely, car-sharing companies (Uber, Lyft, Didi) are already running tests (Kang, 2016; Hawkins, 2017), making considerable investments (Buhr, 2017), and developing important partnerships (Russell, 2017) to put driverless fleets on the road within a few years. Hence, we include an **"AV-oriented" Scenario** and a "**SAV-oriented" Scenario**, representing these two opposite trends. In the AV-oriented scenario, we assume a large portion of the population to switch from car to AV (90% of agents having accessibility to car in the Base Scenario). SAVs are available too, but the fleet size is relatively small (1 vehicle every 30 agents) and they are characterized by prices in line with current shared mobility services and predicted costs of SAVs ($0.5 flat charge, 0.4 $/mile distance charge and 0.1 $/min time charge). In the SAV-oriented scenario, SAVs are largely available (1 vehicle every 10 agents), whereas most of the population is still car-dependent (only 10 % has access to privately owned AVs). Furthermore, we assume a decrease of availability of privately owned vehicles to 60% in order to reflect a decrease of ownership (Litman, 2014). Autonomous vehicle implementation predictions (Victoria Transport Policy Institute, 28). In this case, SAVs are

characterized by lower prices than in the AV-oriented scenario (reduction by 50%), assuming that main companies and local authorities would stipulate agreements on prices concerning the provision of shared autonomous services.

Results of MATSim simulations in terms of modal shift are reported in Figure 13.3. In the Base Scenario car clearly appears as the dominant travel option, in line with current situation. In the AV-Oriented Scenario and SAV-Oriented Scenario, the introduction of two additional travel options (SAVs and AVs) generates significant changes. SAVs replace a considerable amount of PT trips that decrease to 0.5% and 1.8% respectively in the AV-Oriented Scenario and SAV-Oriented Scenario. Also "active trips" are reduced: 1.0% and 4.0% respectively in the AV-Oriented Scenario and SAV-Oriented Scenario. As result of this shift, congestion measured as daily total vehicle-miles traveled (VMT) and daily total travel delay increase in both the scenarios (Table 13.3). Particularly in the SAV-oriented Scenario, the large presence of SAVs seems to determine a considerable increase of delay (about 35% higher). Interestingly, a considerable source of congestion can be identified in empty SAV trips that account for 1.7% and 11.6% of the total VMT in the AV-oriented Scenario and SAV-oriented Scenario.

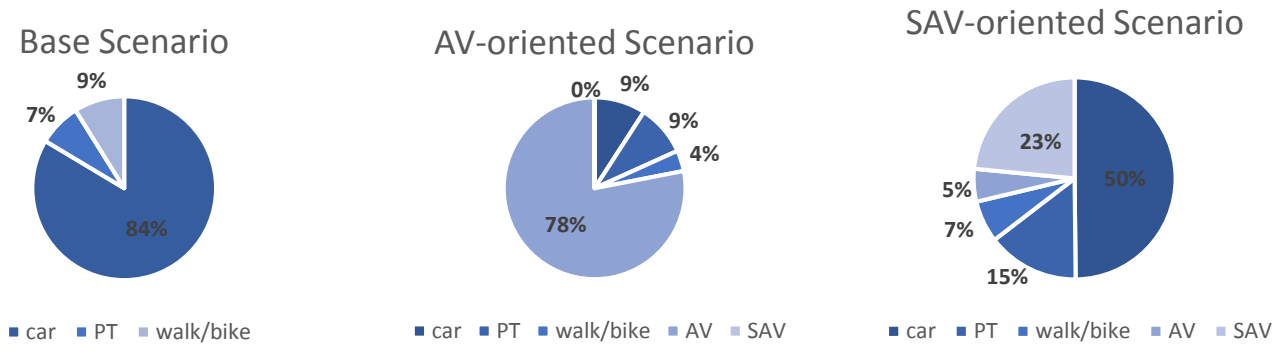


Figure 14.3 Modal share for the three different scenarios

Table 14.3 Traffic conditions of the three different scenarios

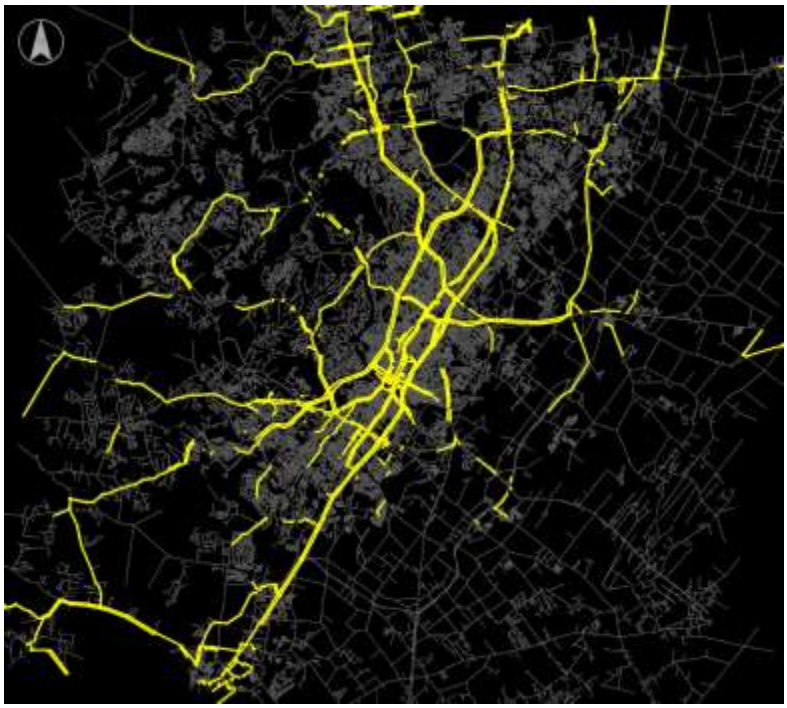
	Base Scenario	AV-oriented Scenario	SAV-oriented Scenario
Total Daily VMT	2,845,406	2,889,913	3,032,629
VMT by Empty SAVs	2,845,406	2,106	201,828
Total Travel Delay (veh-hours per weekday)	437,887	948,845	523,594

14.3. Pricing Strategies and Impact

Facility-based tolls are probably the most common form of CP since they do not require particularly advanced technologies for implementation. In the past, this typology of scheme has been applied mainly to tunnels, bridges, and highway facilities that represent major bottlenecks. In this study, a "**Link-based Scheme**" is applied to the one thousand most congested links during

the morning peak hours (7–9 AM) and evening peak hours (5–7 PM). The tolled links are selected based on the volume/capacity (V/C) ratio calculated on hourly basis and aggregated for the peak hour periods. A minimum threshold V/C ratio of 0.9 is chosen to identify the most congested links, resulting in the selection of about 2 to 4% of the road network (3,911 links in the Base Scenario, 5,100 links in the AV-Oriented Scenario, and 3,820 links in the SAV-Oriented Scenario). As Figure 13.4 demonstrates, the tolled links include the most important segments of road infrastructure of the region such as Interstate 35 and the Texas State Highway Loop 1. A flat toll rate of $0.2 is set to all the selected links regardless of the amount of congestion and the characteristics of the link.

Distance-based fee consists of charges that varies (linearly or not) with the distance traveled. Such type of scheme could be implemented relatively straightforwardly with the support of GPS. In this study, we investigate the effects of a linear "**Distance-based Scheme**" of $0.2 per mile operating between 7AM and 8PM.



Source: VIA:Senozon

Figure 14.4 Selected links in the Link-based Scheme for the base scenario

14.4. Results and Implications

The impacts derived from the different CP schemes in each scenario are discussed in this section. The evaluation of the schemes is carried out by means of a set of commonly used performance indicators such as mode shift, change of traffic delay, and motorized trips.

14.4.1. Mode Choice

All the schemes succeed in reducing car, AV, and SAV trips to a different extent. In all the scenarios PT and slow modes witness a considerable increase of mode share (Tables 13.4 and 13.5).

Overall, the demand for SAVs and AVs seem more elastic than the demand for car given the higher modal shift achieved for all the CP strategies. Because of their higher initial costs, AV and SAV travelers are more incentivized than car travelers to adopt PT or slow modes in presence of tolls. For this reason, CP strategies seem to be more effective in AV-oriented and SAV-oriented scenarios.

Among the traditional schemes, as expected the distance-based scheme generates larger changes in travelers' mode choice than the link-based scheme, particularly for the base scenario. The results are in line with previous studies about distance-based schemes (Litman, 1999). Only in the SAV-oriented scenario, the link-based scheme yields results comparable to the ones of the distance-based scheme. This is an interesting outcome, since the two schemes are conceptually very different from each other and might have different effects in terms of economic gains, distributional effects, and public acceptability.

Table 14.4 Modal shift from the link-based scheme

	AV oriented	SAV oriented	Base (no SAVs-AVs)
Change of car trips (%)	-0.8	-7.92	-6.81
Change of PT trips (%)	28.07	16.2	5.89
Change of walk/bike trips (%)	10.73	7.5	0.91
Change of AV trips (%)	-37.94	-2.3	0.0
Change of SAV trips (%)	-0.05	-13.48	0.0

Table 14.5 Modal shift from the distance-based scheme

	AV oriented	SAV oriented	Base (no SAVs-AVs)
Change of car trips (%)	-3.97	-7.31	-14.1
Change of PT trips (%)	35.55	14.74	2.89
Change of walk/bike trips (%)	11.11	5.74	11.29
Change of AV trips (%)	-42.62	-1.07	0
Change of SAV trips (%)	-0.06	-12.1	0

14.4.2. Traffic Performance of the Network

Both traditional and advanced CP strategies determine a significant reduction of trips traveled by AVs, SAVs, and cars (Figure 13.5). Schemes with a distance-dependent fee component achieve the highest reduction of trip given the larger scale of population affected and the high average trip length in the original scenarios (around 15 miles). The CP schemes seem to yield much higher improvements in the AV-oriented and SAV-oriented scenario because of higher elasticity.

However, this is just a single perspective to evaluate the effects of the strategies, as the changes in terms of network daily travel delay show (Figure 13.6). Interestingly, in the AV-oriented scenario all the CP strategies generate a considerable reduction of delays (above 80%). In the SAV-oriented, results significantly vary according to the scheme.

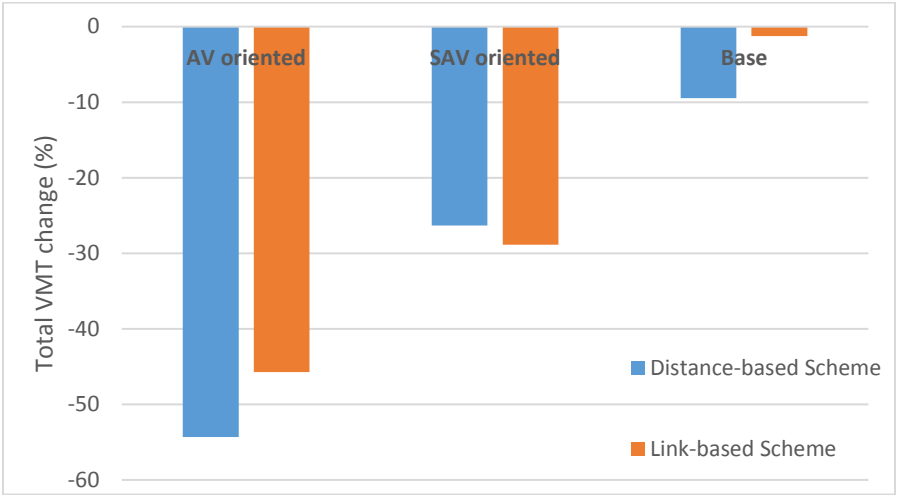


Figure 14.5 Reduction of motorized trips for the different scenarios according to the CP scheme

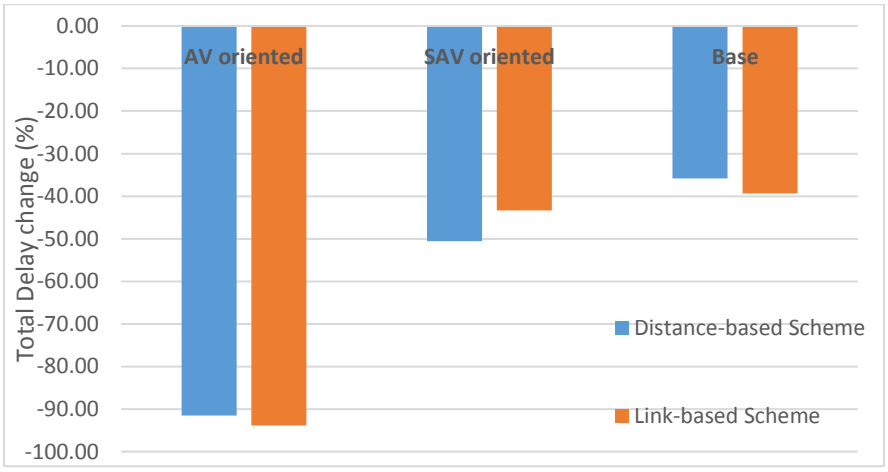


Figure 14.6 Reduction of traffic delay for the different scenarios according to the CP scheme

Chapter 15. Technologies for Congestion Pricing

15.1. Introduction

Traffic congestion is a major problem in all major urban areas, costing citizens valuable time. Congestion is caused by an excess of vehicles on part of a roadway at a given time, leading to vehicle speeds that are slower than the normal "free flow" speeds of that roadway (FHWA 2017). Congestion costs the U.S. economy over $100 billion a year, and this number is rising over time (Cebr 2014, Shrank et al. 2015, Burfeind 2017). This includes the direct costs of the value of fuel and time wasted as well as the indirect costs from the increased cost of doing business. As economies and populations continue to grow, congestion is expected to increase. In order to combat increasing gridlock, it is important to develop policies and implement technologies that reduce congestion.

Roadways are limited by their capacity, which is the maximum flow of traffic that can be handled by a given roadway section. Capacity flow values are affected by the number and width of lanes, median and merge area designs, intersection or interchange frequency, presence of stop signs or signal lights, curvature, grade, and other design variables (FHWA 2017). When demand for travel rises, congestion sets in, slowing travel speeds and lengthening travel times. Congestion can be recurring or non-recurring. Recurring congestion is the result of normal traffic volumes in a typical environment (Hallenbeck et al. 2003), such as peak times of day every weekday upstream of key bottlenecks (like bridge crossings) in urban environments. Non-recurring congestion is caused by unusual events or conditions that result in capacity losses or added demand. Vehicle collisions, construction zones, inclement weather, and special events (like professional football game days) can all result in non-recurring congestion by temporarily reducing capacity or exceeding existing corridor capacities. While transportation network capacities are rather limited by existing infrastructure investments, travel demands fluctuate minute to minute and day to day. Travel demand can be influenced via public policy, special events, and weather, among many other factors.

Without regulation and pricing, the demand-supply equilibrium for roadway space settles at a suboptimal point, because users only consider the direct costs of congestion on their personal travel time (Komanoff 2017). Users ignore the additional marginal cost of their travel on the transportation network, which adds to the travel time of all road users (Kockelman and Kalmanje 2005). This phenomenon is represented in the supply and demand curves shown in Figure 1.

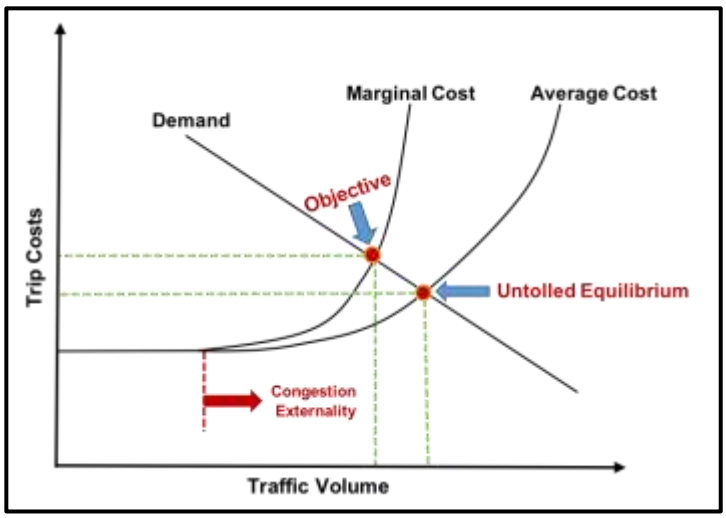


Figure 15.1. Traffic Supply and Demand Curves, for Tolled and Un-tolled Conditions

Congestion pricing is one potential solution to this issue. Such pricing or road tolling involves incentivizing certain link and thus route choices for drivers, to improve the overall efficiency of a congested corridor's or congested network's roadways. By charging a higher price to travel on highly-congested roadway sections or offering tax credits for traveling through less-congested areas, system managers can encourage choices that decrease system-wide costs and improve social welfare or net community benefits. By confronting users with the true cost of their travel (reflecting the delays they impose on other travelers, behind them, essentially), congestion pricing pushes the supply-demand equilibrium point to the left, decreasing traffic volume. Lower volume means less congestion and lower travel times on that link. Without congestion pricing, drivers face only the directly experienced or average cost of travel, resulting in over-consumption of what truly is a socially more expensive good than they realize. With appropriate pricing in place, travel choices become less sub-optimal, and ideally reflect the full cost of added vehicles on each roadway segment, at each time of day.

Recent and emerging developments in communication and computation technology make widespread implementation of congestion pricing systems feasible and potentially highly cost-effective. This paper examines the technologies and policies that could be implemented in a congestion pricing system. With information gathered from various expert sources, this work provides recommendations for the best mix of technology and policy in several transport settings, as well as a roll-out strategy for congestion pricing.

15.2. Policy Implementation

In order to deliver a successful and maximally cost-effective roadway pricing system, an appropriate policy structure is needed. A few major pricing policies are a vehicle-miles-traveled (VMT) tax, cordon- or area-based congestion pricing, and credit-based congestion pricing (CBCP). A VMT tax is simplest, and can fittingly recover general infrastructure investment and maintenance costs, for example; but it does not address congestion directly. Cordon-based and

area-based tolling reduce travel within high-traffic areas by charging for ingress during specific times of day, but they are broad-based and do not reflect over-use or under-use of specific links. Rationing by license plate and day of week or time of day has also been studied and used (see, e.g., Nakamura and Kockelman [2002]), but can lead to perverse outcomes (Nie 2016). CP and CBCP can directly and efficiently (in an economic sense) address congestion costs by location and time of day. Variations in tolling can influence trip generation by time of day, mode choices, destination and route choices; but only CBCP is designed to directly address congestion costs in time and space while addressing equity implications, thereby delivering greater societal benefits.

Both a VMT Tax and CBCP offer the opportunity to decrease congestion and collect additional funding. With the rise of autonomous vehicles (AVs), many sources of public funding may decrease. The 25 largest cities in the U.S. reported $5 billion in auto-related revenues in 2016 (Maciag 2017). If users opt for shared autonomous vehicles (SAVs), the need for parking may decrease as vehicles pick up new passengers. Parking fees and parking tickets make up a large portion of local government revenue used for infrastructure improvements in many cities. AVs will not violate traffic laws as often, decreasing revenues from traffic citations, which average $8.5 million in the largest cities (Maciag 2017). Additionally, the rise of electric vehicles will decrease revenue generated from motor fuel taxes, which accounted for $16 billion spent on local infrastructure or transit in 2015 (Maciag 2017). The implementation of a VMT tax or CBCP could help local governments to maintain the necessary infrastructure budgets. Additionally, decreased congestion would help to limit the need for infrastructure maintenance and construction costs while benefitting citizens through time savings.

15.2.1. Vehicle-Miles-Traveled (VMT) Tax

The concept of a VMT tax involves charging users for the number of miles traveled on roads within the state. VMT taxes have arisen as an alternative to the gas tax, which is a means for states to collect funding. In most states, the gas tax is applied on the cost of a gallon of gas when travelers fill their tanks at gas stations. The increased fuel efficiency of electric and hybrid vehicles has enabled some users to use roadways without contributing to the funding for roadway maintenance, for which costs are increasing (Caltrans 2016). Automakers will continue to improve the fuel efficiency of vehicle fleets, so these challenges will only become more difficult over time. A VMT fee is one way to collect appropriate taxes from all vehicles to gain sufficient funding for roadways and, potentially, to discourage excessive vehicle travel.

One way to charge users for the number of miles traveled is through odometer readings at yearly vehicle inspections. However, this policy assumes all miles traveled are within the state, and some users would be getting double charged if they traveled and purchased gas out of state. A VMT tax can be applied only within the state operating the program by sending Global Positioning System (GPS) data to calculate the number of miles traveled within the state by each vehicle. This can be accomplished by using either dedicated short-range communications (DSRC) or cellular communication to send the GPS data to a central database, where a public or private entity would calculate the amount of money owed by each driver.

California, Washington, and Oregon have started pilot programs to test the feasibility and efficacy of a VMT tax program. These programs track all miles driven on public roads and charge users accordingly. The California Road Charge Pilot program plans to analyze a variety of means for collecting road usage data, with and without the need for electronic vehicle location data (Caltrans 2016). Users can choose from four types of monitoring systems: time permit, mileage permit, odometer charge, and automated mileage reporting. The automated mileage reporting option requires in-vehicle equipment, which reports location data collected from vehicle telematics, smartphone apps, or OBD-II port devices (Caltrans 2016). An advantage of this more advanced option is that participants will not be charged for out-of-state or private road travel (Caltrans 2016). Enforcement of this advanced method can be somewhat challenging, since it would require vehicles to have operational hardware that has not been modified (to reduce toll totals). Participants would need to be randomly audited to ensure they are not misrepresenting their travel data to save money. The Oregon Department of Transportation (ODOT) has implemented a similar pilot, which involves actual payment rather than simulated payment, with a program called OreGO. The permanent program currently accepts 5,000 volunteers, who are also given an option between a GPS tracking and a series of non-tracking options such as odometer readings.

While VMT tax policies are currently in their infancy, they may become increasingly necessary with the rise of more fuel-efficient vehicles. Additionally, they enable more equitable charges for road usage for all types of vehicles. The development of pilot and permanent VMT fee programs that use GPS tracking could lay the foundation for the development of more advanced transportation management policies that would require this location and communication technology.

15.2.2. Cordon-Based Tolling

Cordon-based tolling involves charging users for entry into an enclosed area, commonly downtown business centers, to ease traffic at peak hours. Cordon-based congestion pricing has been used in cities around the world as a means to reduce congestion and emissions in urban centers.

Singapore, London, Stockholm, and Milan have all implemented some form of congestion pricing (Brown 2011). Singapore first introduce a manually-enforced Area Licensing Scheme in 1975, which charged drivers a flat fee to enter into the central business district during peak hours (ITDP 2015). Users showed their purchased license to enforcers at the gantries to ensure compliance. In 1998, Singapore replaced the manual Area Licensing Scheme with Electronic Road Pricing (ITDP 2015). This system requires installation of an in-vehicle unit with a smart card and a DSRC system. The Singapore cordon-based congestion pricing system has resulted in lower traffic volumes, higher average vehicle speeds, and lower carbon dioxide emissions (ITDP 2015). London also employs a cordon-based congestion charging system in the central downtown area between 7:00 AM and 6:30 PM on weekdays. Payments can be made at retail outlets or through electronic means on the same day, or users can purchase weekly, monthly, and annual passes (Litman 2011). Video cameras installed throughout the city record license plates, and the user pays a fine if they do not

pay for downtown road usage. Automobile usage has decreased, public transport usage has increased, and average vehicle speeds have increased in urban centers (Litman 2011). Stockholm and Milan have reported similar results (Croci 2016).

These successful programs around the world suggest that cordon-based congestion pricing is a viable and valuable program to implement in cities with large traffic volumes in dense urban centers. As seen in these examples, cordon-based congestion pricing can be implemented with different technologies that have been around for years. An advantage of this type of system is that installations are only needed at entry points to the congested area. While overall traffic volume may decrease, the users who choose to enter the restricted area may still choose routes that are suboptimal to the congestion of the roadway system. However, cordon-based pricing is an effective means to decrease the general travel volume within highly congested areas during peak hours.

15.2.3. Credit-Based Congestion Pricing (CBCP)

Credit-based congestion pricing (CBCP) involves charging road users a fee that accounts for the marginal cost of congestion they cause (Nie and Liu 2009). Current drivers make route decisions based on the shortest path or time to their destination, and these decisions do not take into account the externality of the cost of vehicle travel to the rest of the transportation network. CBCP adds this cost into the decision-making process, making users aware of their impact on the roadway congestion, and decreasing the volume of traffic along the most congested stretches of road (Kockelman and Kalmanje 2005). CBCP would require a more complex system than a VMT tax, as additional technology would need to be coordinated to communicate vehicle position to a central system and the pricing of travel to vehicle drivers or occupants. However, CBCP would more effectively accomplish the goal of changing user behavior to alleviate congestion, because it would incentivize more optimal route choice rather than incentivizing reduced total miles traveled. Such a system would require effective two-way communication, a fair pricing policy that attracts users, and an auditing procedure that ensures compliance.

CBCP requires communication of vehicle location and velocity data to a database, where vehicle speeds are used to evaluate the state of congestion along a given stretch of roadway. This information is then used to price routes. When certain routes are more congested, the price to travel along these routes increases. In order to alter user behavior and ensure fairness and transparency of the congestion pricing system, the toll operator will communicate the pricing of alternative routes to the users via DSRC or cellular communication. This information can be displayed on a smartphone or other device early enough to allow human operators to alter their route based on this information. The vehicle location data can indicate when individual automobiles pass checkpoints along a route in order to toll each user. Reliable communication and accurate location data are important to ensure the consistency and fairness of those tolls.

Based on the value of each individual's value of time and the time constraints of their travel, users can choose to take an alternative route in exchange for a lower cost or continue on the same path for a larger fee. While many people may choose to continue along their route and pay the fee,

others will be influenced by this charge and opt to take a different route or travel at off-peak times, which will alleviate congestion along the most congested roadways.

One major challenge with establishing a CBCP program is the attraction of users. Many citizens are averse to being tolled in any way. There would have to be sufficient incentives to the volunteers to encourage them to opt in. One way to do this is to provide users a tax deduction that would offset the cost of tolls collected through CBCP. With that monetary incentive, users would realize some value in joining the program.

One possible issue with congestion pricing is that lower income users and people with inflexible schedules could be tolled excessively (Gulipalli et al. 2008). Equity can be improved by allocating a flat budget to each individual to spend on congestion pricing over a certain period of time (Gulipalli et al. 2008). Gullipalli et al. (2008) detail more specific policy recommendations for effective CBCP management. It is essential that the policy be set appropriately to ensure efficiency, equity, and effectiveness.

15.3. Technology Solutions

Research has been conducted on the potential technology solutions for a congestion pricing system through a review of previously-published interviews and a series of expert interviews. Based on the information collected during this research, three leading concepts have been identified for use in a congestion pricing solution: video, DSRC, and cellular. Each of these solutions requires a different mix of technologies, and each has its own advantages and disadvantages. The specifications, cost, and value of each of these systems are discussed below.

15.3.1. Video-Based System

Video is one technology that could be employed to measure congestion and price routes accordingly. Video cameras are already installed in many locations along highways and at intersections, so these feeds could be harnessed to create a real-time model of traffic congestion. The system would consist of a series of video cameras on poles along major roadways, a data connection to send the information to a central system, and algorithms to analyze the video feed. This system would then need a means to communicate and toll users based on the pricing of each route. This could come through the DSRC or cellular networks previously discussed or through license plate recognition and electronic signs indicating the toll for upcoming routes.

The major infrastructure installations would be the camera, cable, and pole along the roadside. Installations including all three of these major components could cost $20,000-50,000, depending on the quality of the camera and pole height (Lange 2017). The camera can differ based on which features are included, such as the ability zoom and pan. The pole could be anywhere from 20 to 50 feet, and taller poles would allow for greater range but also would increase cost (Lange 2017). Based on the average range of cameras, one could be placed approximately every half-mile, depending on the road curvature, buildings, and other obstructions (Lange 2017). A large portion of the cost is the pole itself, and the individual video cameras themselves can be purchased for

$800-1,200 (Lange 2017). In order to toll individual users, the video feed would need to be of high enough quality to capture license plate numbers of passing vehicles. This may require multiple cameras at one location, or a very high-quality, high-speed camera. The processing of these characters from varying angles and speeds would also need to be incorporated into the software evaluating the video feed.

One major challenge with a video-based solution is that the pricing information cannot be communicated to travelers through the same system with which traffic data is collected. Communication of pricing to travelers is essential, as the goal of a congestion pricing system is to alter travel behavior to alleviate congestion. The DSRC or cellular solutions described in the following sections could be combined with the video feed for a comprehensive solution, but this would result in multiple expensive and somewhat redundant infrastructure investments. Alternatively, tolls could be implemented only at a limited number of locations and the pricing could be communicated via electronic signs on the side of the road or above highways. While this additional infrastructure investment limits the number of locations that tolls can be placed, it increases the number of users that can participate in the program because it requires no in-vehicle installation.

One advantage of a video-based solution is that the video infrastructure is already installed in many places in major cities. Another advantage of a video-based solution is the relative ease of obtaining higher levels of market penetration without every user needing a communication device in his/her vehicle. Despite these advantages, additional infrastructure to communicate the real-time pricing to users will be required. Such infrastructure can be prohibitively expensive if added everywhere, so it normally would be implemented in a limited number of locations. The challenge and cost of installing two separate systems for information collection and transmission ultimately render a video-based solution less viable.

15.3.2. DSRC-Based System

Another possible solution is a congestion pricing system that uses Dedicated Short-Range Communication (DSRC). DSRC is a spectrum of 75 MHz in the 5.9 GHz band that has been reserved for use in vehicle safety and mobility applications (ITS 2017). DSRC is currently being used in vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) applications to alert drivers of potential hazards, such as stopped traffic or collisions. The low latency communication of two-way messages makes DSRC useful in time-sensitive situations (ITS 2017). Fast communication is essential for safety applications such as crash avoidance, and it would also be useful for adaptive pricing schemes in which the cost of traveling certain routes changes often. Since the DSRC band is reserved for mobility applications, congestion pricing would be a useful allocation of this bandwidth.

A DSRC system will require roadside units (RSUs) installed along roadways, along with on-board units (OBUs) installed in vehicles. As vehicles pass the RSUs, a message is sent from the vehicle's OBU to the RSU indicating the vehicle's position and speed, and data from all vehicles' messages is compiled to model the amount of congestion in a certain area. With this information, incentives

for certain routes can be generated, and this information can be sent back to the vehicles' OBUs via RSU messaging. Vehicle operators decide which route(s) to take based on travel times and dynamic tolls. Using cloud-based tolling information, travelers can also delay their trips or choose different destinations and modes.

Currently, most vehicles on the road are not equipped with DSRC communication. However, DSRC is beginning to be incorporated into some new vehicles, and it is possible it will be required in all new vehicles in the future along with GPS. Some experts expect that both may be required by the National Highway Traffic Safety Administration (NHTSA) within the next 5-7 years (Sturgeon 2017). A mandate for DSRC to be included in all new vehicles was proposed in December 2016, but the plans were discontinued in November 2017 (Lowy 2017).

Conventional vehicles could take advantage of a congestion pricing system by adding DSRC connectivity through installation of an OBU. An on-board DSRC unit can be small, lightweight, and it can be mounted on the windshield of a vehicle with Velcro or other simple fasteners (Kapsch n.d.). An OBU would cost about $1,500 currently, but this price is likely to decrease as technology improves and production volume increases. OBUs can communicate position and speed, and relatively accurate traffic flow speed can be gathered from a limited number of vehicles. As the number of vehicles equipped with DSRC increases, the accuracy of this data and the benefits of a congestion pricing system will increase.

The other major component of a DSRC-based congestion pricing system is the installation of RSUs. RSUs have a line-of-sight range of about one kilometer. Due to the short range of DSRC RSUs, a high density of these devices would be required. Since communication is limited by line of sight, dense urban environments would require RSUs to be more compact or placed higher, with longer poles and leads. Billboards, buildings, and other objects could block the signal even within a short distance. Currently, RSUs are in the prototype stage and cost around $3,500. With improvements in technology and mass production, that price could go down to $500 to 800. In addition to the cost of producing the RSU, the installation and maintenance costs would add up quickly. The installation cost could vary from $1,000 to tens of thousands of dollars based on a variety of factors. Higher leads and poles for RSUs would cost more money. Connection to a communication network will also increase costs, especially if a data link backbone does not yet exist. RSUs will need routine maintenance for updates or replacement if weather or other external factors cause damage.

While a DSRC-based congestion pricing system would allow for fast communication between vehicles and infrastructure, it does require a large capital investment. DSRC communication is well-suited for transmitting small packets of data accurately in short periods of time. Pairing this communication with a smartphone or device for route decisions would enable an effective congestion pricing system. However, a DSRC-based system is limited by the cost of installing DSRC units both in vehicles and in dense urban environments. Furthermore, the installation and penetration of DSRC devices in infrastructure and vehicles will take a long time. For this reason, some experts believe that connected vehicles may leapfrog DSRC and go straight to using 5G

cellular communication. Benefits can be realized with the installation of a limited number of RSUs at highly congested areas, but the long time frame is an important consideration.

The large monetary and time investments make a DSRC-based congestion pricing system challenging to implement throughout an entire transportation network. However, DSRC solutions are viable for installation at major bottlenecks. A pilot program could be implemented on bridges or stretches of highways that are often highly congested at certain hours. Vehicles could be informed of an upcoming toll and given an alternate route option when passing the DSRC unit. This initial installation would allow testing of an adaptive tolling scheme and route choice data could be collected in response to congestion pricing

15.3.3. Cellular-Based System

A congestion pricing system could also be created with the use of cellular data. Information could be communicated via a smartphone or a device installed in the on-board diagnostic (OBD) port in the vehicle. Each of these solutions would take advantage of the already-widespread cellular network, but use different devices which each have distinct advantages and disadvantages.

A smartphone solution would require an app that would allow users to opt-in to the service. This acceptance of the agreement would allow the user's location to be tracked in order to toll users and gain information about traffic conditions. The communication to the cell tower and to the toll operator would be included in the user's cellular data service plan. This type of system would allow for faster market penetration because many people already own smartphones. Users could download the application that connects them to the congestion pricing system, rather than needing to install additional hardware. Location data would be collected from the phone's GPS and sent to the tolling entity. One potential issue is with the accuracy of the GPS currently installed in smartphones. Smartphone GPS is usually accurate enough to identify the road a user is on, but it can decrease in accuracy in dense urban environments. Smartphone GPS is not accurate enough to monitor lane-by-lane traffic reliably (Claudel 2017). While such a system could be implemented, there would likely be some issues with ensuring appropriate tolling if incorrect location information is used in determining a user's toll.

One potential solution to the location accuracy issue would be to combine the smartphone application with the installation of an inertial measurement unit (IMU) in the vehicle (Claudel 2017). An IMU is a single unit that incorporates an accelerometer and a gyroscope. The accelerometer measures the linear acceleration along three axes (University of Maryland n.d.). The gyroscope, also known as an angular rate sensor, outputs three signals describing the angular rate about each of the axes (University of Maryland n.d.). The IMU data allows the device to calculate its position based on the acceleration measurements, and it can bridge the gap between position estimates when the signal is blocked (Godha and Cannon 2005). While this does improve the location accuracy to improve the likelihood of fair congestion pricing, it also would require an additional installation, possibly deterring potential users.

The final cellular solution involves installing an OBD-II dongle in the onboard diagnostic (OBD) port. A dongle is a small electronic device that traditionally collects emissions and malfunction data (Moran and Baker 2016). Such a device could be configured to receive GPS location data and communicate using cellular data (Moran and Baker 2016). The dongle could be outfitted with a more accurate GPS system to improve the resolution of the congestion pricing system. A GPS unit with lane-by-lane accuracy would cost around $200, while one with road-level accuracy would be less than $50 (Dorfman 2017). The OBD-II dongle would also need a cellular communication modem. A mobile chip costs around $200 at low volume, but this price would decrease at higher volumes (Sturgeon 2017). The major issue with this data cost is determining who will pay the fee. Users may be willing to pay for the monetary or time benefit they gain from opting in to the program. Original equipment manufacturers (OEMs) may accept the cost in order to collect more data on the users. Departments of transportation (DOTs) could enter into agreements with cell carriers to provide this service to improve the efficiency or gather funding from their transportation network. The cost of a small data plan purchased at high volume by an OEM or DOT is estimated to be $3 to 4 per month (Dorfman 2017). This may increase at higher volumes of data communicated, but advances in technology could also decrease the cost of data. Alternatively, a third-party vendor may see an opportunity in providing the service and take on the cost of data communication.

The OBD-II dongle solution improves the problem of low-accuracy GPS included in current smartphones. This solution would allow for increased standardization and ensure greater fairness of a congestion pricing system. The use of OBD-II dongles does present some challenges, however. Users would need to purchase and install the hardware to enable this system, and they could unplug the device to avoid tolling. Additionally, the entity that would be willing to pay for the cellular connection is not clear, and sufficient incentives to encourage that additional cost would need to exist. Another issue is that the inclusion of OBD ports by OEMs is mandated by emissions standards, so many electric vehicles do not come equipped with the appropriate hardware (Dorfman 2017). So, if congestion pricing was implemented only through OBD installations, electric vehicles would either need to start including a similar port or their users could not participate in the congestion pricing program. Additionally, older vehicles predating the OBD requirements would not be able to use this program.

15.4. Additional Technology Considerations

15.4.1. 5G Network

While some level of congestion pricing could be implemented with current 4G or LTE cellular communication, the development of a 5G network will increase the effectiveness of congestion pricing. Applying congestion throughout an urban transportation network would put a large load on the current networks and may challenge the available bandwidth (Claudel 2017). While the development of a functioning, widespread 5G network is many years down the line, it will further improve the performance of connected vehicles (CVs) and congestion pricing.

A 5G network is expected to begin to be available 5-10 or more years from now, and there are a few major differences between 5G and current cellular communications. New, unlicensed frequencies of the electromagnetic spectrum, such as millimeter wave spectrum (> 25GHz), are expected to be released by the FCC for use in 5G networks (Andrews 2017). 5G would allow for information to pass between individuals and between vehicles without having to connect through a cell tower. Information about upcoming traffic, hazards, or road pricing on routes ahead could be passed backwards along sequences of vehicles on a roadway. Additionally, 5G will allow for high throughput (> 10 Gigabit per second per user) and low latency (< 1 ms RTT) communication (Fettweis 2015). Faster, larger data transfers can allow important, time sensitive travel information to be communicated more quickly and reliably. Vehicles can receive congestion, safety, and road pricing information in a timely manner, and the network will be able to handle the communication required for connected vehicles and congestion pricing more easily.

There are many challenges with the development and adoption of a 5G network. First, a business model must be developed for the distribution of a 5G network. The public value of safety-critical applications in CVs will provide value for government entities to invest in 5G. The private telecommunications sector will need to provide the service, however, and their investment will need to be profitable. Telecommunications companies could charge individual users, automobile OEMs, or government entities depending on the value to each of these groups. The ideal way in which to structure a profitable case for 5G is unclear at this point, but will need to be determined before benefits can be realized. Different people and companies differ in their opinions on 5G for CV applications, independently and in comparison to DSRC. GM has installed DSRC in some vehicle models, while Daimler (Mercedes-Benz) has focused more on preparing its vehicles for 5G (Sturgeon 2017). The debate between 5G and DSRC will continue, and it is important to stay informed about the developing value of each when considering CVs and congestion pricing.

15.4.2. Global Positioning System (GPS)

Accurate location information from global positioning systems (GPS) is key to an effective congestion pricing system. The accuracy of this data is important for obtaining a good understanding of the traffic conditions and tolling individual users fairly for road usage. The communication between satellites and GPS devices can often be interrupted or obstructed, especially in dense urban areas. This phenomenon often causes a wider location radius or inaccurate estimation of the user. High accuracy is important for congestion pricing, and there are varying types of GPS that offer different levels of accuracy.

Road-level accuracy is relatively easy to achieve with the current standard of GPS, and it should be sufficient for most forms of congestion pricing. Road-level accuracy would allow users to be tolled for travel on a certain route or stretch of road. Lane-level accuracy would enable greater precision and allow for specially assigned lanes, which could incentivize high-occupancy travel. While this would be a nice feature to add in some areas, it is not essential to effective congestion pricing.

There are four combinations of GPS satellites and technologies that carry increasing levels of accuracy. The standard GPS (SPS) included in most smartphones has 1-sigma accuracy of approximately 3 meters. The standard lane is also around 3-meters, so this provides enough accuracy for roads that include at least two lanes in either direction (Humphreys 2017). 31 SPS satellites are currently in orbit, providing sufficient coverage. With the addition of an antenna, wide area augmentation service (WAAS) enables 1-sigma accuracy of approximately 1.5 meters (Humphreys 2017). This service allows for nearly lane-level accuracy, but there would be significant potential error.

Additionally, the United States-built GPS system does not offer as wide a bandwidth or as high accuracy as the Galileo satellite system being built by the European Union. With WAAS and good visibility, the Galileo GPS offers 1-sigma accuracy of 1 meter (Humphreys 2017). This system is sufficiently accurate to collect lane-level accuracy but could present some issues in dense urban areas with poor visibility. There are currently 11 Galileo satellites in orbit, and 30 satellites are expected to be in orbit within five years. The ideal GPS system would be GPS L2C, which allows the GPS to communicate with a smartphone over Bluetooth. 19 of the 31 SPS satellites are currently equipped with L2C capability, and all 31 are expected to be L2C compatible within 5 years (Humphreys 2017). GPS L2C allows for 1-meter accuracy even in poor visibility, making lane-level congestion pricing possible even in dense urban centers. The current GPS systems are capable of road-level accuracy that would allow for some level of congestion pricing, and the advancement of GPS technology will allow for lane-level accuracy. While high accuracy is possible, the solutions do require an installation of a GPS antenna in addition to a smartphone or other device. So, the accuracy of location information is a challenge to congestion pricing, but current technology is sufficiently accurate for a basic system. With the correct systems in place, congestion pricing can be implemented fairly and accurately.

15.5. Privacy & Security

Privacy and security are major concerns when handling personal location data of a large pool of users. The privacy and security issues with each solution differ based on the method of data collection and communication used. These potential problems are important when evaluating the reliability and safety of congestion pricing applications.

For a video-based system, there is some concern about capturing images of users and non-users along roadways at all times. Monitoring users who do not opt in to the congestion pricing service seems to be a small invasion of privacy. However, cameras are installed along many roadways, and are not an illegal invasion of privacy in many places (Claudel 2017). While video cameras may cause some backlash from citizens who are especially concerned about privacy, the concern is not as great as applications using GPS location data.

For cellular and DSRC solutions that use GPS location data to track the routes of users, the privacy concern is greater. For these location-based applications, it would be essential to offer users the opportunity to opt in rather than mandating sharing of location information. Allowing the

government to handle personal location information at all times would likely deter some users. A private-sector, third party service provider could handle the data, which may ease the worries of some consumers, but data privacy would still be a concern.

Cellular communication would carry the same risks that current cellular service does. 3G has known security issues, and it can be spoofed relatively easily (Sturgeon 2017). 4G is more secure and is the most common technology insurance companies and OEMs currently use for vehicle monitoring (Sturgeon 2017). While location and speed information are anonymized for many DSRC safety applications, applications that toll individual users cannot be truly anonymous. Encryption and decryption of user information would be necessary to prevent hacking, and this would add to the overhead cost of implementing a congestion pricing system (Sturgeon 2017).

Malicious users could gain access to sensitive personal location information if the CP communications are not robustly designed or monitored properly. Concerns about people hacking into and assuming some control of automated vehicles are unlikely to be valid in CP applications, since in-vehicle installations for data remittance should be designed to be only "push" (rather than receive) information and should be partitioned from vehicle controls (Claudel 2017). In other words, as with connected vehicles, communications should and will be separated from vehicle control programs. Security and privacy are important concerns of a location-based congestion pricing application, and they must be priorities during implementation and operations. Fortunately, many systems exist, in Singapore and New York, Copenhagen and Southern California, with third party account managers and scrambled IDs providing meaningful privacy protections.

Furthermore, creating a centralized system for managing CP policy creates a single point of vulnerability that could be subject to attack on a system-wide level. To mitigate this, a CP system can be designed with great care to allow for decentralization, i.e. distributing the responsibilities of the system across multiple hardware units in multiple locations. In doing so, a system-wide attack becomes more difficult for a malicious agent, and any such attack would likely disable only a small portion of the system at any given time.

15.6. Compliance & Auditing

In order to ensure compliance with a congestion pricing system, an auditing process would need to be created. Users could tamper with the GPS location or communication devices in order to avoid toll payment. At the state level, vehicle inspections required by some states for registration offer the opportunity to ensure correct operation of the devices. If a congestion pricing user is not compliant with the required standards, he could be denied vehicle approval and the incident would be reported.

While inspection may catch some malfunctioning devices, users who are intentionally avoiding fees would likely fix their vehicles before taking them into registration. An auditing process with an external check on location could be added to the congestion pricing policy. Video cameras are one possible check on a vehicle's location. A few video cameras at major bottlenecks could capture vehicle license plates, and this information could be matched with the location data transmitted by

the vehicle. If the GPS data does not indicate the same vehicle location at the time and date the video was captured, the user would be noncompliant with the congestion pricing system. The vehicle would then be investigated for tampering, and a fine would be issued to the user of that vehicle if it is found to be illegally altered.

While it is not economically feasible to audit every vehicle regularly, a selection of vehicles could be audited periodically. A portion of license plate numbers would be chosen, and this number would be searched for in video footage. While this would not necessarily catch all people using GPS or communication jammers, it would likely deter people due to the chance of being caught and fined.

15.7. A Hardware Migration Path Forward

For an implementation of a policy such as is recommended in this paper to be successful, it is necessary that the implementation plan be as straightforward as possible. Consider the three facets necessary for a successful management scheme: impact measurement, agency communication, and driver interaction. This section proposes a potential migration path that improves these facets incrementally. With careful design, the technology required for the various schemas discussed in this paper can be made modular, allowing for the development of a migration plan such as is detailed below in Figure 2. This plan allows for the reuse of hardware until such point as it becomes obsolete, when it can be replaced by technology that can provide for more advanced management methodologies.

As illustrated in this figure, the most basic ("bare-bones") form of VMT discussed in this paper depends on an odometer keeping track of the miles driven by a vehicle, which informs the driver of how much of an impact their driving has and must be read in person by an agent. This method's drawbacks, discussed above, can be a significant disincentive for those that regularly drive in such a way that an odometer-based method would overcharge for their driving behavior.

To improve upon these flaws in impact management, a "headless" (i.e. without a display) GPS system can be implemented which would track users without any form of telecommunication equipment. The position tracking equipment can be of various degrees of accuracy at this point in such a plan, as lane-level accuracy is not yet needed (however, such hardware will be needed later in this process, and sufficient planning should allow for an easy upgrade to achieve this). In this "low-end" VMT scenario, an in-person reading of the vehicle's mileage is still necessary, but the accuracy of such a measurement will be much more reflective of a driver's impact on the road network. On the other hand, the drawback to this is that there is no method in place at this time for vehicle owners to be made aware of how many miles the GPS system has recorded until such time that a reading is taken, either by an agent or by the owner themselves.

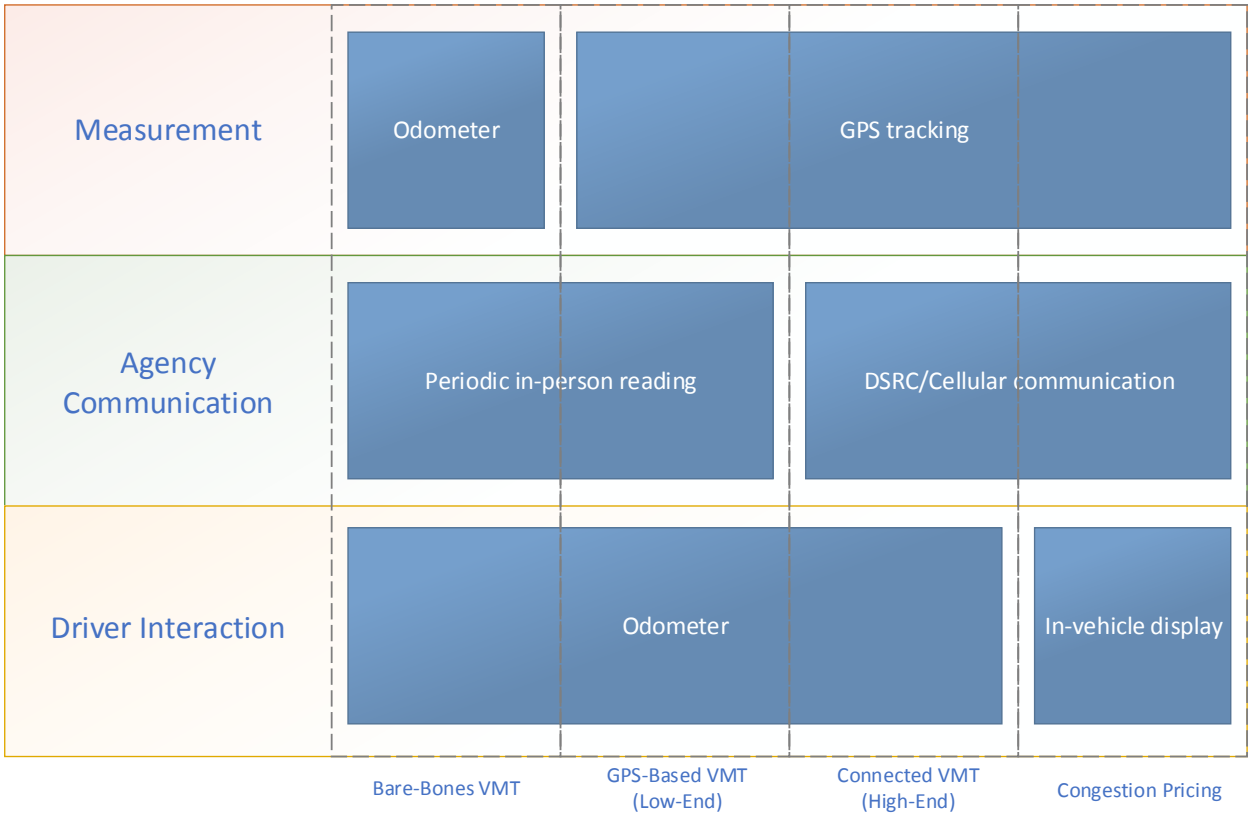


Figure 15.2 Technology Migration Plan for VMT and CBCP Tolling Schemas

The drawbacks of necessitating an in-person reading to discover the logged mileage are addressed in the next phase of this plan—the "high-end" VMT system. This next step provides for a telecommunications module (DSRC, cellular) such as is discussed above, thus allowing for agencies and vehicle owners to be updated regularly regarding a vehicle's mileage. This can be accomplished using an application or through regular email updates. These methodologies can limit the data usage needed based on the update frequency provided and can assist in automating the auditing process by reducing the human element in the process. However, these would not provide a manner for the driver to be updated in real time regarding their mileage, so the driver must use the odometer to measure their mileage approximately while mid-trip.

This drawback also prohibits a CBCP system, as drivers must be able to see road prices in real time. To address this, the final addition to the system is a display mounted in the vehicle that can provide pricing information to the driver as necessary for a CBCP implementation. Such a display removes the need for an old-fashioned odometer display, as this functionality can be handled by the GPS system and display. However, the amount of data transmitted through this system will be substantially higher than that of the VMT systems. Additionally, while the added accuracy of more advanced GPS modules as discussed above becomes imperative at this point, such an advanced module could be considered optional for the VMT phases as well and could be rolled out as resources become available.

15.8. Conclusion

VMT fees and CBCP are related, but they are separate types of policies. VMT fees primarily help state and local governments gain funding for roadways with declining revenue from the gas tax, with a small possible congestion benefit. Some users may opt to travel fewer miles with their VMT being monitored, but this does not alter the routes they would take. CBCP would be far more effective in alleviating congestion, as this policy is focused on route selection based on the congestion at a given time. However, the technologies and systems required for each program are similar.

VMT fees and CBCP could both be implemented with DSRC or cellular technology. VMT fees are a simpler system, so they could be implemented first. This would allow DOTs to gain additional revenue, gain experience with collecting vehicle location data, and identify potential compliance challenges. If DSRC units are placed along major roadways, the location data held in the vehicle can be transferred to the central database periodically to charge VMT fees. The funds collected from VMT fees can be used to improve roadways or invest in additional technology. With this experience, the same entities could move into implementing CBCP for additional benefits to the transportation network. The same DSRC units could be used to collect information on vehicle speed and location and communicate route pricing at these highly congested locations. If the CBCP program proves valuable, it could be expanded through cellular communication. VMT fees would be a good first step in technology-based roadway management, and CBCP could take advantage of the technology in place to further improve the efficiency of the transportation network.

15.8.1. Technology Recommendation

Analysis was conducted on the viability of DSRC, cellular, and video technologies for use in congestion pricing. The value of each of these technologies is based solely on its value for congestion pricing, rather than use in connected vehicles generally. The technologies were evaluated based on their effectiveness for this application, current level of market penetration, and the scalability throughout a transportation network. These criteria were evaluated on a scale representing their relative values. The ability of each technology to be applied to the major policies of VMT fees and CBCP was also taken into account. First, the effectiveness of each technology when applied to congestion pricing was considered. DSRC and cellular solutions are both similarly effective in transferring information to and from vehicles. Both systems are able to transfer small data packets known as basic safety messages (BSMs), which include vehicle location and speed. DSRC currently allows for lower latency communication, but this is not as important for congestion pricing as it is for vehicle safety applications because routing decisions are not as time-sensitive as collision avoidance maneuvers. Video can collect congestion information in order to price roads, but it lacks the ability to communicate information back to the users, which is required to change user behavior and reduce congestion. A video solution would need to be combined with electronic signage indicating the price of upcoming routes or with DSRC or cellular communication. This limits the scope of a video-based solution since it requires costly, redundant technology. Therefore, cellular or DSRC solutions will be most effective for the longer term.

Cellular technology is widespread in urban centers, as it employs cell towers that enable long-distance communication. The infrastructure is already in place for 4G communication, and it is currently used to transfer data between smartphones and other connected devices. Video cameras are installed along some stretches of roads and intersections, but they are not widely installed along roadsides throughout cities. DSRC is also not widely available, and RSUs would need to be installed densely along roadways. Additionally, both video and DSRC systems would require installations at short intervals along the roadways, while cellular communication has much longer range. Both DSRC units and video cameras are recommended about every half-mile, so installing these throughout a transportation network could be costly (Lange 2017).

DSRC is recommended to be implemented at locations with high congestion in the short term, as a pilot system. Bridges, major highways, and other commonly congested stretches of roads are terrific locations for such pilots. RSUs could be placed a mile or two before these major bottlenecks to communicate route and pricing options to arriving travelers, and then again at section entrance, to notify on-board devices of toll charges. The use of congestion pricing at each region's most congested points and corridors will encourage use of alternative routes, driving at off-peak times of day, and/or changes in trip destination, mode and generation decisions, in order to reduce the travel costs. In the long term, however, cellular systems will be more effective in tolling entire urban transportation networks. With the ability to toll large areas using cellular networks, congestion pricing can be effectively scaled to decrease congestion throughout rather than just a few key nodes.

Chapter 16. Traffic Flow Estimation Using Fast-Algorithms for Fast Forward Simulations of Macroscopic Traffic Flow Models

16.1. Introduction

Traffic flow models are commonly used to describe the propagation of traffic on transportation networks. Depending on the scale of the problem and on the type of traffic phenomena that need to be reproduced by the model, it is possible to identify three main classes of traffic flow models: microscopic, mesoscopic, and macroscopic (Peeta and Ziliaskopoulos, 2001). In macroscopic models, traffic is modeled as a fluid stream described by a density and flow function, defined on all points of a road network, and for all times. Macroscopic models encode both the propagation of traffic on network links (resulting in macroscopic link models), as well as the splitting and merging of vehicle flows at junctions (resulting in junction models, or node models). One of the most commonly used macroscopic link model used in the literature is the Lighthill–Whitham–Richards (LWR) model (Lighthill and Whitham, 1955; Richards, 1956). This model is based on two main assumptions: the conservation of vehicles and the existence of a univocal flow-density relationship (fundamental diagram). Assuming that links can be described by space-independent parameters (homogenous problem), the propagation of queues and shockwaves can be then modeled by means of a partial differential equation (PDE), known as the LWR PDE. The LWR model is often used for studies involving large simulations since it is relatively straightforward and robust, depending on a low number of model parameters that are easy to calibrate. Furthermore, its computational time that is independent of the number of vehicles to model (Wageningen-Kessel, 2016), unlike microscopic or mesoscopic models. Junction models have also been studied extensively to reproduce traffic behavior at merges/diverges (Daganzo, 1995), to investigate the propagation of kinematic waves (Garavello and Piccoli, 2006), and to identify general methods (Tampère et al., 2011; Flötteröd and Rohde, 2011; Jabari, 2016).

In the past two decades, a considerable number of numerical schemes have been proposed to solve the LWR model on networks, striving for higher computational efficiency and accuracy. The most popular ones include the cell transmission model (CTM) (Daganzo, 1994), a particular case of Godunov discretization (Godunov, 1959), and the link transmission model (LTM) (Yperman et al., 2006; Yperman, 2007), based on earlier work by Newell (Newell, 1993). Among the event-based numerical schemes, the wave-front tracking methods (Bressan, 2000; Garavello and Piccoli, 2006) reproduce the propagation of expansion waves and shocks using Riemann solvers and the Rankine Hugoniot formula (Baiti and Jenssen, 1998). Raadsen et al. (2016) propose another promising event-based algorithm suitable for large simulations, based on semi-analytical solutions of the LWR PDE. Event-based approaches can be very fast; however, their efficiency and accuracy depend on the initial and boundary conditions and flux function of the problem to solve.

Alternative computational methods are based on the Hamilton-Jacobi formulation of the LWR model. Dynamic programming (DP) methods (also referred as *variational principle* in the

transportation literature) solve a network least cost problem (through DP) on space-time grid, resulting in the so-called variational method (Daganzo, 2005; Daganzo 2006). Alternatively, the Lax-Hopf (LH) algorithm (Lax, 1957; Hopf, 1969) uses a specific structure of the DP problem to decompose the solution into the minimum of a finite number of explicit functions, resulting in an exact computational method to compute the solution on a single link, and a faster computational time than DP. Unfortunately, the LH algorithm does not perform well over large time horizons and is in general slower than most of the traditional link models.

In this study, we propose a modification of the LH algorithm, referred as Fast Lax-Hopf (FLH) to solve the LWR model more efficiently while retaining the exactness of the LH. We show that its computational performance is comparable to the LTM, which is used, together with the CTM and the original LH, as a benchmark for our study.

Given known initial conditions in all links of the network, and given traffic demand and supply functions at the boundaries of the network, the objective of the present algorithm is twofold. First, compute the boundary conditions (inflows and outflows) on each of the link as fast as possible. Second, once these boundary conditions are computed, the solution can be found at any point in space and time required by the particular problem (for example at a precise point of space and time where a measurement data point is generated for estimation problems, or at a given time horizon for forward simulation problems), by minimizing explicitly computed functions. For this present task, we restrict our algorithm to triangular fundamental diagrams, which are widely used in the literature. This specific fundamental diagram allows the original formulation of the LH algorithm to be considerably simplified.

The FLH algorithm, which is particularly suitable for network simulations, requires lower number of operations than the original version of LH without compromising its accuracy. Furthermore, we show that the FLH can be further simplified for specific space-time discretizations, while remaining exact. The resulting formulation shares similarities (in terms of the formulation) with the LTM, though it is slightly slower than the latter. Nonetheless, the solution computed by the FLH algorithm can be computed everywhere in the computational domain, for arbitrary initial conditions, unlike the former. In some situations, such as network loading problems with zero initial conditions, the LTM also provides a solution inside the computational domain, and retains a slight computational advantage over the FLH scheme.

The rest of this chapter is organized as follows. First, we describe some of the main computational methods available, and discuss their advantages and drawbacks. We then derive the FLH algorithm using a set of theorems that simplify the original LH formulation for triangular fundamental diagrams. We also show that the FLH algorithm can be further simplified for specific discretizations, and show that the resulting algorithm has a similar (but not identical) expression as the original LTM. In the second part of the chapter, we provide numerical validation of this algorithm by means of network traffic simulations, and comparisons with the original LH, CTM, and LTM formulations. Finally, we present some considerations and conclusions based on the results.

16.2. Background: Link Models

Network simulation algorithms require a link model to reproduce traffic flow on each link. In this study, we focus on computational methods that solve the LWR model with triangular fundamental diagrams on each link of the domain (the parameters of the fundamental diagram can change across the links of the network). In this section, after introducing the LWR model, we provide an overview of some of the main computational methods: LTM, CTM, variational theory, and LH (going through the details of each formulation is beyond the scope of the study).

16.2.1. The LWR Model and the Hamilton-Jacobi PDE

For a given time *t* and position *x*, we define the local traffic density *k(x,t)* as the number of vehicles per unit length, and the instantaneous flow *Q(x,t)* in vehicles per unit time. The conservation of vehicles on the highway is formulated as the following PDE (Lighthill and Whitman, 1956; Richards, 1956):

$$\frac{\partial k(t,x)}{\partial t} + \frac{\partial Q(t,x)}{\partial x} = 0 \tag{16.1}$$

In the LWR model, the fundamental diagram relates the flow and density; in this work, we consider the triangular fundamental diagram (Daganzo, 1994). The fundamental diagram is a positive and concave function defined on $[0,k_j]$ where k_j is the maximal density (jam density). It ranges between $[0,q_{max}]$ where q_{max} is the maximum flow (capacity). The triangular fundamental diagram, which is a concave function with derivative $Q'(k) = v$ (free flow speed) for $k < k_c$ (critical density) and $Q'(k) = w < 0$ (congested wave speed) for $k > k_c$. Hence, the triangular fundamental diagram is defined as follows:

$$Q(k) = \begin{cases} v\,k & : \ 0 \le k \le k_c \\ -w\left(k - k_j\right) & : \ k_c \le k \le k_j \end{cases} \tag{16.2}$$

Since the fundamental diagram is concave, it is continuous in the interior of its domain of definition, and therefore its parameters satisfy $vk_c = -w(k_c - k_j)$.

While the flow of traffic can be described by the density function $k(\cdot,\cdot)$, it can alternatively be described using the Moskovitz function $N(x,t)$ that expresses the cumulated vehicle count through a location *x*, at time *t*. The Moskowitz function (also called *cumulative number of vehicles* function) is defined as follows. All vehicles on and entering the road link are labeled by increasing integers as they pass the entry point x_0 of a highway section, and are assumed not to pass each other. The Moskowitz function at location x and time t is defined as $N(x,t) = n$, where n corresponds to the label of the vehicle closest to x at time t. The derivatives of the Moskowitz function are related to the density and flow functions (Daganzo, 2006).

Replacing *k* and *q* with *N* yields to the following Hamilton-Jacobi PDE (Newell, 1993; Daganzo, 2005a, 2006; Claudel and Bayen, 2010a, b):

$$\frac{\partial N(x,t)}{\partial t} - Q\left(-\frac{\partial N(x,t)}{\partial x}\right) = 0 \tag{16.3}$$

16.2.2. Computational Methods

The LWR PDE is a first order hyperbolic scalar conservation law that can be solved using a number of computational methods.

In the CTM both time and space are discretized, as each link is divided into a given number of cells of size Δx. This size is constrained by the Courant-Friedrichs-Lewy (CFL) condition (Bretti et al., 2006), according to which, for a given time discretization Δt the inequality $\Delta x \geq v\Delta t$ must hold, where v is the free flow velocity. The CTM is essentially a Godunov discretization of the original LWR equation, and assumes that the density of vehicles in each cell is constant across space. For every time interval the number of vehicles leaving a given cell and entering in the cell immediately downstream is computed using the Godunov flux. The maximum number of vehicles that can fit into a cell is a function of the jam density. The CTM requires calculating flows for all the cells of the link in order to compute the upstream and downstream boundary conditions of this link. In addition, the CTM does not yield exact solutions to the LWR model in general, due to numerical diffusion errors (Leclercq et al., 2007). The discretization in cells leads to an approximation in the speed of shockwaves that can propagate over the network, and ultimately can yield considerable cumulated errors. Several extensions of the CTM have been proposed to model other properties of traffic, such as the capacity drop (Schreiter et al., 2010; Srivastava and Geroliminis, 2013), different shapes of the fundamental diagram (Lo, 1999), and to reduce the discretization error (Daganzo, 1999; Szeto, 2008). Although the CTM allows to fairly reproduce important traffic phenomena like the forming and propagation of queues, the spatial discretization of links represents a main limitation in terms of efficiency and accuracy (Gentile, 2010).

Instead, the LTM only requires time to be discretized. The main feature of this model based the simplified theory of Newell (1993a; b) consists in using the characteristic speeds (free-flow and congested flow) to derive the upstream and downstream boundary conditions. Recently, a close variant of the original LTM formulation has been proposed by Himpe et al. (2016) to allow for larger time steps. In recent years, the LTM has become very popular for the dynamic network loading procedure within the dynamic traffic assignment (DTA), where simulations can involve thousands of links, and where the solution only needs to be computed on the link boundaries. However, a limitation of the LTM is that the solution cannot be computed inside each link, which makes it unsuitable to problems involving estimation and calculation of traffic indicators inside the links (e.g., in estimation, traffic optimization or control problems). In some specific situations, in which no expansion wave is present (for example, in a constant initial density scenario) the LTM allows computation of the solution inside the link, though this procedure does not converge towards the solution of the LWR model for general initial conditions.

The Variational Theory introduced by Daganzo (2005) consists in applying DP to solve the Hamilton-Jacobi PDE [Equation (16.3)] through the classical LH formula. The solution can

equivalently be computed using the viability theory (Aubin et al., 2008). Both approaches are similar, with the exception that the viability approach allows more general (discontinuous) initial conditions to be considered, and allows the computation of lower-constrained solutions to the Hamilton-Jacobi PDE.

The LH algorithm exploits a particular structure of the DP problem used in the Variational Theory to compute the solutions more efficiently in the case where the fundamental diagram is space and time independent. In this situation, the solution can be obtained without discretizing the computational domain, and it corresponds to the minimum of a finite number of functions associated to the initial and boundary conditions. By definition, this method is analytical and yields exact results in simulations of single links. In the network simulations errors can occur due to the temporal discretization of the boundary conditions, since boundary conditions are not necessarily constant over a given time step.

Because it uses an additional structure of the DP problem, the LH algorithm is always faster than the Variational Theory (although it has less general applications since it cannot handle situations in which the fundamental diagram depends on space or time). Nevertheless, its computational performance is comparable (if not worse) to that of the CTM (Mazare et al., 2011) and thus offers no speed improvement over the abovementioned algorithms (unless the boundary conditions are known in advance).

The FLH described in the following section allows one to compute solutions (at the boundaries) with lower computational requirements than the original LH. Its performance is similar to the LTM, but, unlike the former, it converges towards the solution of the LWR model inside the link, in network simulations, and for any given initial conditions. We achieve this by proving that some initial condition blocks appearing in the minimization problem (considered in the original LH) can be neglected to save time, without affecting the results. We demonstrate that these excluded blocks cannot theoretically influence the solution, and thus, the solution computed by this algorithm remains exact (for single link problems), as in Mazare et al. (2011). Once the sets of upstream and downstream boundary conditions have been derived, they can be used to solve Equation (16.3) in any point of the computational domain, without relying on a computational grid like in the original LH. This a particularly important aspect for estimation and control applications. For example, in estimation problems, one only needs to compute the solution on the space-time points corresponding to sensor measurements, which are in general considerably less than the total number of grid points (assuming a uniform grid in space and time).

16.3. FLH Algorithm for Computing Solutions to the LWR Model on Networks

In this section, we describe the main features of the LH algorithm used to compute the solutions of the LWR model semi-analytically. We then derive the FLH algorithm using a set of theorems that can be used to reduce the number of calculations compared with the original LH algorithm.

16.3.1. The Generalized LH Formula and Boundary Conditions

Let a value condition function $c(\cdot,\cdot)$ be defined. This value condition can encode for example initial and boundary conditions. Aubin et al. (2008) showed that the solution associated with the value condition $c(\cdot,\cdot)$, denoted here by $N_c(\cdot,\cdot)$, is the solution to the following optimization problem involving the value condition:

$$N_c = inf\{c(t-T, x-Tu) + TR(-u)\} \tag{16.4}$$

$$s.t.\,(u,T) \in [w,v] \times \mathbb{R}_+ \; and \; (t-T, x-Tu) \in Dom(c)$$

The value condition $c(\cdot,\cdot)$ corresponds to initial, upstream, and downstream boundary condition functions:

$$c(x,t) = \begin{cases} N_{ini}(x) & t = 0 \\ N_{up}(t) & x = x_0 \\ N_{down}(t) & x = x_n \end{cases} \tag{16.5}$$

The optimization problem (4) involves the function $R(\cdot)$, which is defined as the convex transform associated with the fundamental diagram $Q(\cdot)$:

$$R(u) = \sup_{k \epsilon [0, k_j]} (Q(k) - u \cdot k) \tag{16.6}$$

Equation (16.4) is well known in the Hamilton-Jacobi literature and often referred to as Lax-Hopf (LH) formula (Lax, 1973; Evans, 1998; Daganzo, 2006; Aubin et al., 2008; Claudel and Bayen, 2010 a,b).

Assuming a triangular fundamental diagram (2), the calculation of its convex transform $R(\cdot)$ yields:

$$\forall u \in [-w, v], R(u) = k_c(v-u) \tag{16.7}$$

The LH algorithm assumes that the initial and boundary conditions $c_{ini}(\cdot,\cdot), c_{up}(\cdot,\cdot)$ and $c_{down}(\cdot,\cdot)$ are piecewise linear (Mazare et al. 2011), and can thus be written as:

$$c_{ini}(0,x) = c_{ini}^i(x) = -k_i x + b_i \quad if \; x_i \le x \le x_{i+1}$$

$$c_{up}(x_0,t) = c_{up}^j(t) = q_j t + d_j \quad if \; t_j \le t \le t_{j+1}$$

$$c_{down}(x_L,t) = c_{down}^j(t) = p_j t + c_j \quad if \; t_j \le t \le t_{j+1}$$

In this situation, the solutions associated with the $c_{ini}^i(x)$, $c_{up}^j(t)$ and $c_{down}^j(t)$ can be computed explicitly (Appendix 1). The solution at any point (t,x) of the space time domain can then be computed by taking the minimum of the solutions taken in (t,x) and associated with each initial

and boundary condition block. This comes from the inf-morphism property, initially derived in Aubin et al. (2008).

16.3.2. Fast Lax-Hopf Algorithm

The primary objective of the proposed algorithm is to quickly compute the outflows and inflows at every time step, by using a minimum number of operations, and maintaining exactness. Once the boundary conditions are known on all links, the solutions inside the computational domain can be found by minimizing a number of explicitly computed functions. The FLH algorithm speeds both the computation of the boundary conditions, and the computation of the solution inside the computational domain.

This algorithm relies on the specific structure of the partial solutions to the Hamilton-Jacobi PDE (Eq. 16.3) with triangular fundamental diagrams. From (Claudel and Bayen, 2010a, b), the partial solutions associated with affine blocks are convex (this property is valid for any concave fundamental diagrams) functions of (t, x). Furthermore, (Daganzo 2005) showed that these solutions are Lipschitz continuous on their domain of definition for general diagrams. In the case of a Triangular diagram, it is easy to verify from the expression of the solutions (given in Appendix 1) that these solutions are indeed Lipschitz continuous.

Furthermore, the partial solutions associated with linear initial or boundary conditions, for a triangular fundamental diagram, are piecewise linear functions of space and time. This property is very important in the present situation, and would not be true for example in the case of a Greenshields fundamental diagram.

In the present case, we consider a general mixed initial-boundary condition problem on a given stretch of highway limited by upstream and downstream boundaries. We also assume that the boundary conditions that apply on the domain are not known in advance, unlike in the LH case. These boundary conditions have to be computed at each time step through junction models relating the demands of the incoming links to the supplies of the outgoing links, across each junction. These junction models have the effect of coupling the solutions computed over adjacent links. To compute these boundary conditions, our objective is to compute the inputs to the junction models as fast as possible. These inputs are upstream demands and downstream supplies of each link (for a given time step).

More precisely, let the initial condition be expressed as a piecewise linear function, with each linear piece on intervals (x_i, x_{i+1}) defined by:

$$c_{ini}^{i}(x) = \begin{cases} -k_i x + b_i & : x_i \le x \le x_{i+1} \\ +\infty & : otherwise \end{cases} \tag{16.8}$$

Where $i \in \{0, \dots, n_{ini} - 1\}$. As described in (Daganzo, 2006), the initial condition has to satisfy some growth and continuity conditions, which can be expressed as:

$$0 \le k_i \le k_j \text{ for all } i \in \{0, \dots, n-1\} \tag{16.9}$$

$$-k_i x_i + b_i = -k_{i+1} x_i + b_{i+1}\ , \forall i \in \{1, \dots, n-1\} \tag{16.10}$$

The associated solutions on the upstream boundary are defined by:

$$\text{If } \ k_i \le k_c,\ N_{c_{ini}^i}(x,t) = k_c(tv - x_0) + b_i + x_i(k_c - k_i)\ : \frac{x_i - x_0}{w} \le t \tag{16.11}$$

$$\text{If } \ k_i \ge k_c, N_{c_{ini}^i}(x,t) = \begin{cases} k_i(tw - x_0) - tk_j w + b_i & : \frac{x_i - x_0}{w} \le t \le \frac{x_{i+1} - x_0}{w} \\ k_c(tw - x_0) - tk_j w + x_{i+1}(k_c - k_i) + b_i & : \frac{x_{i+1} - x_0}{w} \le t \end{cases} \tag{16.12}$$

The associated solutions on the downstream boundary are defined by:

$$\text{If } \ k_i \le k_c \quad N_{c_{ini}^i}(x_L,t) = \begin{cases} k_i\left(tv - x_{n_{ini}}\right) + b_i & : \frac{x_{n_{ini}} - x_{i+1}}{v} \le t \le \frac{x_{n_{ini}} - x_i}{v} \\ k_c\left(tv - x_{n_{ini}}\right) + b_i + x_i(k_c - k_i) & : \frac{x_{n_{ini}} - x_i}{v} \le t \end{cases} \tag{16.13}$$

If

$$k_i \ge k_c \qquad N_{c_{ini}^i}\left(x_{n_{ini}}, t\right) = k_c\left(tw - x_{n_{ini}}\right) - tk_{jam} w + x_{i+1}(k_c - k_i) + b_i : \frac{x_{n_{ini}} - x_{i+1}}{v} \le t \tag{16.14}$$

By the inf-morphism property (Mazare et al, 2011), the solution $N(x,t)$ associated with the Hamilton-Jacobi PDE (3) with triangular fundamental diagram [as in Equation (16.2)] can be computed at any point (x,t) of the space-time domain by the following formula:

$$N(x,t) = \min\left(\min_i N_{c_{ini}^i}(x,t), N_{c_{up}{}^j}(x,t), N_{c_{down}}^k(x,t)\right) \tag{16.15}$$

In Equation (16.15), the indices j and k are respectively the index of the upstream boundary condition and the index of the downstream boundary condition influencing the chosen point (x,t), defined by $j = \left\lfloor \frac{t - \frac{x - x_0}{v}}{\Delta t} \right\rfloor$ and $k = \left\lfloor \frac{t - \frac{x_{n_{ini}} - x}{w}}{\Delta t} \right\rfloor$, where the $\lfloor a \rfloor$ denotes the floor of a (largest integer that is lower or equal to a). Note that if $j < 0$ (respectively $k < 0$), the solution $N(x,t)$ does not depend upon the upstream (respectively downstream) boundary condition.

To compute the upstream boundary block for a given time interval $[t, t + \Delta t]$, we first need to determine the supply of this particular link over the time interval $[t, t + \Delta t]$, defined by $s(t, t + \Delta t) = \frac{N(x_0, t + \Delta t) - N(x_0, t)}{\Delta t}$. The actual flow over the time interval $[t, t + \Delta t]$ is then determined using the other demand and supplies acting on this junction, through the junction model.

Hence, assuming that $N(x_0, t)$ is known, and using the classical LH algorithm (Mazare et al., 2011), we can compute $N(x_0, t + \Delta t)$ as:

$$N(x_0, t+\Delta t) = \min(min_{i\le j\le n_{ini}} N_{c_{ini}^j}(x_0, t+\Delta t), N_{down}^{\left[t+\Delta t-\frac{x_{n_{ini}}-x_0}{w}\right]}(x_0, t+\Delta t), N(x_0, t+\Delta t) + v\cdot k_c \cdot t) \quad (16.16)$$

The process required to compute the upstream boundary condition block at time $t+\Delta t$ is shown in Figure 16.1.

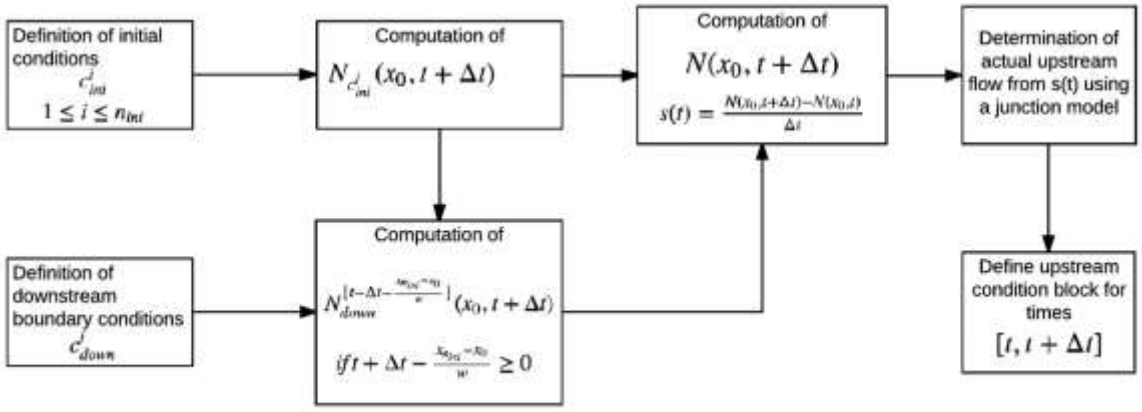


Figure 16.1: Required operations to determine the entering flow (upstream) over the time interval $[\boldsymbol{t}, \boldsymbol{t}+\boldsymbol{\Delta t}]$ *using the classical LH algorithm*

Equation (16.15) requires the minimization of $(n_{ini}+1)$ explicitly computed functions to derive the upstream supply of the link. The objective of the Fast Lax Hopf algorithm is instead to decrease the required number of operations, while still computing the average demand and supply functions exactly.

To decrease the set of required operations to compute the solution at the boundaries of the computational domain, we shall introduce four theorems: Theorems 1 through 4.

Theorem 1: Let a set of n_{ini} initial conditions be defined as in Equation (16.8), with Lipschitz continuity constraints from Equations (16.9) and (16.10). Let us further assume that $N_{c_{ini}^i}(x_{n_{ini}}, t') \le N_{c_{ini}^j}(x_{n_{ini}}, t')$ for a time $t' \ge \frac{x_{n_{ini}}-x_{i+1}}{v}$, with $i<j$. Then:

$$\forall\ t \ge t_s, N_{c_{ini}^i}(x_L, t) \le N_{c_{ini}^j}(x_{n_{ini}}, t) \quad (16.17)$$

Proof: using the structure of the solutions $N_{c_{ini}^j}(x_{n_{ini}}, t)$, we have that $N_{c_{ini}^j}(x_{n_{ini}}, t) = N_{c_{ini}^j}(x_{n_{ini}}, t') + (t'-t)v\,k_c$ if $t' \ge \frac{x_{n_{ini}}-x_{i+1}}{v}$ and $i<j$, irrespective of the value of k_j. We also have that: $N_{c_{ini}^i}(x_{n_{ini}}, t) \le N_{c_{ini}^i}(x_{n_{ini}}, t') + (t_s-t)v\,k_c$. Since $N_{c_{ini}^i}(x_{n_{ini}}, t') \le N_{c_{ini}^j}(x_{n_{ini}}, t')$, we have that: $\forall\ t \ge t', N_{c_{ini}^i}(x_{n_{ini}}, t) \le N_{c_{ini}^j}(x_{n_{ini}}, t)$.

This theorem implies that at the downstream boundary, if the solution associated to a particular initial condition piece *i* is lower than the solution associated with another piece of initial condition

j (with *i*<*j*), for a time *t*, then the solution associated with piece *j* cannot influence the solution (at the downstream boundary) for subsequent times. Indeed, the solution associated to the piece *j* is always greater or equal than the solution associated with piece *i*. Hence, the piece $N_{c_{ini}{}^j}(x,t)$ can be ignored from Equation (16.15) for times greater than *t*.

Theorem 2: Let a set of n_{ini} initial conditions be defined as in (8), with Lipschitz continuity constraints from Equations (16.9) and (16.10). Let us further assume that $N_{c_{ini}^i}(x_0,t_s) \leq N_{c_{ini}^j}(x_0,t'')$ for some $t'' \geq -\frac{x_0-x_i}{w}$, with $i > j$. Then:

$$\forall\, t \geq t'', N_{c_{ini}^i}(x_0,t) \leq N_{c_{ini}^j}(x_0,t) \tag{16.18}$$

Proof: using the structure of the solutions $N_{c_{ini}^j}(x_0,t)$, we have that $N_{c_{ini}^j}(x_0,t) = N_{c_{ini}^j}(x_0,t'') + (t''-t)v\,k_c$ if $t'' \geq -\frac{x_0-x_i}{w}$ and $i > j$, irrespective of the value of k_j. We also have that $N_{c_{ini}^i}(x_0,t) \leq N_{c_{ini}^i}(x_0,t'') + (t''-t)v\,k_c$. Hence, we have that $\forall\, t \geq t'', N_{c_{ini}^i}(x_0,t) \leq N_{c_{ini}^j}(x_0,t)$.

This theorem implies that at the upstream boundary, if the solution associated to a particular initial condition piece *i* is less than the solution associated with another piece of initial condition *j* (with *j*<*i*), for a time t'', then the solution associated with piece *j* cannot influence the solution (at the upstream boundary) for subsequent times. Indeed, the solution associated to the piece *j* is always greater or equal than the solution associated with piece *i*. Hence, the piece $N_{c_{ini}{}^j}(x,t)$ can be ignored in Equation (16.15) for times $t \geq t''$.

Theorem 3: Let $t_M = \lceil \frac{x_{n_{ini}}-x_0}{w} \rceil$. For any time $t > t_M$, we have that

$$\forall\, i \in \{1, n_{ini}\}, N_{c_{ini}^i}(x_0,t) \geq \min(N_{c_{up}{}^j}(x_0,t), N_{down}{}^k(x_0,t))\ , \tag{16.19}$$

Where $j = \left\lfloor \frac{t-\frac{x_{n_{ini}}-x_0}{v}}{\Delta t} \right\rfloor$ and $k = \left\lfloor \frac{t-\frac{x_{n_{ini}}-x_0}{w}}{\Delta t} \right\rfloor$

Proof: for any $t \geq t_M$, we have that $N_{c_{ini}^i}(x_0,t) = N_{c_{ini}^i}(x_0,t_M) + (t-t_M)v\,k_c$, irrespective of the value of k_i and b_i. Since $N(x_0,t_M) = \min\left(\min_i N_{c_{ini}^i}(x_0,t_M), N_{c_{up}{}^j}(x_0,t_M), N_{down}{}^k(x_0,t_M)\right)$, we have in particular that $N_{c_{ini}^i}(x_0,t_M) \geq N(x_0,t_M)$ for all $i \in \{1, n_{ini}\}$. We also have by Lipschitz continuity of the upstream boundary condition that $N_{c_{up}{}^j}(x_0,t) \leq N(x_0,t_M) + (t-t_M)vk_c$. Hence, we have that $N_{c_{up}^j}(x_0,t) \leq N_{c_{ini}^i}(x_0,t)$, which concludes the proof.

This theorem implies that the initial condition blocks has no influence on the upstream condition after time t_M, and can thus be discarded for computations of the upstream boundary condition past t_M.

Theorem 4: Let $t_N = \lceil \frac{x_{n_{ini}} - x_0}{v} \rceil$. For any time $t > t_N$, we have that

$$\forall\, i \in \{1, n_{ini}\}, N_{c_{ini}^i}(x_{n_{ini}}, t) \geq \min(N_{c_{up}{}^j}(x_{n_{ini}}, t), N_{down}{}^k(x_{n_{ini}}, t)) \,, \tag{16.20}$$

Where $j = \left\lfloor \frac{t - \frac{x_{n_{ini}} - x_0}{v}}{\Delta t} \right\rfloor$ and $k = \left\lfloor \frac{t - \frac{x_{n_{ini}} - x_0}{w}}{\Delta t} \right\rfloor$

Proof: for any $t \geq t_N$, we have that $N_{c_{ini}^i}(x_{n_{ini}}, t) = N_{c_{ini}^i}(x_{n_{ini}}, t_N) + (t - t_N) v\, k_c$, irrespective of the value of i, k_i and b_i. Since $N(x_{n_{ini}}, t_N) = \min\left(\min_i N_{c_{ini}^i}(x_{n_{ini}}, t_N), N_{c_{up}{}^j}(x_{n_{ini}}, t_N), N_{down}{}^k(x_{n_{ini}}, t_N) \right)$, we have in particular that $N_{c_{ini}^i}(x_{n_{ini}}, t_N) \geq N(x_{n_{ini}}, t_N)$ for all $i \in \{1, n_{ini}\}$. We also have by Lipschitz continuity of the upstream boundary condition that $N_{c_{up}{}^j}(x_{n_{ini}}, t) \leq N(x_{n_{ini}}, t_N) + (t - t_N) v k_c$. Hence, we have that $N_{c_{up}^j}(x_{n_{ini}}, t) \leq N_{c_{ini}^i}(x_{n_{ini}}, t)$, which concludes the proof.

This theorem implies that the initial condition blocks has no influence on the downstream condition after time t_N, and can thus be discarded for computations of the downstream boundary condition past t_N.

Using the above four theorems, we can significantly reduce the number of operations required to compute the solution at both the upstream and downstream boundary. This extension of the LH algorithm for computing the evolution of the upstream boundary condition (a similarly of the downstream boundary condition) is illustrated in Figure 16.2 and corresponds to:

$$N(x_0, t + \Delta t) = \min(min_{i \leq j \leq n_{i,max}} N_{c_{ini}^j}(x_0, t + \Delta t), N_{down}^{\left\lceil t + \Delta t - \frac{x_{n_{ini}} - x_0}{w} \right\rceil}(x_0, t + \Delta t), N(x_0, t + \Delta t) + v \cdot k_c \cdot t) \tag{16.21}$$

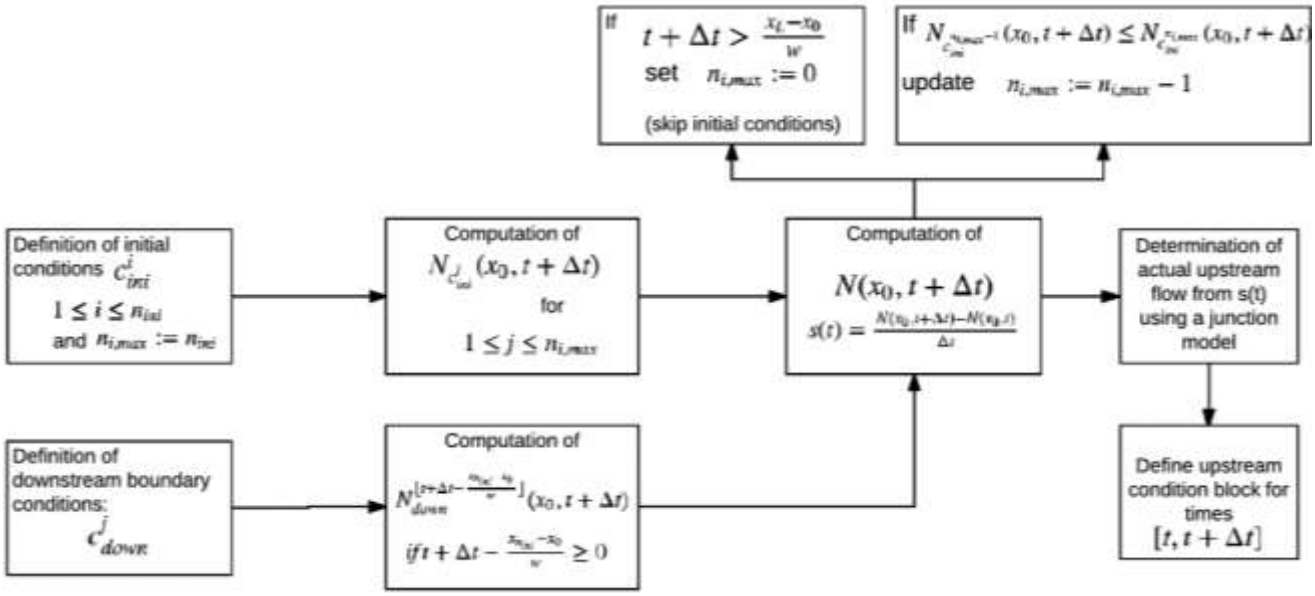


Figure 16.2: Required computations to determine the entering flow (upstream) during the interval $[\boldsymbol{t}, \boldsymbol{t} + \boldsymbol{\Delta t}]$ *according to the FLH algorithm*

Furthermore, the number of required operations required to compute the solution at an arbitrary point (t, x) of the computational domain can also be reduced, using the following two theorems: Theorems 5 and 6.

Theorem 5: Let a set of n_{ini} initial conditions be defined as in (8), with Lipschitz continuity constraints from Equations (16.9) and (16.10). Let us further assume that $N_{c_{ini}^j}(x, t_s) \leq N_{c_{ini}^i}(x, t_s)$ for a time $t_s \geq \frac{x_{i+1} - x}{w}$, with $i < j$. Then:

$$\forall\, t \geq s, N_{c_{ini}^j}(x, t) \leq N_{c_{ini}^i}(x, t) \tag{16.22}$$

Proof: using the structure of the solutions $N_{c_{ini}^j}\left(x_{n_{ini}}, t\right)$, we have that $N_{c_{ini}^j}(x, t) \leq N_{c_{ini}^j}(x, t_s) + (t_s - t) v\, k_c$ if $t_s \geq \frac{x_{i+1} - x}{w}$ and $i < j$, irrespective of the value of k_j. We also have that $N_{c_{ini}^i}(x, t) = N_{c_{ini}^i}(x, t_s) + (t_s - t) v\, k_c$. Since $N_{c_{ini}^i}(x_L, t_s) \leq N_{c_{ini}^j}\left(x_{n_{ini}}, t_s\right)$, we have that $\forall\, t \geq s, N_{c_{ini}^j}\left(x_{n_{ini}}, t\right) \leq N_{c_{ini}^i}\left(x_{n_{ini}}, t\right)$.

This theorem implies that inside the computational domain, if the solution associated to a particular initial condition piece *j* is lower than the solution associated with another piece of initial condition i (with $i < j$), for a location x and time s such that $t_s \geq \frac{x_{i+1} - x}{w}$, then the solution associated with piece i cannot influence the solution (at the same location) at subsequent times.

Theorem 6: Let a set of n_{ini} initial conditions be defined as in (8), with Lipschitz continuity constraints from Equations (16.9) and (16.10). Let us further assume that $N_{c_{ini}^{j}}(x, t_V) \leq N_{c_{ini}^{i}}(x, t_V)$ for some $t_V \geq \frac{x-x_i}{v}$, with $i > j$. Then:

$$\forall\, t \geq t_V, N_{c_{ini}^{j}}(x_0, t) \leq N_{c_{ini}^{i}}(x_0, t) \tag{16.23}$$

Proof: using the structure of the solutions $N_{c_{ini}^{j}}(x, t)$, we have that $N_{c_{ini}^{i}}(x, t) = N_{c_{ini}^{j}}(x, t_V) + (t_V - t)v\, k_c$ if $t_V \geq \frac{x-x_i}{v}$ and $i > j$, irrespective of the value of k_j. We also have that $N_{c_{ini}^{j}}(x, t) \leq N_{c_{ini}^{j}}(x, t_V) + (t_V - t)v\, k_c$. Hence, we have that $\forall\, t \geq t_V, N_{c_{ini}^{j}}(x_0, t) \leq N_{c_{ini}^{i}}(x_0, t)$.

This theorem similarly allows us to exclude a priori some terms from Equation (16.16), to speed up computations inside the computational domain.

16.3.3. Formulation of the FLH Algorithm for Specific Spatio-temporal Discretizations

In this section, we assume that the domains of the initial condition satisfy $x_i = x_0 + i\Delta x$ (where $i \in N$), that is, that the initial conditions are piecewise constant on domains of constant size Δx. We further assume that the space and time steps satisfy the CFL condition: $\Delta t \leq \frac{\Delta x}{v}$. In this situation, we can prove the following two theorems (Theorems 7 and 8), which further simplify the computation of the solution at the upstream and downstream boundaries:

Theorem 7: Let a set of n_{ini} initial conditions be defined as in (8), with Lipschitz continuity constraints from Equations (16.9) and (16.10). Let us further assume that $x_i = x_0 + i\Delta x$ and $\Delta t \leq \frac{\Delta x}{v}$. For any discrete time $t = i \cdot \Delta t, i \in N$ we have that:

$$N(x_0, t) = \begin{cases} \min\left(N_{c_{ini}^{l}}(x_0, t), N_{c_{ini}^{l-1}}(x_0, t), N_{c_{up}{}^{j}}(x_0, (i-1)\Delta t) + v \cdot k_c \cdot \Delta t\right) \; if\; t \leq \frac{x_{n_{ini}} - x_0}{w} \\ \min\left(N_{c_{up}{}^{j}}(x_0, (i-1)\Delta t) + v \cdot k_c \cdot \Delta t, N_{down}{}^{k}(x_0, t)\right) \qquad else \end{cases} \tag{16.24}$$

where $j = i - 1$, $k = \left\lfloor \frac{t - \frac{x_{n_{ini}} - x_0}{w}}{\Delta t} \right\rfloor$, $l = \left\lfloor \frac{wt}{\Delta x} \right\rfloor$

Proof: The first case corresponds to the situation where only initial components and upstream boundary condition components can influence the upstream condition ($t \leq \frac{x_{n_{ini}} - x_0}{w}$). In this situation, we have that $N_{c_{ini}^{k}}(x_0, t) = +\infty$ if $k > l$. Hence, we can write that $N(x_0, t) =$

$\min\left(N_{c_{ini}^0}(x_0,t),\dots,N_{c_{ini}^{l-1}}(x_0,t),N_{c_{ini}^l}(x_0,t),N_{c_{up}{}^j}(x_0,(i-1)\Delta t)+v\cdot k_c\cdot\Delta t\right)$. However, by the structure of the initial condition solution components (12), we have that for any $k\in[0,l-2]$, $N_{c_{ini}^k}(x_0,t)=N_{c_{ini}^k}(x_0,(i-1)\Delta t)+vk_c(t-(i-1)\Delta t)$. By the inf-morphism property $N_{c_{ini}^k}(x_0,(i-1)\Delta t)\geq N(x_0,(i-1)\Delta t)$, and thus, since $N(x_0,t)\leq N(x_0,(i-1)\Delta t)+k_cv(t-(i-1)\Delta t)$, we have that $N(x_0,t)\leq N_{c_{ini}^k}(x_0,t)$ for any $k\in\{0,\dots,l-2\}$, which shows that only $N_{c_{ini}^{l-1}}$ or $N_{c_{ini}^l}$ can influence the solution in (x_0,t).The proof of the second case is the result of Theorem 3.

Theorem 8: Let a set of n_{ini} initial conditions be defined as in Equation (16.8), with Lipschitz continuity constraints from Equations (16.9) and (16.10). Let us further assume that $x_i=x_0+i\Delta x$ and $\Delta t\leq\frac{\Delta x}{v}$. For any discrete time $t=i\cdot\Delta t$ $(i\in N)$ we have that:

$$N\left(x_{n_{ini}},t\right)=\begin{cases}\min\left(N_{c_{ini}^l}\left(x_{n_{ini}},t\right),N_{c_{ini}^{l+1}}\left(x_{n_{ini}},t\right),N_{c_{down}{}^j}\left(x_{n_{ini}},(i-1)\Delta t\right)+v\cdot k_c\cdot\Delta t\right) & if\ t\leq\dfrac{x_{n_{ini}}-x_0}{v}\\ \min\left(N_{c_{down}{}^j}\left(x_{n_{ini}},(i-1)\Delta t\right)+v\cdot k_c\cdot\Delta t,N_{up}{}^k\left(x_{n_{ini}},t\right)\right) & else\end{cases}$$

(16.25)

where $j=i-1$, $k=\left\lfloor\frac{t-\frac{x_{n_{ini}}-x_0}{v}}{\Delta t}\right\rfloor$, $l=\left\lfloor\frac{vt}{\Delta x}\right\rfloor$

Proof: The first case corresponds to the situation where only initial components and upstream boundary condition components can influence the upstream condition ($t\leq\frac{x_{n_{ini}}-x_0}{v}$). In this situation, we have that $N_{c_{ini}^k}\left(x_{n_{ini}},t\right)=+\infty$ if $k>l$. Hence, we can write that $N\left(x_{n_{ini}},t\right)=\min\left(N_{c_{ini}^0}\left(x_{n_{ini}},t\right),\dots,N_{c_{ini}^{l-1}}\left(x_{n_{ini}},t\right),N_{c_{ini}^l}\left(x_{n_{ini}},t\right),N_{c_{up}{}^j}\left(x_{n_{ini}},(i-1)\Delta t\right)+v\cdot k_c\cdot\Delta t\right)$. However, by the structure of the initial condition solution components (12), we have that for any $k\in[0,l-2]$, $N_{c_{ini}^k}\left(x_{n_{ini}},t\right)=N_{c_{ini}^k}\left(x_{n_{ini}},(i-1)\Delta t\right)+vk_c(t-(i-1)\Delta t)$. By the inf-morphism property $N_{c_{ini}^k}\left(x_{n_{ini}},(i-1)\Delta t\right)\geq N(x_{n_{ini}},(i-1)\Delta t)$, and thus, since $N\left(x_{n_{ini}},t\right)\leq N\left(x_{n_{ini}},(i-1)\Delta t\right)+k_cv(t-(i-1)\Delta t)$, we have that $N\left(x_{n_{ini}},t\right)\leq N_{c_{ini}^k}\left(x_{n_{ini}},t\right)$ for any $k\in\{0,\dots,l-2\}$, which shows that only $N_{c_{ini}^{l+1}}$ or $N_{c_{ini}^l}$ can influence the solution in $(x_{n_{ini}},t)$. The proof of the second case is the result of Theorem 4.

The above theorems imply that, when computing the upstream and downstream conditions in the initial phase of the computation, the computation of their associated solutions can be restricted to just two consecutive blocks (Figure 16.3a-b). According to Theorem 3 and Theorem 4, the subsequent computations of the solutions at the upstream and downstream boundaries (outside of

the area of influence of the initial conditions) can be reduced to those in the classical LTM formulation. The FLH scheme thus computes the boundary conditions with a slightly higher computational cost as the LTM during the initial phase of the simulation, requiring just two additional operations to account for the initial conditions. For subsequent times both formulations (LTM and FLH) are identical, and thus have the same computational cost.

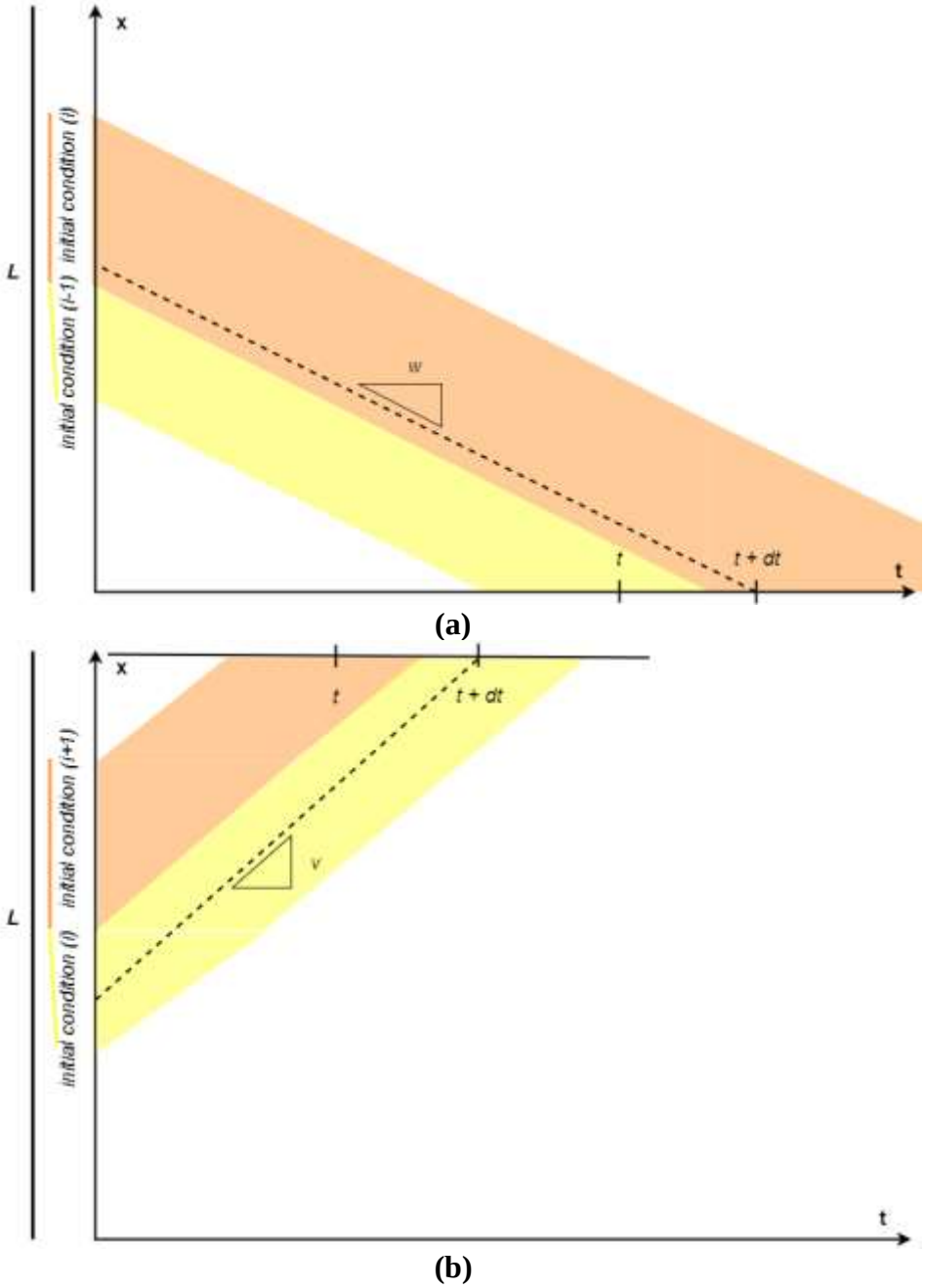


Figure 16.3: Initial conditions considered for computation of flows upstream (a) and downstream (b) according to Theorem 5 and Theorem 6

16.3.4. Theoretical Comparison of FLH Computational Complexity and Accuracy with Other Methods

The computational time required by the different algorithms outlined in the previous section depend on the type of problem that needs to be solved. In this study, we consider two different problems:

1. Computing the solution to the LWR model at the **boundaries** of each link within a road network, with temporal step Δt, over some time horizon T.

2. Computing the solution to the LWR model inside the link, on a **uniform grid** of spatial resolution Δx and temporal resolution Δt, over some time horizon T.

The first problem is typically encountered in forward simulations involving network loading, or network control when the objective function depends only upon the state computed at the boundaries of each computational domain. In contrast, the second problem is common in the applications such as traffic estimation (Cristiani et al., 2011), control (Ferrara et al., 2015), and estimation (Work et al., 2008). In some of these problems the solution only needs to be computed at specific points of the space-time grid, which are known in advance (for example, in estimation problems the solution only needs to be computed on points corresponding to sensors' locations). In this type of situation, the FLH can be used to solve the solution just at these specific locations without relying on the entire grid like in the second problem, unlike CTM and Variational Theory.

The computational performance of all algorithms is illustrated in Table 16.1, where n_{ini} represents the number of initial conditions (or the number of grid points in the x axis), and n_t represents the number of time steps ($n_t = \frac{T}{\Delta t}$).

In the first problem, although the algorithms differ by their accuracy and computational cost, they all converge to the true solution to the LWR model when $\Delta t \to 0$ (and when both $\Delta t \to 0$ and $\Delta x \to 0$ for the CTM and DP). The LTM is the fastest algorithm, requiring 2 calculations per time step. In contrast, the CTM requires at least $4 \cdot n_{ini}$ calculations (computing demand and supplies, and computing the Godunov flux) per time step. The LH algorithm and DP both require on the order of n_{ini} computations per time step (less during the first time steps), and are thus not significantly improving over the CTM. In contrast, the FLH algorithm requires three calculations per time step when $t \leq \frac{x_{n_{ini}} - x_0}{w}$, and two calculations per time step when $t > \frac{x_{n_{ini}} - x_0}{w}$. It thus has a computational complexity comparable to that of the LTM.

The second problem can be solved by all algorithms except the LTM, which is designed by definition for computations only at links' boundaries and it is not convergent whenever the initial condition contains expansion waves (see Section 16.1 for further explanations). Similarly to the first problem, the CTM requires $4 \cdot n_{ini}$ calculations per time step. DP methods require a number of calculations on the order of n_{ini}^2 per time step, while the classical LH algorithm requires on the order of $n_{ini} \cdot (n_{ini} + 2)$ calculations per step, which is similar to the DP. In contrast, the FLH algorithm requires less than $3 \cdot n_{ini}$ calculations per time step, which is a considerable

improvement, and on par with the classical CTM. Note that in practice the FLH can be considerably faster than the CTM when the solution does not have to be computed in all cells (for example, in most estimation, control, or optimization problems).

Table 16.1: Computational performance and accuracy of different algorithms

Numerical scheme	DP	LH	FLH	CTM	LTM
Computational complexity (1)	$\sim n_{ini} \cdot n_t$	$\sim n_{ini} \cdot n_t$	$2 \cdot n_t \leq \leq 3 \cdot n_t$	$\sim n_{ini} \cdot n_t$	$\sim 2 \cdot n_t$
Computational complexity (2)	$\sim n_{ini}^2 \cdot n_t$	$\sim n_{ini}^2 \cdot n_t$	$\sim n_{ini} \cdot n_t$	$\sim n_{ini} \cdot n_t$	X
Accuracy	Convergent	Exact on single links problems	Exact on single links problems	Convergent	Convergent on boundary conditions simulation problems

16.4. Numerical Implementation

In this section, we numerically implement the FLH algorithm presented in Section 16.3, and compare it to the LH, LTM, and CTM, when possible. We show that the FLH algorithm has favorable characteristics in comparison with existing algorithms, particularly when solutions do not need to be calculated everywhere. The interested reader may refer to Mazare et al. (2011) for a comparison between LH, CTM, and DP.

16.4.1. Single Link Case

We now illustrate the problem of computing the solution in a single link problem, where upstream demand and downstream supplies are arbitrary functions. We consider the example of Yperman (2008), and illustrate the cumulative curves obtained with the different algorithms in Figure 16.4. The example corresponds to a two-hour simulation of a single homogeneous link with length L = 10 km, with a triangular fundamental diagram of parameters $q_{max} = v \cdot k_c = 3600\ veh/h$, $v = 120\ km/h$ and jam density of $k_j = 225\ veh/km$. The road is affected by a full blockage downstream that prevents any vehicle to leave (zero supply) and is fed by a constant flow entering the link at capacity rate (constant demand equal to link capacity). Furthermore, the link is initially assumed to be empty. The LH, LTM, and the FLH use a time step of 1 minute while the CTM is characterized by a time step of 30 seconds. As it possible to see from the cumulative numbers of vehicles at the upstream end of the link, for the chosen time interval, the LH, FLH, and LTM are more accurate than the CTM. The CTM, even for a relatively fine time discretization, yields a significant numerical error (shockwaves do not travel at the correct speed). This phenomenon is due to the assumption of homogeneous density in each cell (Daganzo, 1994; Yperman, 2007).

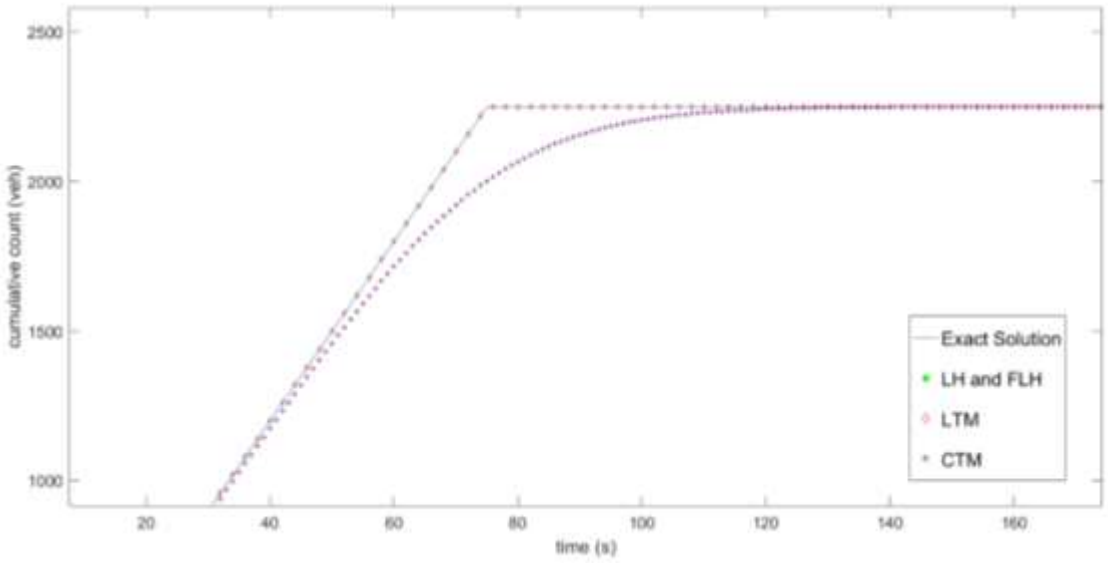
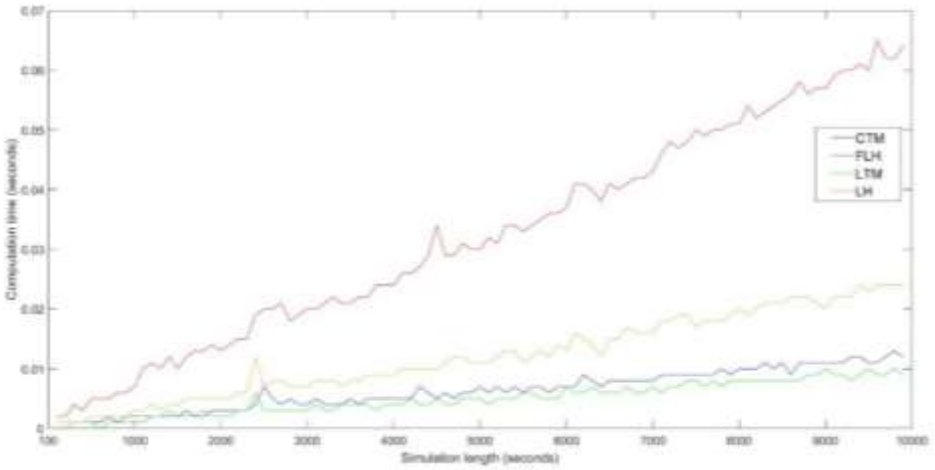


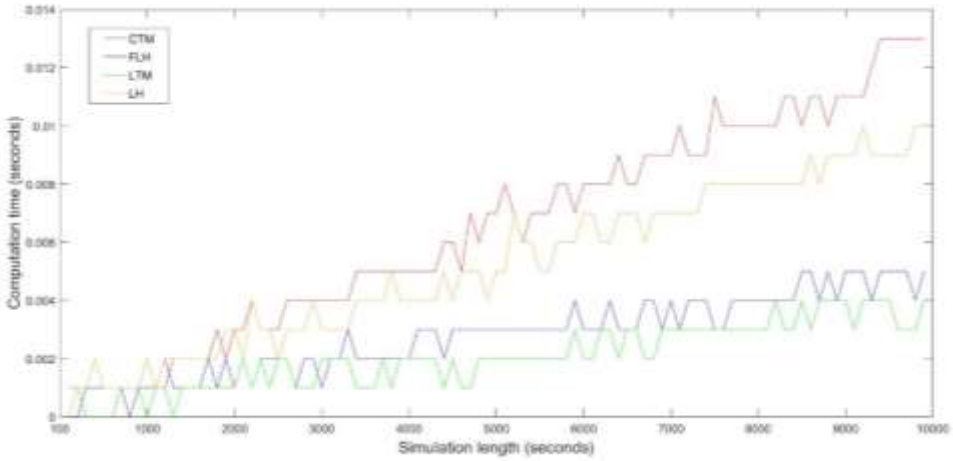
Figure 16.4: Cumulative curves at the upstream end of the link obtained with different algorithms

With respect to computational times, we compare in Figure 16.5a-b the LH, FLH, LTM, and CTM algorithms for computing the solution at the boundaries of the link, with increasing time horizons, using time steps of 2 and 5 seconds. The simulation involves a link of 500 meters and two lanes. The parameters (per lane) of the fundamental diagram are the same as in the previous example. The upstream demands and downstream supplies are randomly chosen in the interval $[0, q_{max}]$. The initial condition consists in a single block, with random initial value. As expected, the results of this experiment show that the LTM has the lowest computation times, followed by the FLH, the LH and the CTM. However, the FLH's computational performance is comparable to one of the LTM.

As proved mathematically in the previous section, the FLH is faster than the original LH, thanks to the elimination of solution blocks through the computation process. This is illustrated in Figure 16.6, where the computational cost is calculated for both models during a simulation of 200 seconds. As the figure demonstrates, the FLH gains advantage over the LH over the simulation. The number of operations per time step required to compute the solution using the FLH algorithm decreases over time, as indicated by the changes of slope (around 15 and 60 seconds), as an increasing number of initial condition blocks can be discarded.



(a)



(b)

Figure 16.5: comparison of computation times for different models using a time step of 2 seconds (a) and of 5 seconds (b)

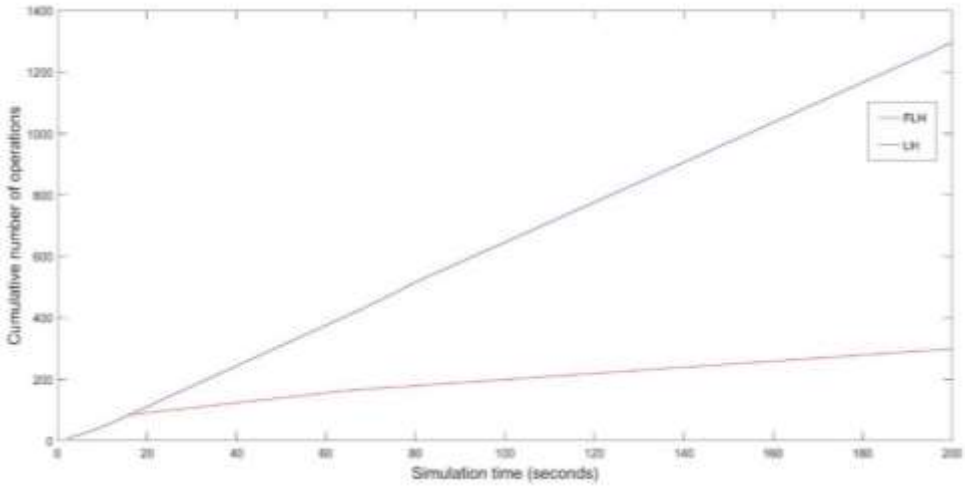


Figure 16.6 Computational cost calculated for both models during a simulation of 200 seconds

16.4.2. Network Case Studies

Also in case of network simulations, the higher accuracy of methods like the FLH and LTM compared to the CTM becomes apparent. As an example we show the simulation results of a five-link highway network (Figure 16.7) composed of a three-lane major highway section (Links 1,2 and 3), a two-lane off-ramp (Link 4) and a two-lane on-ramp (Link 5). A triangular fundamental diagram with capacity q_{max}=0.556 veh/s/lane, free-flow speed v=30 m/s, and jam density of k_j=0.1297 veh/m/lane is adopted for all three models. All links are characterized by initial free-flow density k_1=0.004 veh/m, with the exception of Link 2; that is characterized by two initial condition blocks associated with densities k_1=0.004 veh/m in the downstream half and $k_2 = 0.01$ veh/m in the upstream half of the link.

In order to model traffic throughout intersections, there is need for a generic macroscopic node model that respects some critical conditions: satisfaction of links' capacity constraints; conservation of flows; satisfaction of demand distribution constraints; maximization of flows (vehicles should proceed if there is available supply downstream); satisfaction of invariance principle (if the flows are restricted by demands, solutions cannot vary by increasing supplies and vice versa); and non-simultaneity of conflicting flows. In this study, we adopt the "I-HFS algorithm" by Jabari (2016), which respects the abovementioned properties and efficiently derives solutions by staging movements according to any arbitrary priory rules.

As Figure 16.8 indicates, the results of the LTM and LH algorithms are close to the exact solution to the problem, while the solution computed by the CTM exhibits significant errors. We then compare the performance of all three schemes (CTM, LTM, and FLH) in computing the solutions at the boundaries of each link of the network, averaged over random initial conditions, and random boundary demand and supplies at the edge of the road network. The results are averaged over 100 simulations, where the initial condition densities, demand, and supply flows are drawn independently from uniform distributions. The average root mean square error of all three schemes

is shown in Figure 16.9. As this figure demonstrates, FLH outperforms both the CTM and LTM in terms of error.

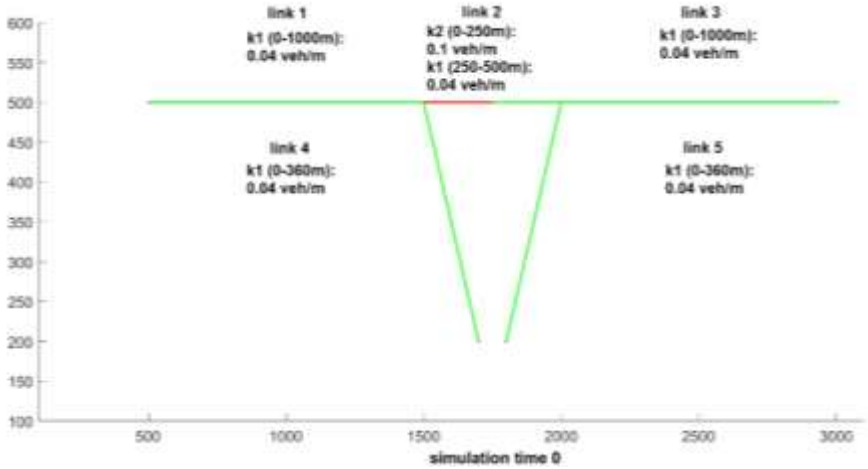


Figure 16.7: Simulation of the highway network at t=0 seconds.

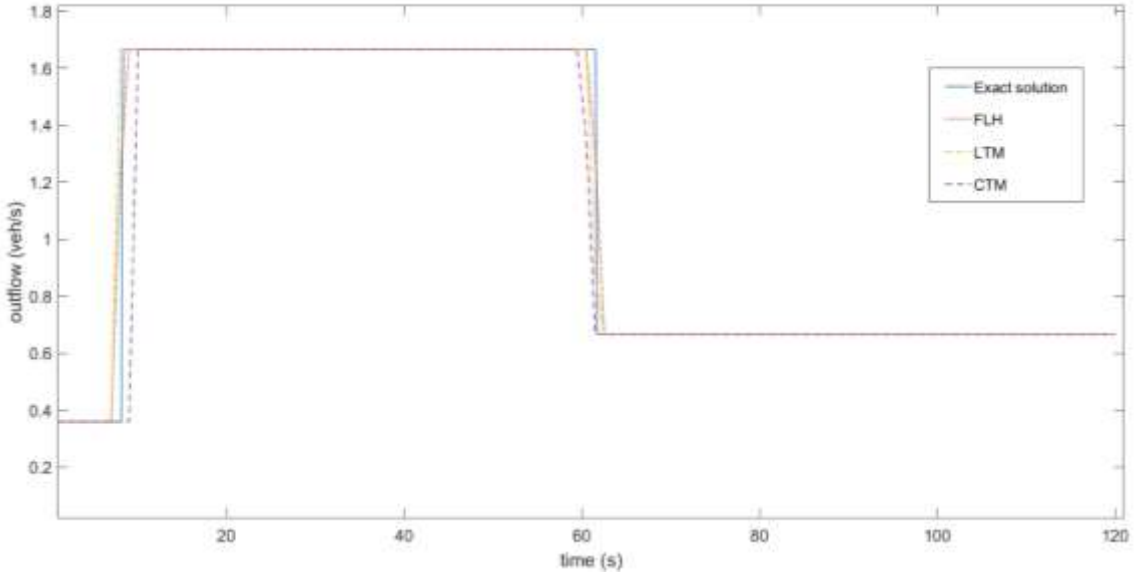


Figure 16.8: Comparison of the outflows of link 2 with the three methods (using a time step of 1 second)

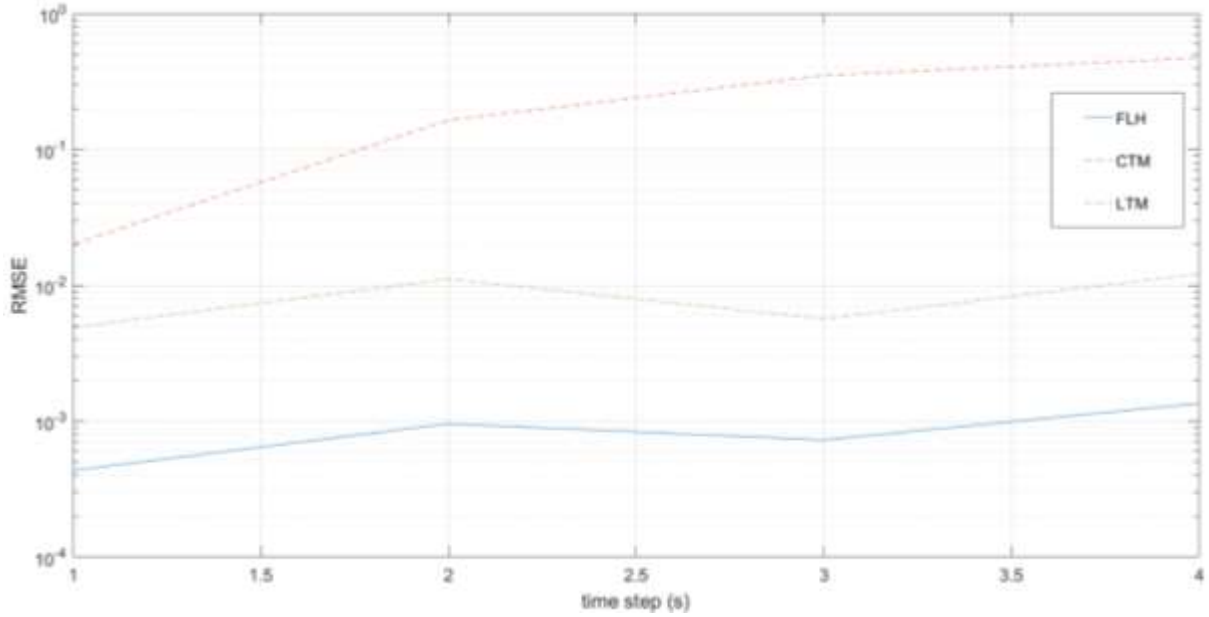


Figure 16.9: Accuracy of outflows calculated with the three different methods according to increasing time step

The favourable computational time properties of the FLH appear also in large network simulations. To demonstrate the scalability of the algorithm, we present and discuss the results of its application to a subset of the Austin downtown network (Figure 16.10). The network is characterized by 201 links and 110 nodes. Streets have between two to three lanes and the majority of the intersections is signalized (about 90%). For simplicity, in this study we only model green/red phases and we adopt the same triangular fundamental diagram for all links with: q_{max}=0.4625 veh/s, v=12.5 m/s, and k_j=0.1295 veh/m.

We report in Table 16.2 and Table 16.3 the (average) computation times obtained for increasing simulation horizons, using different time steps for the four models. The simulations were performed on Matlab running on a laptop with a 2.8 GHz processor. The results are consistent with those obtained for single-link simulations. FLH and LTM have comparable performances when the initial conditions are explicitly considered in the computation. The CTM, for larger time steps (e.g., 5 seconds), is equivalent to the other methods since the links of this network are relatively short (resulting in a low numbers of cells). The numerical approximation of CTM, however, amplifies on large networks, leading to significant divergence from the exact solution after relatively short simulation horizons. We illustrate in Figure 16.11 the average densities across the links of the network, at a given time horizon, using either the CTM or the FLH algorithm.

Original source: Google Earth

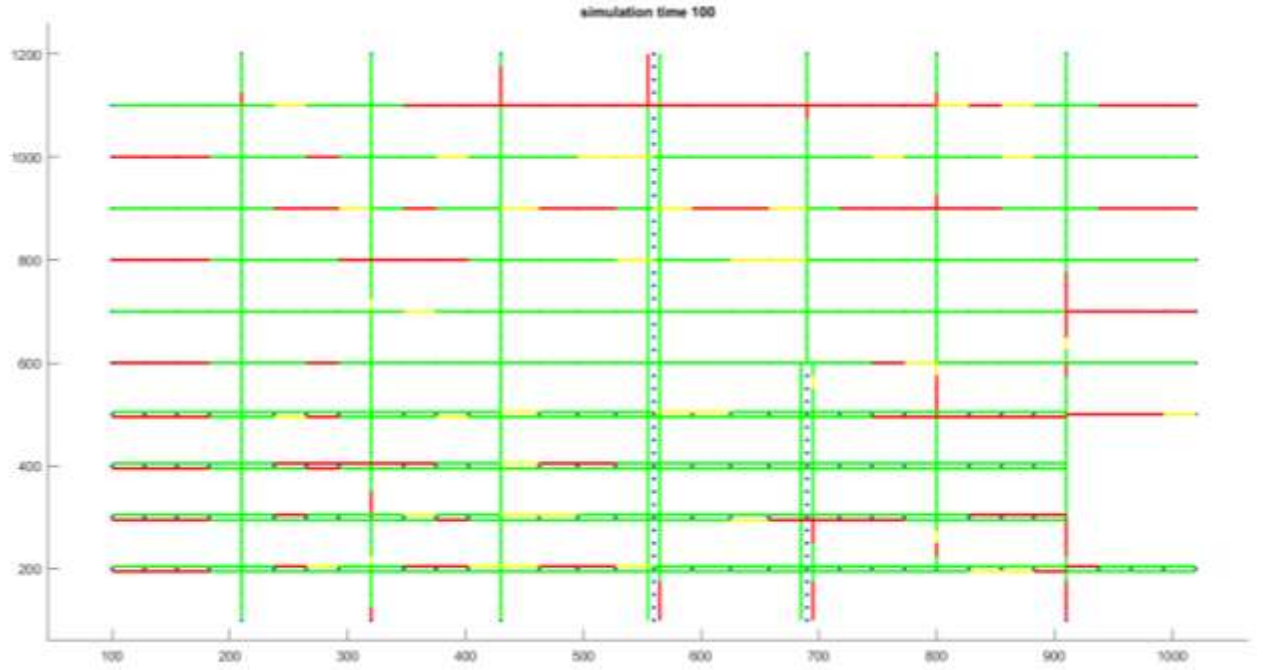
Figure 16.10 Austin downtown network

Table 16.2: Comparison of computational times (in seconds) for different simulation lengths in Austin downtown network using a time step of 1 second

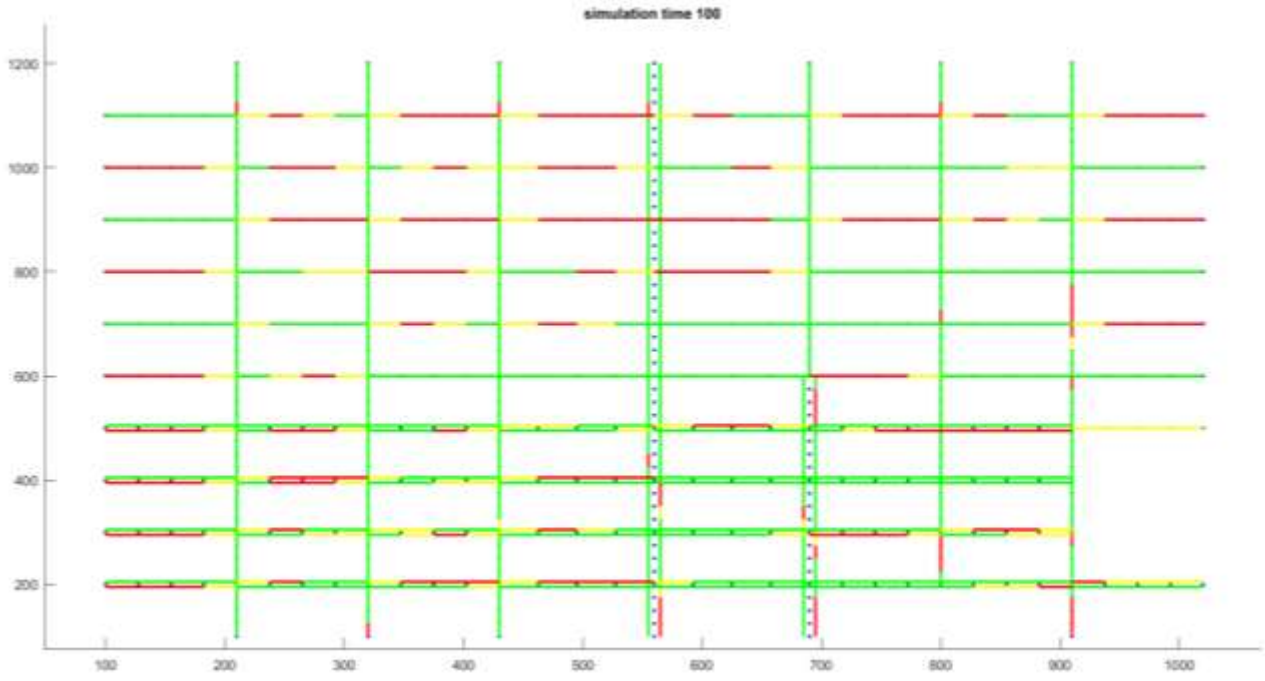
	CTM		LTM		FLH	
Simulation Horizon (s)	**link model**	**node model**	**link model**	**node model**	**link model**	**node model**
200	1.411	0.485	0.359	0.520	0.354	0.527
500	3.582	1.159	0.938	1.368	0.856	1.309
1000	8.334	2.562	1.888	2.570	1.799	2.596

Table 16.3: Comparison of computational times (in seconds) for different simulation lengths in Austin downtown network using a time step of 5 seconds

	CTM		LTM		FLH	
Simulation Horizon (s)	**link model**	**node model**	**link model**	**node model**	**link model**	**node model**
200	0.066	0.110	0.074	0.106	0.079	0.113
500	0.164	0.260	0.177	0.266	0.169	0.257
1000	0.308	0.467	0.333	0.483	0.341	0.526



(a)



(b)

Figure 16.11 Average densities across the links of the network

16.4.3. Discussion

An important difference among the discussed models is that, while at any point (x, t), the solutions generated by the CTM and the FLH converge toward the solution of the LWR PDE, the solution generated by the LTM converges only in specific cases.

We illustrate this in Figure 16.12, where we present a scenario in which we consider a triangular fundamental diagram of parameters $k_c = 0.037\ veh/m$, $k_j = 0.1297\ veh/m$, $u = 20\ m/s$, and $w = 3.5\ m/s$. We assume that the upstream half of a link is congested ($k_1 = 0.1297\ veh/m$), while the downstream half of the link is free flow ($k_2 = 0.01\ veh/m$). In this specific case, calculating *N* at the point A (10,600) by using the LTM procedure (Newell's method) would yield: $N_A = \min\{N_D + (x_A - x_D) \cdot k_j; N_U\} = -51.88$. The correct solution, instead, would correspond to: $N_A = N_{U'} = -63.65$.

A common procedure to avoid expansion waves would be partitioning the network by splitting links wherever an expansion wave would occur, that is, wherever the density would decrease over space. However, dividing the link presents two issues.

First, this requires a modification of the topology of the network, which becomes a function of the choice of the initial conditions. This is problematic in case of unknown initial conditions (for example, in estimation, optimization, or robust control problems[29]). This would increase the computational overhead before the actual computation process.

Second, splitting the link in two or more links would increase the computational time to find the solution by a factor of two or more, since the demand and supplies at the boundaries would have to be derived for each split link.

The proposed FLH algorithm avoids these computational issues by imposing a minor computational penalty on the original LTM (three computations per time step instead of two computations per time step, while in the domain of influence of the initial conditions). Splitting a link in two for example would require four computations per time step. More than one split may be required, depending on the number of initial condition blocks and their configuration.

It would still be possible to apply Newell's method to derive the solution at the link boundaries as we did in previous examples. Because of that, the LTM would have a slightly increased computational time, comparable to the FLH.

[29] In these situations, often a large number of random initial conditions is drawn according to a certain distribution and the simulation outcomes are used for the estimation or control process. Examples include Ensemble Kalman Filtering for the prediction step, or Model Predictive Control, or Particle Filtering.

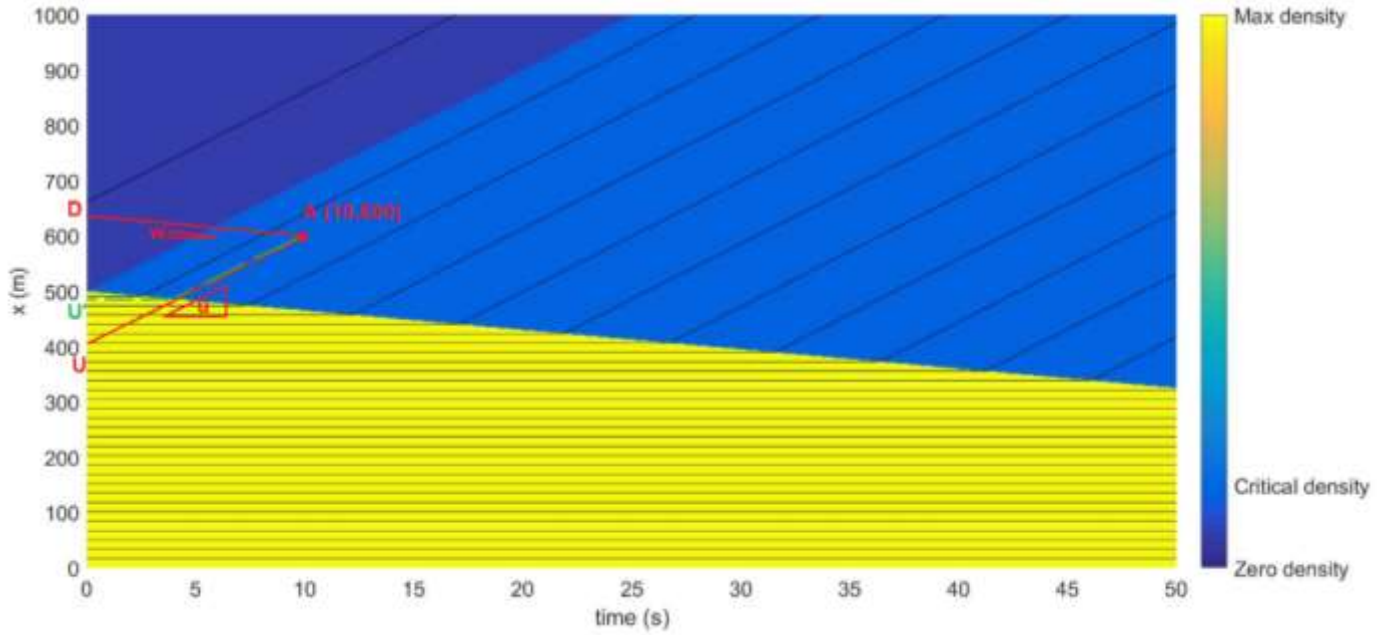


Figure 16.12 Derivation of the solution at point A by using Newell's method (solid lines) and correct approach (dashed line)

Chapter 17. Development of an IMU-based Traffic and Road Condition Monitoring System

17.1. Introduction

17.1.1. Traffic and Road Surface Monitoring

As traffic congestion becomes an increasing burden all over the world, creation of real-time traffic monitoring system becomes an essential step in mitigating the effects of traffic congestion. Traffic monitoring is a critical component in traffic estimation and forecasting systems (Wang and Work, 2014) that generate traffic maps (Canepa and Claudel, 2012), travel time estimates (Mazare et al., 2014), optimal routes for vehicles, or optimal control policies (Li et al., 2014), for traffic control systems and help prepare roadway geometric, intersection, and pavement design.

One important application of traffic monitoring is to estimate traffic conditions (congestion detection or trajectory estimation). Such applications need to estimate the present traffic situation and that of the near future at a forecasting horizon based on data that are available in real time (Canepa and Claudel, 2017). However, such sensing systems can be expensive, require high amounts of power, or have to be in range of satellites, which may degrade the accuracy in urban areas due to the reduction of the number of visible satellites caused by the urban canyon effect (Ojeda and Borenstein, 2007; Weiss et al., 2011; Bachrach et al., 2011). Road surface monitoring is another important application of traffic estimation systems, which aims at detecting any surface imperfections (such as road surface cracks) at early stages in order to apply road maintenance on time (Fendia et al., 2014). For this project, we developed a low-cost and high-accuracy traffic monitoring system based on use of inertial measurement units (IMUs). This system can realize both estimate traffic conditions and monitor road surfaces. Since IMUs do not generate the absolute position measurement data needed for traffic estimation or direct information to indicate road surface condition, some essential computational tasks should be operated to realize such functions, which requires the development of a new platform and new software capable of meeting these specifications.

17.1.2. Lagrangian Sensing

Traffic monitoring systems are often categorized as either fixed (Eulerian) or mobile/probe-based (Lagrangian) systems. The former often includes a variety of sensors, such as radars, inductive loop detectors, and traffic cameras (Alessandretti et al., 2007; Ki and Baik, 2006; Braberger et al., 2004), which are always installed at a fixed point. In contrast, a Lagrangian-based system, such as GPS, relies on the data generated by vehicle themselves, which measure traffic conditions along the path of a vehicle. Over the past decades, Lagrangian sensing has become increasing prevalent in modern traffic monitoring systems.

As one of the Lagrangian sensing methods, vehicle probe technology is emerging as a viable means for traffic monitoring, delivering speed, location, and time information for the purposes of advanced traffic management systems (Zheng and Zuylen, 2013; Young, 2007). While probe data is relatively accurate and has an extremely low marginal cost, many issues remain associated with this technology—in particular the low penetration rate of participating users due to weak user privacy guarantee (for instance, a GPS-based system needs the users to share their location data). Other issues such as high power consumption (as with GPS in cellphones) or higher cost also prevent the large-scale deployment of such systems. Based on these concerns, our objective is to design a low-cost, low-power Lagrangian system, which at the same time guarantees the privacy of users.

17.1.3. Wireless Sensor Network

Wireless sensor networks have emerged as a solution for urban monitoring applications due to their computation, communication, and sensing capabilities. As for most of the urban sensing systems, the cost for deployment and maintenance is usually higher than the hardware cost, compared with which WSNs feature easier deployment and better flexibility of devices. Due to all these advantages, WSNs have been used for countless applications in many different fields such as environmental monitoring, structural health monitoring and seismic activity detection (Ye et al., 2009; Kijewski-Correra et al., 2006; Lopes Pereira et al., 2014). In our case, the proposed system should also be fully wireless in order to meet the flexibility requirement of probe vehicles as well as to minimize the deployment cost. In other words, a wireless sensor network need to be built so that the data generated by the sensing device can be sent to a wireless sensor network for processing, then estimated information can be matched to real road network through the sensor network. At last, the sensor network will send the corresponding data to a traffic state estimation server.

17.2. Computational Requirements

Even though the low-cost IMU can generate high frequency and accurate sensing data, it does not generate speed or positioning data directly, which is necessary for most of the traffic related applications. Also, the data generated by IMU cannot indicate road surface condition directly without any further analysis. Thus, an efficient computational platform is needed for this proposed sensing system to generate the essential information through conduction of some mathematical algorithms. For the proposed system, each node has to carry out a number of computational tasks for traffic sensing purpose:

1) Automatic calibration of IMU;
2) Traffic estimation and sensor bias detection;
3) Road surface condition monitoring based on IMU data, which is the focus of the present chapter.

17.3. Platform Architecture and Design

In order to support a low-cost, distributed, and reliable traffic monitoring system to solve the aforementioned computational tasks in real time, we designed a hardware platform, on which we are porting an operating system to simplify programming. Following these requirements, the proposed platform should have some specific features in comparison with other reported hardware platforms:

1) Low node and deployment cost, low operation power needed;
2) Small size and straightforward installation for probe vehicles, to allow real-time sensing;
3) Advanced computational capabilities, with enough free memory to allocate relatively large matrices required for self-calibration, attitude estimation, and other computational tasks needed for the application.

Figure 17.1 shows the block diagram of the hardware platform. Figure 17.2 illustrates the third generation hardware platform over the hardware development period (September 2015 to December 2017). We now give a detailed description of this hardware platform by focusing on the following areas: processing unit (the core element), communications, data storage, sensors, and peripherals.

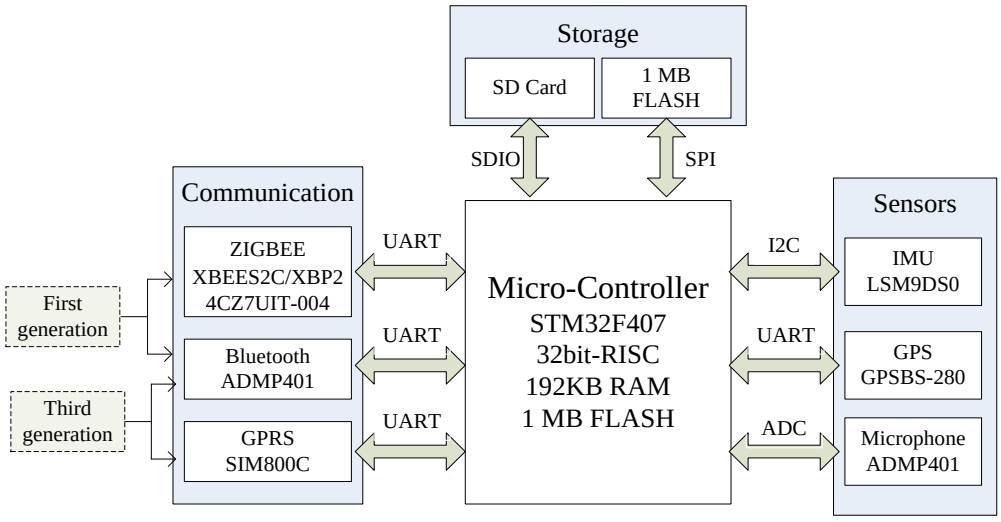


Figure 17.1 Block diagram of the IMU platform (different versions)

Figure 17.2 Third-generation platform

17.3.1. Core Element

The core component of the sensing platform is the microcontroller unit (MCU), which handles sensing (analog-to-digital converter [ADC] and digital buses), computation, and control. We selected for this application the STM32F407, a 32-bit ARM Cortex-M4-based MCU from ST since it satisfies the requirements described above and best balances the tradeoff among computation, RAM, power consumption, and cost. We considered a wide range of MCUs, in which the ATmega1281 is on the low end (low performance, low power consumption) and the TI TMS570 on the high end (high performance, high power consumption). The MCUs in the low end are not able to provide with sufficient internal data RAM (8 KB), program memory (128 KB), and computational power (16 MHz). On the other hand, while the high end exhibits a fast frequency (180MHz), they also have higher power consumption and higher prices, which are unsuitable for extensive deployment. In contrast, the STM32F407 provides a comparable performance with the TMS570 at only one-third of the price; it is even less expensive than an 8-bit ATmega1281. In addition, at its lowest frequency setting, its power consumption is comparable to the power consumption of the ATmega1281.

The STM32F407 includes a 1-MB Flash memory and 196 KB of data RAM. It supports up to 17 timers, a 24-channel ADC, and two 12-bit DACs for peripherals.

On this platform, the MCU is configured to have three universal asynchronous receivers/transmitters for communication and positioning modules, and one I2C interface bus and one ADC interface for sensors. Furthermore, a SDIO and a USB OTG bus are configured to provide MicroSD Flash storage and USB host access. The STM32F407 supports a maximum frequency of 168 MHz, which is sufficient to run the envisioned traffic sensing and estimation algorithms in real time.

17.3.2. Communications

The transmission of data between different sensor nodes requires the use of radio transceiver. For the first generation, we choose XBEE S2C/XBP24CZ7UIT-004 from Digi working at 2.4 GHz, using the IEEE 802.15.4 standard. This transceiver is capable of generating signals up to +18 dBmm, which is the maximal legally allowed transmission power in the 2.4-GHz band (equivalent

to 100-mW EIRP when combined with a 2-dB dipole antenna). While there are a number of 802.15.4-compliant radio transceivers available, such as the TI CC2500, their maximal radiated power is insufficient for our application.

In addition, a Bluetooth module is used for dynamic data transfer from the sensor to the mobile device. The selected module is SH-HC-06, which has the Blue Core4-Ext chip, and follows the Bluetooth V2.0 + EDR Bluetooth standard. The maximum transfer rate of this transceiver is 2.1Mb/s, and the transmission distance is more than 20 meters. Compared with other possible choices, this module is low cost, small size, and with a high sensitivity—up to -80dBm at 0.1% BER (bit error rate). It supports AT commands to modify serial baud rate, device name, pairing passwords, and other parameters. As for the software, a mobile client was developed with an Android operating system for data transmission and storage.

Figure 17.3 Transmission of data to an Android phone via Bluetooth

For the third-generation platform, the GPRS SIM800C module is used for remote monitoring purpose and communicating the sensing data to remote server via the cellular phone network. This transceiver is selected mainly for its low power consumption and small size. The SIM800C is designed with power saving technique, and the current consumption can be as low as 0.6 mA while in sleep mode. It is with a tiny configuration of only 17.6*15.7*2.3 mm, which can meet the space saving requirement for our platform. This module supports 4G GSM/GPRS working on frequencies GSM850MHz, EGSM900MHz, DCS1800MHz, and PCS1900MHz. And the SIM card interface support 1.8V and 3.0V SIM cards. A micro SIM card is used in this transceiver for data storage and communication.

17.3.3. Data Storage

The MCU has an internal 1-MB Flash memory for storing the bootloader, firmware, the operating system, and a 196-KB internal SRAM for data during firmware execution. While this amount of memory is sufficient for real-time processing to perform their computations, we need additional storage for non-volatile data storage (for instance, measurement data needed for trajectory estimation, historical acceleration from IMU, and audio spectrum information from the Audio

Processing module needed for pavement condition monitoring). We thus added a microSD slot (MicroSD Flash Socket), which is accessed through a SDIO interface, and can support 8-GB microSD cards (FAT 16).

17.3.4. Sensors

The real-time traffic estimation and road surface condition monitoring is based on the use of IMUs. IMUs are based on a combination of accelerometers and gyrometers, which can be used to determine the accelerations and rotation rates of a vehicle. IMUs do not require any external infrastructure to work and do not receive or transmit data wirelessly. They just require an extremely low power to operate, considerably less than GPS or cellphone-based systems. Because they are much less complex than GPS systems, IMUs are less expensive to manufacture (Mousa et al., 2015; Jimenez et al., 2009; Wan and Foxlin, 2001). They do not require an antenna for receiving signals, and are not at risk of losing connectivity with positioning satellites, which frequently happens with GPS systems, particularly if obstructions are present between the receiver and the satellites. They are also immune to environmental noise effects, in particular to the multi-path effect encountered in cities. Because of their very high accuracy (over short time windows), IMUs are very good at detecting and classifying the type of congestion encountered (such as traffic lights, stop-and-go waves, slow and continuous traffic) (Wan and Foxlin, 2001; Heng et al., 2015). In addition, such a system offers strong guarantee s for the privacy of the participating users when used in conjunction with a short-range wireless sensor network (Fuke and Krotkov, 1996; Canepa et al., 2014). Those features of IMUs make it more reliable than a GPS-based positioning system.

In our platform, IMU GY-85 is used for the first generation, which consists of ITG3205, ADXL345, and HMC5883 chips. For the third generation we use the LSM9DS0 instead, which is of higher accuracy compared with GY-85. The LSM9DS0 is a system-in-package featuring a 3D digital linear acceleration sensor, a 3D digital angular rate sensor, and a 3D digital magnetic sensor. It has a linear acceleration full scale of ±2g/±4g/±6g/±8g/±16g, a magnetic field full scale of ±2/±4/±8/±12 Gauss and an angular rate of ±245/±500/±2000 degrees/s. The embedded self-test allows the linear acceleration sensor functionality to be tested without moving. This module is connected to the MCU with I2C serial interface and provides a 16-bit data output.

The IMU is the main sensor used in our platform to obtain position data for traffic monitoring—for instance, in the context of trajectory estimation. In order to verify the accuracy of the estimated results by IMU, the exact position information of the vehicle along its path is needed as validation. Thus, a GPS receiver is used in our platform for getting vehicles' localization, which is used only for validation. The Beitian BS-280 GPS was selected for our platform for its relatively high output frequency with a low price and low power requirement. The module is integrated with the UBLOX G7020-KT chip, and can reach an output frequency of up to 10 Hz.

For pavement condition monitoring purposes, one important computational task is to estimate and quantify the audio noise that could be a predictor of the road condition. Thus, an omnidirectional microphone was embedded in our platform for noise detection. The main computational task to be conducted for this module is to sample the sensing signal and divide it into different frequency

components through the Fast Fourier Transform (FFT) algorithm. The audio processing module selected for our platform is the ADMP401, which is designed as a 4.72 mm*3.76 mm*1.0 mm surface-mount package that can meet our space constraints as well. Besides, the signal-to-noise ratio of the ADMP401 can be as high as 60 dBA and the sensitivity is of -42 Dba with a relatively low price. Its current consumption is extremely low.

17.3.5. Other Embedded Auxiliary Equipment/Peripherals

The peripherals consist of several functioning blocks: a self-resetting circuitry, LEDs, and a USB monitoring unit. As the proposed platform works with high modularity, different modules works for different functions (sensing, communicating, or storage), it is also important to make sure each part is functioning well to support the overall functioning of the whole system. Thus, several LED lights were incorporated into our platform to indicate the normal functioning of each module in case of running error. In the first generation, three LED lights were embedded in our platform to indicate the functioning of Bluetooth, GPS, and IMU modules.

Software bugs are a risk for every firmware, particularly in embedded systems. To anticipate the presence of software bugs, a self-resetting circuitry was included in our platform to prevent node failures. The functionality of the circuitry is to reset the whole system while any part of it is not functioning normally, and a reset button was designed on the platform for convenient resetting the system if any module does not function properly. For instance, if any of the LED lights associated with each peripheral does not blink as expected, we can use the reset button to restart the entire system.

Since this platform is mainly designed for probe vehicles, it should be installed in a moving vehicle easily and get continuous power supply for real-time sensing. A USB port was added for powering the sensor, via a USB car charger. The USB port also plays the role of a structural support that maintains the device in a constant (albeit unknown) orientation with respect to the vehicle frame, despite the accelerations, turns, and presence of road bumps.

17.4. Software

This platform runs Keil RTX, which is a deterministic and high-speed real-time multitasking operating system for ARM and Cortex-M processor-based devices. It allows to create programs that simultaneously perform multiple functions and to create applications which are well structured and easily maintained with low interrupt travel time. Also, this system can manage unlimited number of tasks each with up to 254 priority levels as well as unlimited number of mailboxes, semaphores, mutex, and timers. The source code is mainly written by ANCI C; thus, developers can write the application code in ANSI, with no need to learn a specialized language for programing (such as NesC for Tiny OS). Compared with other event-driven embedded OSs such as Tiny OS or Contiki, an important advantage of Keil RTX is its flexible scheduling, which is round-robin, pre-emptive, and more collaborative. For our application, both traffic estimation and pavement condition monitoring require certain levels of real-time operation. Thus, the flexible task scheduling mechanism of Keil RTX is an appropriate selection for this platform.

17.5. Platform Cost Evaluation

The costs and functions of some major components used for the proposed platform are listed in Table 17.1. The entire cost of the subsystems of our third version device is around $55, including all the sensing, communication, and storage modules. The proposed platform can be used for probe vehicles. It should be put into use with mass production in the future, which will make the cost reduce to less than $40 once the breakdown quantity is reached.

Table 17.1 Cost of the major components in the different platforms (excludes manufacturing costs)

Version	Item	Quantity	Price $	Breakdown price $	Remarks
1st	STM32F407	1	11.05	7.18@1000	Micro-controller
	SH-HC-06	1	8.99	5.46@500	Bluetooth Transceiver
	Beitian BS-280	1	12.19	10.79@100	GPS
	XBP24CZ7UIT-004	1	10.403	8.53@500	XBEE Transceiver
	GY-85	1	8.45	5.82@300	IMU Sensor
	Total price $51.083				
2nd	STM32F407	1	11.05	7.18@1000	Micro-controller
	SH-HC-06	1	8.99	5.46@500	Bluetooth Transceiver
	XBP24CZ7UIT-004	1	10.403	8.53@500	XBEE Transceiver
	GY-85	1	8.45	6.20@300	IMU Sensor
	Total price $40.23				
3rd	STM32F407	1	11.05	7.18@1000	Micro-controller
	SH-HC-06	1	8.99	5.46@500	Bluetooth Transceiver
	SIM800C	1	8.75	5@500	GPRS Transceiver
	LSM9DS0	1	7.11	3.83@3000	IMU Sensor
	Beitian BS-280	1	12.19	$10.79@100	GPS Sensor
	ADMP401	1	6.99	4.12@500	Microphone
	Total price $55.08				

17.6. Applications of the Platform

17.6.1. Automatic Calibration of the IMU

The creation of this low-price and high-accuracy platform for traffic sensing and road surface monitoring is based on the use of IMU. However, IMUs do not generate absolute position measurement data compared with GPS, and the initial acceleration and rotation rate generated by the IMU is in its own frame, not in the frame of a vehicle. Thus, we needed to measure the acceleration along the longitudinal, lateral, and vertical axes of the vehicle to determine the orientation of the device with respect to the vehicle automatically, which is referred as *automatic calibration of the IMUs* (Mustafa et al., 2002). The automatic calibration is a basic computational task for our platform to conduct applications such as stop detection, congestion classification, trajectory estimation, and road condition monitoring.

In order to determine the orientation of the device in the vehicle (assumed to be constant, since the device is rigidly connected to an USB port), we need to compute a corresponding rotation matrix mapping the coordinates of the device to the coordinates of the vehicle. Fortunately, the dynamics of ground vehicles are constrained, which allows us to develop an algorithm that automatically computes the rotation matrix transforming the sensor coordinates to the vehicle coordinates (Gustafsson et al., 2017).

Let i_c, j_c, and k_c be the unit vectors associated with the longitudinal, lateral, and vertical axes of the vehicle, and let $R_{d/c}$ (v_1, v_2, v_3 are three column vector of the matrix) be the rotation matrix mapping the coordinates of the vehicle into the coordinates of the device. Then, we have:

$$\begin{bmatrix} a_x(t) \\ a_y(t) \\ a_z(t) \end{bmatrix} = R_{d/c} \begin{bmatrix} a_x'(t) \\ a_y'(t) \\ a_z'(t) \end{bmatrix} \tag{17.1}$$

Firstly, based on an average assumption that the attitude of the vehicle on Earth is flat (that is the vehicle has on average a zero pitch and roll angle), we can get the third column of the matrix by averaging the acceleration measurements of the IMU.

$$\frac{1}{T}\int_0^T \begin{bmatrix} a_x'(t) \\ a_y'(t) \\ a_z'(t) \end{bmatrix} dt \approx \begin{bmatrix} 0 \\ 0 \\ g \end{bmatrix} \tag{17.2}$$

$$\overrightarrow{v_3} = \frac{1}{gT}\int_0^T \begin{bmatrix} a_x(t) \\ a_y(t) \\ a_z(t) \end{bmatrix} dt \tag{17.3}$$

Then the first column of the matrix can be achieved though determining the orientation of the longitudinal axes of the vehicle in the sensor coordinates, which can be obtained by performing a linear fit on the values of residual acceleration vector. The residual acceleration vector is the projection of the acceleration vector on a plane perpendicular to v3.

Having the first and the third column of the rotation matrix, the second column is obtained by cross product: $v_2 = v_3 \times v_1$, allowing us to determine the rotation matrix univocally. The complete process is outlined in Figure 17.4. Based on this algorithm, we imply the code within our proposed platform using Keil v5.0 from the ARM group.

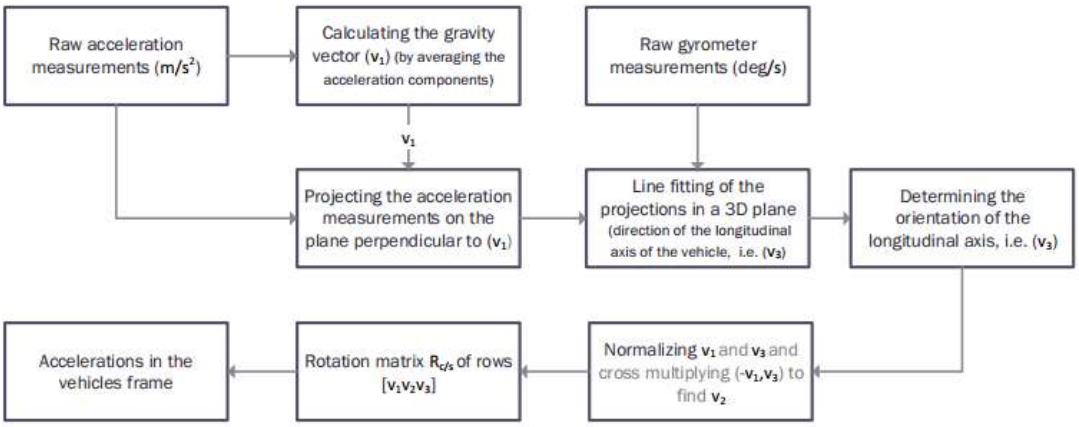


Figure 17.4 The automatic calibration algorithm

17.6.2. Road Surface Monitoring and PSR Estimation

Road surface monitoring is the process of detecting the surface imperfections on paved or unpaved road surfaces, which plays a key role in ensuring safety and comfort to the various road users, from pedestrians to drivers. Furthermore, having information on infrastructure quality allows road managers to guarantee an adequate maintenance (Astarita et al., 2012; Wael et al., 2016). Pavement roughness is an important pavement characteristic used to indicate the condition of road surface, which is generally defined as an expression of irregularities in the pavement surface that adversely affect the ride quality of a vehicle (and thus the user). Pavement roughness affects not only ride quality but also vehicle delay costs, fuel consumption, and maintenance costs.

Pavement roughness is typically quantified using some form of either present serviceability rating (PSR), international roughness index (IRI), or other index, with IRI being most prevalent. One of the earliest pavement condition indices developed at the AASHO Road Test, PSR is defined as "The judgment of an observer as to the current ability of a pavement to serve the traffic it is meant to serve." To generate the original AASHO Road Test PSR scores, observers rode around the test tracks and rated their ride using the quantitative scale ranges from 5 (excellent) to 0 (essentially impassable). Since the PSR is based on passenger interpretations of ride quality, it generally reflects road roughness because roughness largely determines ride quality.

Road surface condition monitoring is an important application of our proposed platform. In this chapter, PSR is used as the indicator of road surface condition. As shown in Figure 17.5, a series of routes in Austin area with certain PSR values are chosen as the test routes for vehicles equipped with our platform. Those routes are color-coded for different levels of PSR (Red: 1<PSR<2; Blue: 2<PSR<3; Purple: 3<PSR<4; Green: 4=PSR). The tests were conducted with the same type of vehicle (the Ford F150) used to obtain the PSR value of these routes.

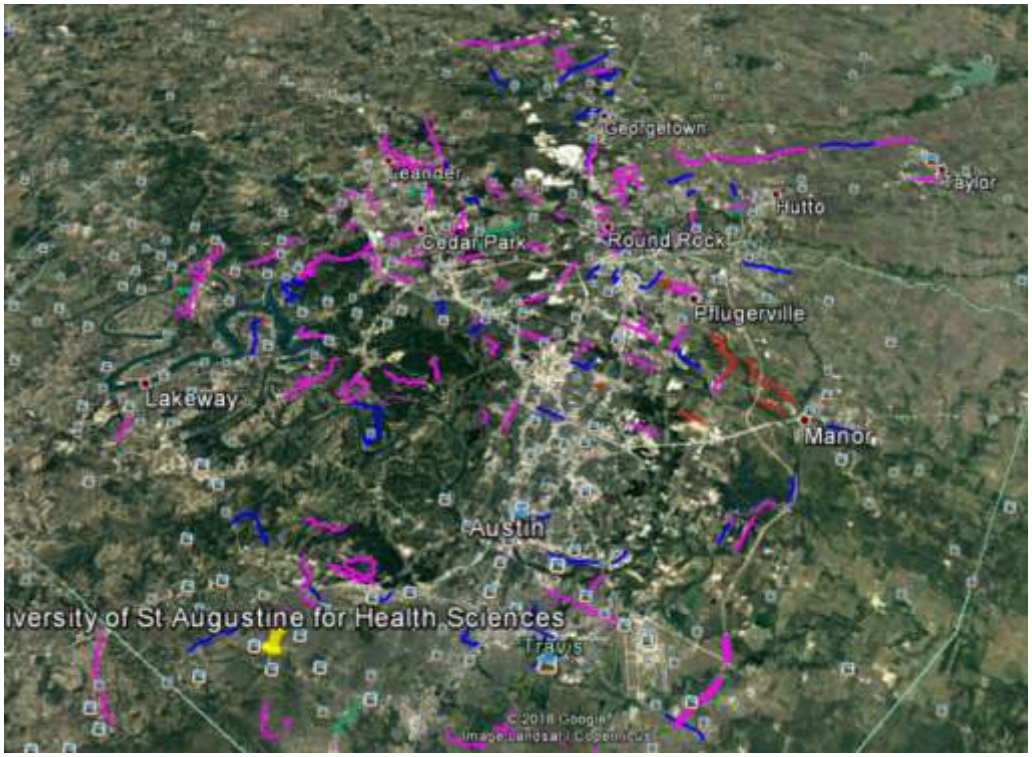


Figure 17.5 Routes with different levels of PSR in Austin Area

17.6.3. Road Surface Monitoring with the Proposed Platform

The main idea of road condition monitoring with our proposed platform is to detect the pavement roughness based on the use of IMU that includes accelerometer data. This is a continuous sensing approach that data is continuous sampled from the probe vehicle (without the explicit involvement of users). The main data we used for this process is the vertical acceleration of the vehicle along the route, and we tried to explore the possible relationship between the vertical acceleration rate and the PSR.

The first computational task for our platform is to sample the acceleration data (mainly the vertical acceleration rate) and separate them into different frequency components through the FFT algorithm. In our case, we separate those samples into six different bands which are 0.5~1.5Hz, 1.5~2.5Hz, 2.5~3.5Hz, 3.5~4.5Hz, 4.5~5.5Hz, 5.5~6.5Hz (as shown in Figure 17.6). We exclude the 0–0.5Hz band that mainly corresponds to fluctuations in the slope of the road, and is not related to the pavement condition.

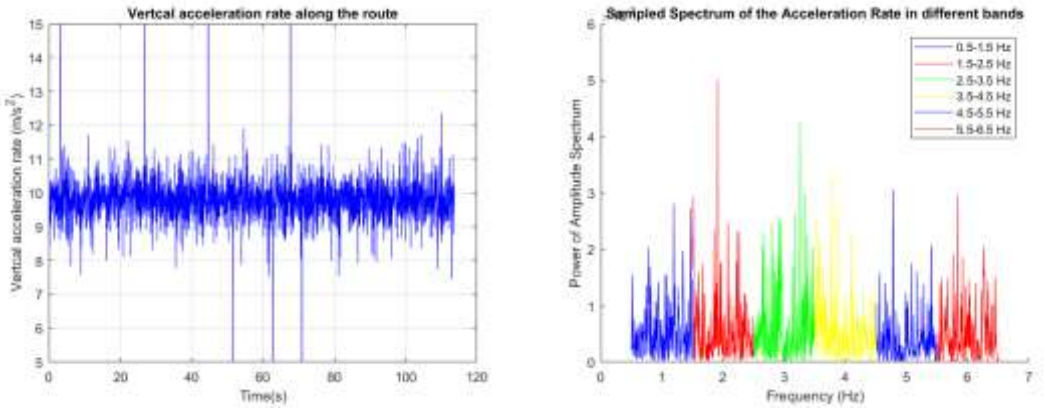


Figure 17.6 The vertical acceleration rate along time (left) and the distribution of the sampled spectrum of the acceleration rate after FFT (right)

Once we get the frequency distribution of the sampled acceleration, we can build the relationship between the data in different bands with the PSR. As running speed is a parameter that would affect the driver/passenger experience while defining the PSR value, we also considered the vehicle's speed difference while running at different routes. If 'a_z' means the average value of vertical acceleration after FFT in one band, we tried to build the relationship between "$a_z/(v^2)$" and the PSR, where 'v' stands for the average running speed along the route. Actually, during the test, we almost kept the same speed running through one route during the test (the route's speed limit) as was used when obtaining the PSR value. Figure 17.7 shows some preliminary results after collecting the data for 37 routes in the Austin area.

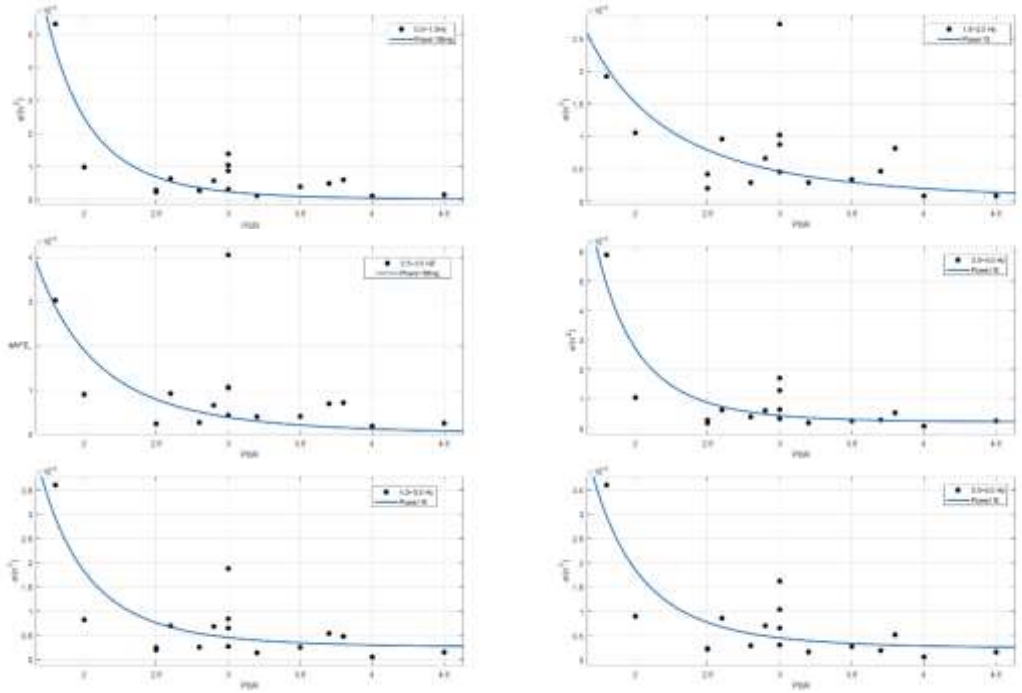


Figure 17.7 Preliminary results for the relationship between acceleration rate and PSR

In those preliminary results, we can easily see the trend that with lower vertical acceleration rate, the PSR value is higher, which means that the road surface condition is better (smoother). There are, however, several outlier points that have not been explained based on the measurement data.

Chapter 18. Cybersecurity Analysis of Connected Vehicles Using Deep Learning

18.1. Introduction

This chapter describes the research team's efforts to meet two objectives: 1) solve an inverse modeling problem (predicting the vehicle dynamics from vehicle input commands, and from vehicle measurement data) and 2) use this dynamical model to detect input faults or spoofing, or sensor faults or spoofing (particularly for GPS position sensors). For the latter problem, the higher the precision of the dynamical model, the smaller the extent to which sensors can fail or be spoofed undetected.

For this specific problem, we used the data generated by instrumented vehicles from SwRI. These vehicles monitor their states 20 times per second (with a 50-millisecond step). The measurement data is then extracted, read, and used to solve the two components of the problem.

18.2. Data Description

The data consists of a set of timeseries, containing a large number of vehicle parameters. To simplify the learning, we choose a subset of relevant parameters as predictors for the learning framework. For the present task, we have chosen the following predictors (inputs) and targets (outputs).

- **Inputs:** Break, Throttle, Steering, and Initial Speed
- **Outputs: Speed and Location (x,y) (starting from local frame origin)**

18.3. Deep Learning for Modeling Vehicle Input-Output Response

18.3.1. Deep Learning Review

Deep learning is a subset of machine learning, which mainly focuses on the optimization of artificial neural networks (ANNs) to reproduce an input-output relationship. The ANN consists of neurons organized in layers. The hyperparameters of this neural model consist in the number of neurons, number of layers of neurons, and activation functions used to describe the nonlinear aspect of the activation of each neuron. The objective of the training phase is then to determine the neurons weights to find the best possible relationship between inputs and outputs.

Deep learning focuses specifically on large ANNs that have a significant number of layers. They usually require large amounts of data and large computational power for the optimization of the neuron weights, since the associated computational cost is exponential in the number of layers.

In our approach, we use a type of recurrent neural network called LSTM (long short-term memory), which has the ability to memorize longer sequences in an encoded cell within it, called a *memory cell* (Figure 15.1).

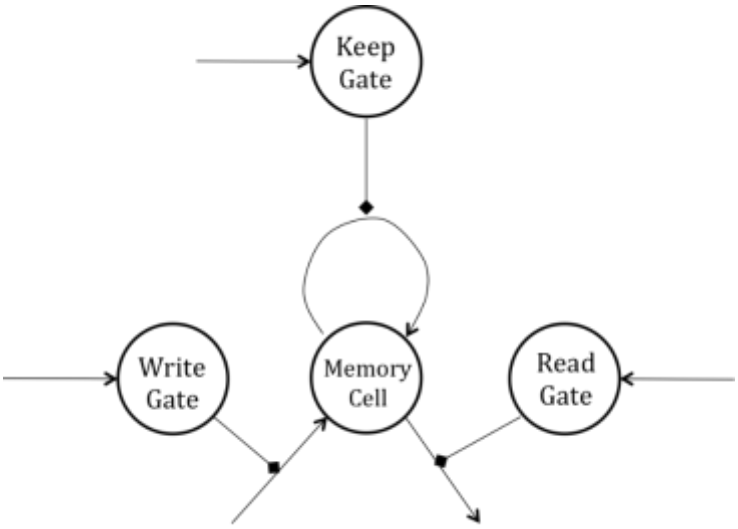


Figure 18.1 Layout of an LSTM cell

One advantage of LSTM networks is that they are immune to vanishing and exploding gradients problems found in standard ('plain vanilla') RNNs (recurrent neural networks).

18.3.2. Basic Deep Learning Principle

Following the principle of 'divide and conquer' popular in the machine learning community, we used the following steps to find a suitable model:

- First, we devised a model that overfits the training dataset, in that it is able to reproduce the input-output relationship of the training data, even if it also captures the randomness of this dataset (and thus loses robustness).
- If we could not achieve overfitting, we turned to the following techniques: normalization, scaling, using different learning rates, or using different activation functions until overfitting is achieved
- We then checked the behavior of the model on the validation dataset. If the behavior was unsatisfactory (high bias), we used different regularization techniques to reduce the number of free parameters of the model, or increased the training dataset if some features of the model were present in the validation dataset but not in the training dataset.

18.4. Data Pre-Processing

We used 250,000 samples from the vehicle original data, split into training and validation datasets.

The original data had specific problems, including missing readings by some sensors, which are replaced by interpolated values between the closest two valid readings, mis-synchronization problems, in which the data is not properly synchronized. We solved the latter issue by determining and using the average time step as a proxy for the actual time step.

The training set was normalized and shifted to be above zero (nonnegative). Normalization is an important process used in machine learning, to ensure that data elements that have different orders of magnitude can be used simultaneously as predictors.

This normalization is achieved by using the mean and standard deviation from the particular training dataset. Figures 15.2 through 15.5 provide examples of normalized data timeseries.

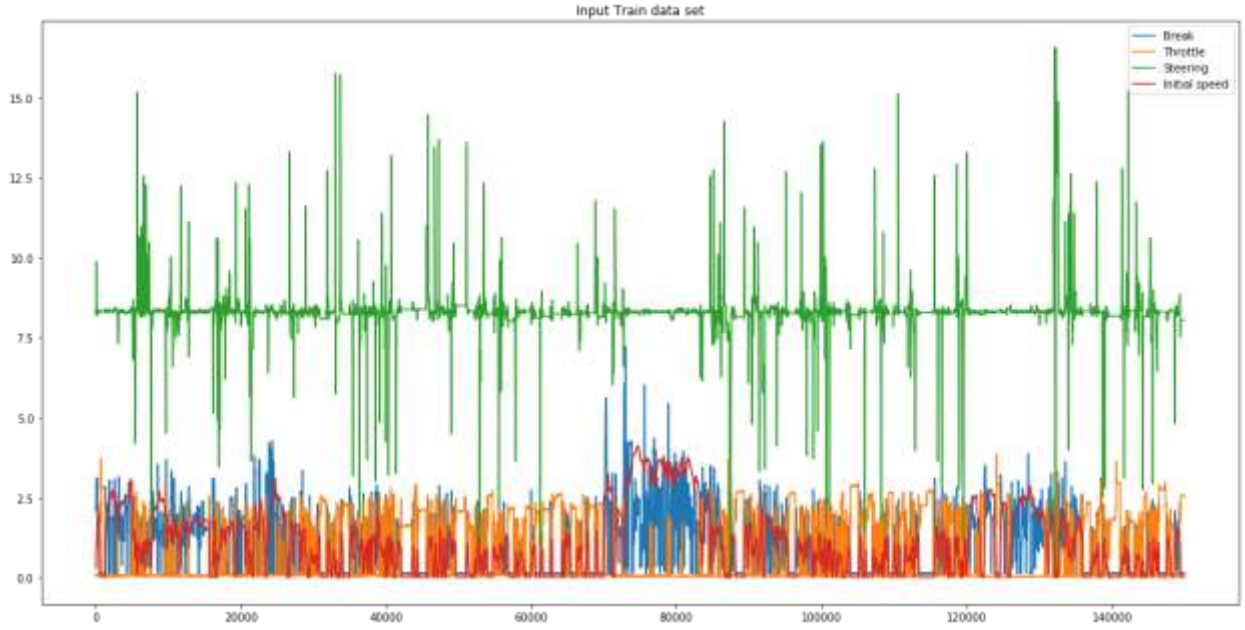


Figure 18.2 Normalized input training timeseries data

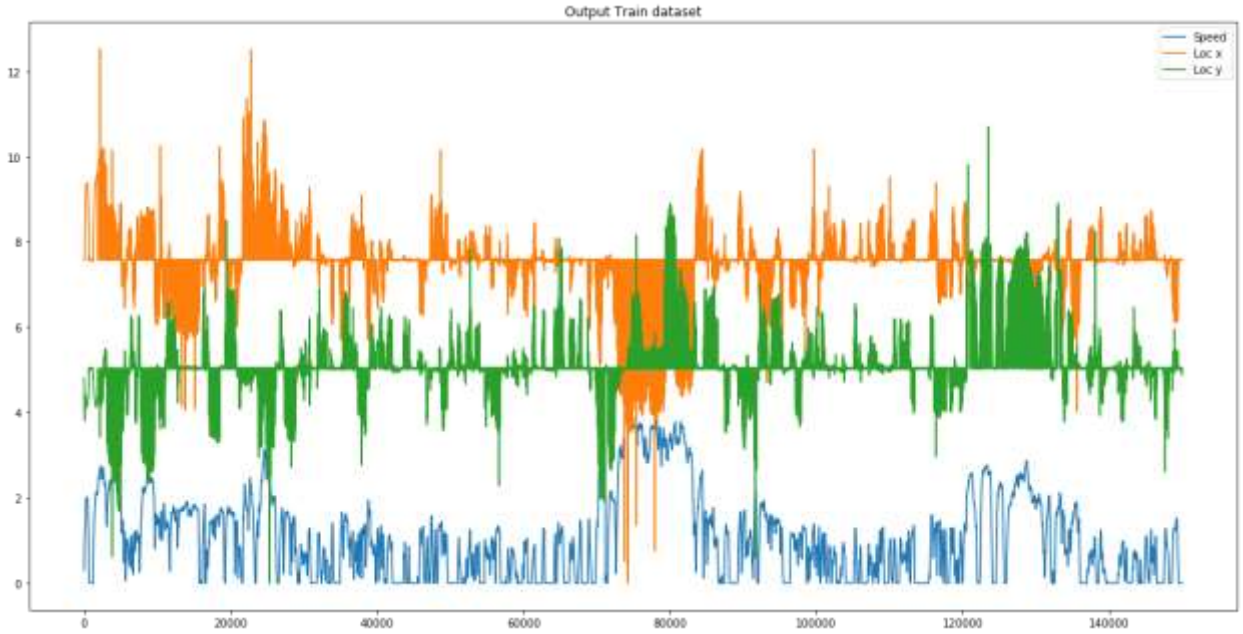


Figure 18.3 Normalized output training timeseries data

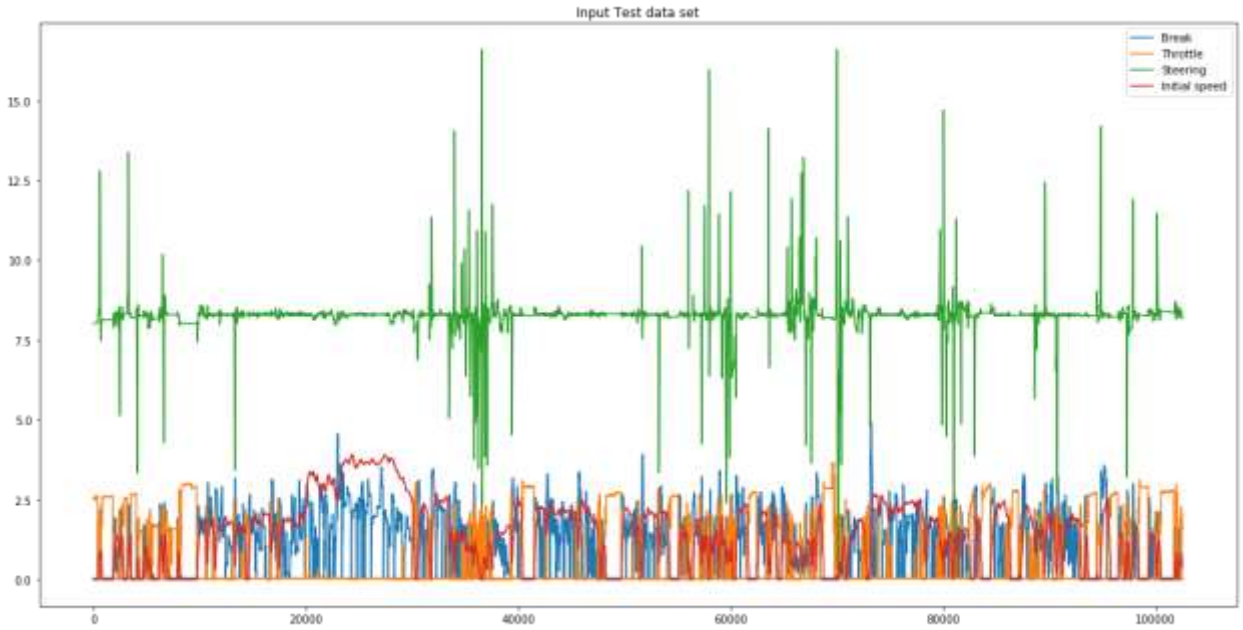


Figure 18.4 Normalized test timeseries data

Figure 18.5 Normalized output test timeseries data

18.5. Deep Model Implementation

As mentioned earlier, we used the LSTM deep model for this work. The model consists of three parts: a feature extraction layer, an LSTM layer, and two regression heads. We used a scaled exponential linear unit (SELU) activation function without the proper SELU initialization methodology, and also used a SmoothL1 Loss function. Adam optimization was used with a learning rate of 0.001, and the LSTM biases were set to 1.0 initially. This enhances the performance of LSTM.

18.5.1. SELU Functions

The SELU functions are defined as follows:

$$\operatorname{selu}(x) = \lambda \begin{cases} x & \text{if } x > 0 \\ \alpha e^x - \alpha & \text{if } x \leqslant 0 \end{cases}.$$

```
VehNet(
  (inp): Linear(in_features=4, out_features=50)
  (lstm): LSTM(50, 50, batch_first=True)
  (hidden2int): Linear(in_features=50, out_features=4)
  (intm_0): Linear(in_features=4, out_features=4)
  (intm_1): Linear(in_features=4, out_features=4)
  (intm_2): Linear(in_features=4, out_features=4)
  (op): Linear(in_features=4, out_features=1)
  (hidden2int2): Linear(in_features=4, out_features=4)
  (intm2_0): Linear(in_features=4, out_features=4)
  (intm2_1): Linear(in_features=4, out_features=4)
  (intm2_2): Linear(in_features=4, out_features=4)
  (intm2_3): Linear(in_features=4, out_features=4)
  (intm2_4): Linear(in_features=4, out_features=4)
  (intm2_5): Linear(in_features=4, out_features=4)
  (intm2_6): Linear(in_features=4, out_features=4)
  (intm2_7): Linear(in_features=4, out_features=4)
  (intm2_8): Linear(in_features=4, out_features=4)
  (intm2_9): Linear(in_features=4, out_features=4)
  (intm2_10): Linear(in_features=4, out_features=4)
  (intm2_11): Linear(in_features=4, out_features=4)
  (op2): Linear(in_features=4, out_features=2)
)
```

18.5.2. SmoothL1 Loss Function

The SmoothL1 Loss is defined as follows:

Creates a criterion that uses a squared term if the absolute element-wise error falls below 1 and an L1 term otherwise. It is less sensitive to outliers than the *MSELoss* and in some cases prevents exploding gradients (e.g. see "Fast R-CNN" paper by Ross Girshick). Also known as the Huber loss:

$$\text{loss}(x, y) = \frac{1}{n} \sum_i z_i$$

where z_i is given by:

$$z_i = \begin{cases} 0.5(x_i - y_i)^2, & \text{if } |x_i - y_i| < 1 \\ |x_i - y_i| - 0.5, & \text{otherwise} \end{cases}$$

x and *y* arbitrary shapes with a total of *n* elements each the sum operation still operates over all the elements, and divides by *n*.

18.6. Errors

The numbers contained in this section explain the rate at which errors occurred in running the LSTM deep model.

[Train] Speed loss = 0.0005266328640690547 , Location loss = 0.3718912093549646, Avg loss =0.37241784122682386

[Test] Speed loss = 0.0005441447103207576 , Location loss = 0.6851853229223736, Avg loss =0.6857294662604256

18.7. Spoof Detection Algorithm

In this application, the spoof detection algorithm is based on determining the Euclidean distance between the predicted location and the GPS location. The predicted location is obtained using the initial state, and using the inputs of the driver (of the AV controller), and the dynamical model developed in the previous section. If the Euclidean distance difference is larger than a specific threshold, then spoofing is detected. This approach is preliminary, and has several disadvantages:

- The choice of the threshold associated with the prediction error is complex, since the dynamical model of the vehicle is not perfectly known. Thus, a large prediction error is not necessarily caused by spoofing of the inputs and outputs, but could be also caused by modeling errors.

- Even if the prediction error is caused by input output spoofing, there is no possibility to determine if this spoofing is accidental (sensor or actuator fault, for example GPS fault or pedal transducer fault) or intentional (malicious spoofing).

Figure 15.6 illustrates the detection of output (sensor) spoofing on simulated positioning data. In this example, the GPS positioning data input into the system is spoofed (purple line, as opposed to the green line, on the top subfigure). The prediction error increases, and once a threshold is reached, the algorithm detects the spoofing.

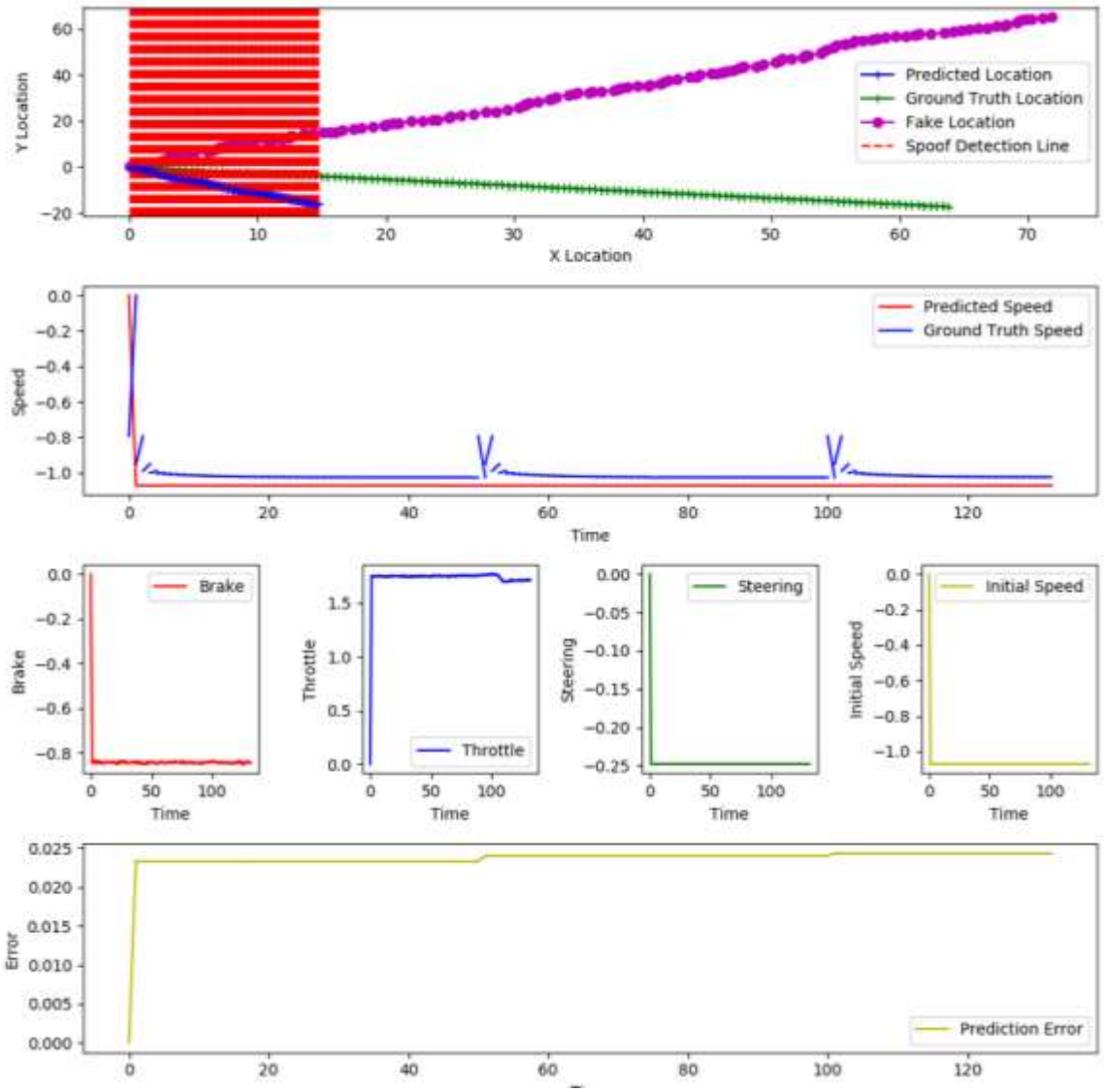


Figure 18.6 Detection of output (sensor) spoofing on simulated positioning data

Chapter 19. Prototype Development and Limited Deployment of CAV Technologies on Texas Roadways

19.1. Introduction

The previous work performed under Project 0-6838 included V2V applications for emergency vehicle alert, emergency electronic brake lights, and intelligent message propagation and V2I applications for static and dynamic wrong-way driver (WWD) detection and road condition monitoring. Leveraging these systems and applications, two core primary tasks were performed under this phase of the project. The tasks included an extension to the WWD detection system and the porting and transition of UT Austin's AIM system (detailed in Chapter 4) onto physical vehicles and infrastructure.

19.2. Roadside and Vehicle DSRC Hardware

Hardware utilized during this phase of the project included re-use of the previous onboard units (OBUs) and roadside units (RSUs) from Phase 1 and the addition of a representative traffic management center (TMC) and two Southwest Research Institute (SwRI)-owned AVs.

19.2.1. Roadside Units

The RSUs used for the demonstration were the latest release of the MK5 DSRC (dedicated short-range communications) radio from Cohda Wireless, shown in Figure 18.1. The equipment has been updated from the previous phase in order to be compliant with the currently accepted 2016 SAE and IEEE standards. As before, the RSUs are connected via PoE (Power-over-Ethernet) to the backhaul network and/or TMC system as appropriate.

Source: www.cohdawireless.com

Figure 19.1 An example of an RSU device, a Cohda MK5 RSU

19.2.2. Traffic Management Center

The representative TMC used for the demonstrations was a modified version of the TxDOT Lonestar Advanced Traffic Management System (ATMS). This included preliminary versions of a CV subsystem to facilitate handling basic safety messages (BSMs) from vehicles and sending alert messages back out to vehicles and a temporary module to execute the WWD detection algorithms.

19.2.3. Onboard Units

SwRI's CV portable onboard devices (PODs) were used for rapidly turning traditional vehicles into CVs. Examples of these PODs are shown in Figure 18.2 and Figure 18.3.

Figure 19.2 SwRI-developed PODs with all of the components necessary to turn any vehicle into a DSRC-enabled CV

Figure 19.3 PODs ready for deployment

The core component of the POD is the DSRC radio, shown in Figure 18.4. The DSRC OBU utilizes the same MK5 Cohda radio chipset as the RSU but is in a different form factor.

Figure 19.4 Example of an OBU, a Cohda Wireless MK5 DSRC unit

19.2.4. Autonomous Vehicle Architecture

Two SwRI-owned AVs were used during testing and demonstration of this phase of the project: a Freightliner Century shown in Figure 18.5 and a Ford Explorer shown in Figure 18.6. Automation of the vehicles was facilitated by integrating aftermarket drive-by-wire systems along with SwRI's existing AV software architecture to facilitate command and control of the vehicle.

Figure 19.5 SwRI Freightliner Century

Figure 19.6 SwRI Ford Explorer

19.3. Connected Vehicle Applications

Two CV applications were demonstrated during this phase of the project: Wrong-Way Driver – Safe Disable and AIM.

19.3.1. Wrong-way Driving – Safe Disable

The WWD Safe Disable demonstration builds upon the previous phase of work to detect and alert a WWD (and other nearby drivers) of the dangerous driving situation. In the previous phase, an alert was provided to the driver of the wrong-way vehicle. In this phase, the alert was still present; however, another module also received the message and issued commands to the AV control architecture, steering the vehicle to a designated safe harbor location.

Wrong-way zones were configured in the backhaul infrastructure, specifically isolating simulated highway exit ramps and a portion of a simulated one-way road. For this phase, the backhaul infrastructure was running the modified Lonestar ATMS described above. Vehicles were traversing the local area broadcast BSMs at the standard rate of 10Hz. BSMs were received by the RSU and forwarded to the ATMS, where they were processed and evaluated against the configured wrong-way zones. Once a vehicle was determined to be driving the wrong way, an alert was generated and broadcast back to the vehicle via the RSU. A process onboard the vehicle displayed the alert to the driver and sent a command to the AV architecture that disables the driver's control (steering, brake, and throttle) and steers the vehicle to a safe harbor location included in the WWD alert message.

Figure 18.7 shows an example in which a vehicle was detected driving the wrong way up an exit ramp, from the right. Once detected by the infrastructure, an alert message was created that included the suggested route the vehicle should follow to the designated safe harbor location. The route was encoded using the message structure shown in Figure 18.8.

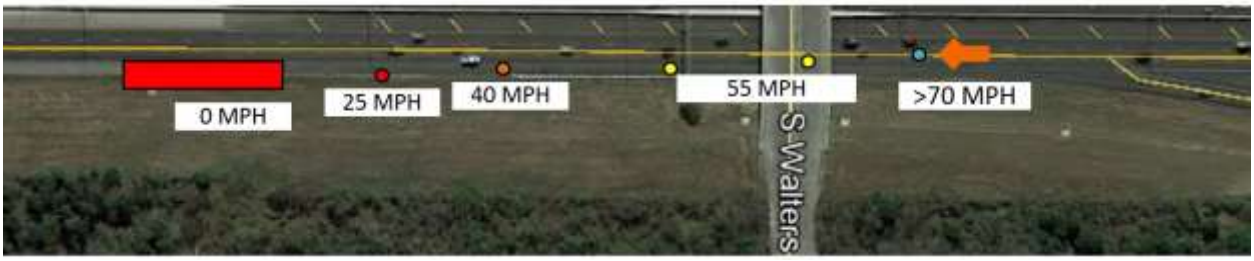


Figure 19.7 Waypoints with desired speed. Blue: 70+, Yellow: 55, Orange: 40, Red: 25, and the safe harbor location

```
-- /--------------------------------------------------------\
-- |List of suggested waypoints to follow for a safe stop   |
-- | Implemented as a sequence to allow for expansion       |
-- \--------------------------------------------------------/
SuggestedWaypoints ::= SEQUENCE {
    waypoints       NodeSetLL -- sequence of 2..63 lat/lons along the route
}

-- /--------------------------------------------------------\
-- |Speed to follow for a safe stop                         |
-- | If tempId is specified but no area is provided,        |
-- | the tempId vehicle should respect the currently        |
-- | transmitted speedLimit.                                |
-- | If an area is provided but no tempId, all vehicles     |
-- | that travel in that area should respect the speedLimit.|
-- | If neither is provided, speedLimit is a max speed for  |
-- | all vehicles within reception range.                   |
-- \--------------------------------------------------------/
SpeedZones ::= SEQUENCE {
    tempId          TemporaryID OPTIONAL, -- BSM tempId of the vehicle that should follow the indicated speed limit
    area CHOICE OPTIONAL {
         roadSegments   RoadSegmentList,
         mapData    MapData,
         },
    speedLimit  RegulatorySpeedLimit
}
```

Figure 19.8 Waypoint message frame

19.3.2. Autonomous Intersection Management

The AIM demonstration is a part of the simulation system developed by UT CTR. The concept behind the AIM system is to demonstrate an infrastructure process that can manage vehicle flow through an intersection such that traditional signal controllers are not necessary.

In this demonstration, a control interface was integrated into the SwRI AV architecture allowing a driver to request a trajectory through an intersection. A tablet displayed a map of the SwRI test track to the driver, seen in Figure 18.9, on which the driver selected a target destination. A process onboard the vehicle calculated a route based on a known underlying road network. The route was encoded into the trajectory request and sent to the infrastructure for evaluation. The request was sent via the DSRC OBU to the RSU, which forwarded the message to a backhaul process running the core AIM algorithms. The request message structure is shown in Figure 18.10.

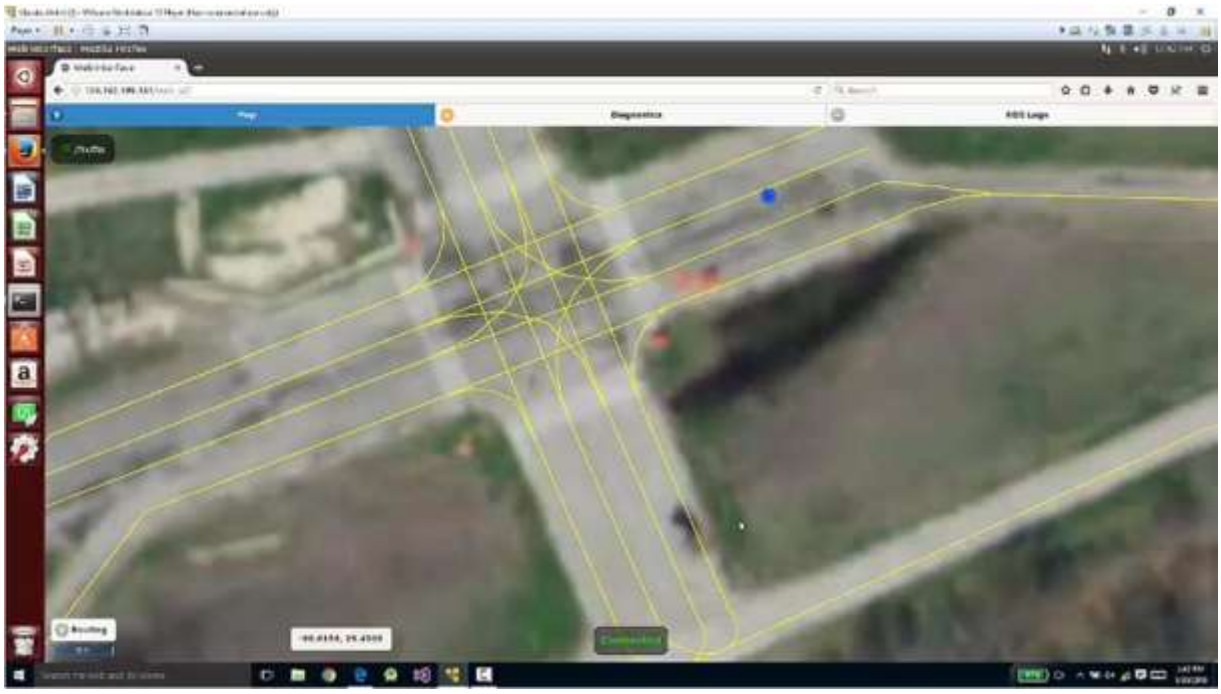


Figure 19.9 In-vehicle AIM interface

```
TrajectoryRequest ::= SEQUENCE {
        tempId              TemporaryID, -- vehicle generated 32-bit ID, should match BSM tempId
        requestTime         DSecond, -- current time (milliseconds within the minute, should match BSM time)
        vehicleSize         VehicleSize, -- length and width in cm
        accelCapacity   Acceleration    OPTIONAL, -- anticipated accel capabilities of the vehicle, if provided, infrastructure may provide speed recommendations
        decelCapacity   Acceleration    OPTIONAL, -- anticipated decel capabilities of the vehicle, if provided, infrastructure may provide speed recommendations
        targetSpeed         Speed, -- requested speed throughout trajectory (assuming from first waypoint onward)
        firstPointTime  DSecond, -- now if first point is vehicle's current position, otherwise calculation of time at which vehicle will reach first node
        waypoints           NodeSetLL -- sequence of lat/lons that define the segments
}
```

Figure 19.10 Abstract Syntax Notification (ASN) definition of a TrajectoryRequest

Within the infrastructure process, the requested path is projected forward in time and evaluated for collisions against other previously approved trajectories. If a collision is detected, a message is sent to the vehicle indicating the rejection. The response message structure (both for approving and rejecting a request) is shown in Figure 18.11. Upon receiving the rejection, the driver can change the approach parameters (generally the current speed) and request passage again. This process is repeated until a rejection is not received. On the infrastructure, once the requested path is approved, it is added to a list of approved trajectories and is included in the list with which later requests from other vehicles are compared.

```
TrajectoryResponse ::= SEQUENCE {
        tempId                  TemporaryID,
        requestTime             DSecond, -- relativeH unique ID to correlate a vehicle request
        approved                INTEGER(0..1), -- 0: false; 1: true
        recommendedSpeeds   SEQUENCE (SIZE(1..4)) OF Speed OPTIONAL, -- limited set of speeds that the infrastructure can provide as recommendations
        waypoints               NodeSetLL
}
```

Figure 19.11 ASN definition of a TrajectoryResponse

The proposed high-level system architecture is shown in Figure 18.12. Up to the point of the demonstrations in December 2017, the AV control had not been integrated into the trajectory request and response process. Rather than autonomous control, a human driver would be provided speed and trajectory recommendations to follow once a route was approved. A rejected route request would be displayed to the driver as a red line overlaid on the map (following the calculated path that was requested), shown in Figure 18.13, while an approved route request would be displayed as a green line overlaying the path to the destination, shown in Figure 18.14.

Unseen to the vehicle, the backhaul system also tracks the vehicle's BSMs and compares them against the vehicle's approved route to watch for deviations from the requested speed or path. No action was taken at this time in the system; however, an operator on the backhaul system would be notified of the deviation. It is expected that the system would generate a message to send to the vehicle when the deviation exceeded a defined threshold, revoking a previously approved route plan through the intersection if a situation was determined to be unsafe; i.e., a collision was more likely at that point due to the path and/or speed deviation.

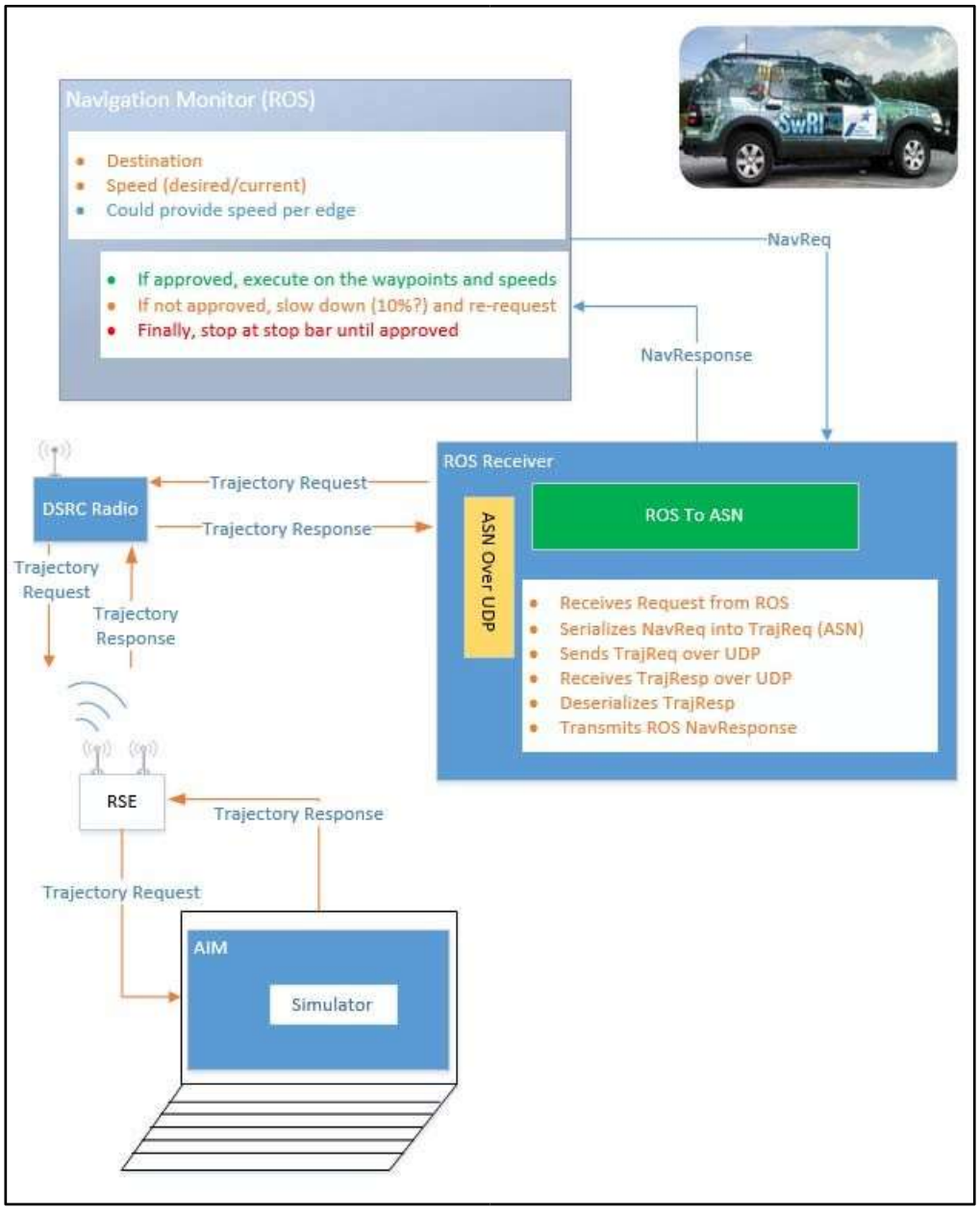


Figure 19.12 System architecture using AIM on one of SwRI's automated vehicles

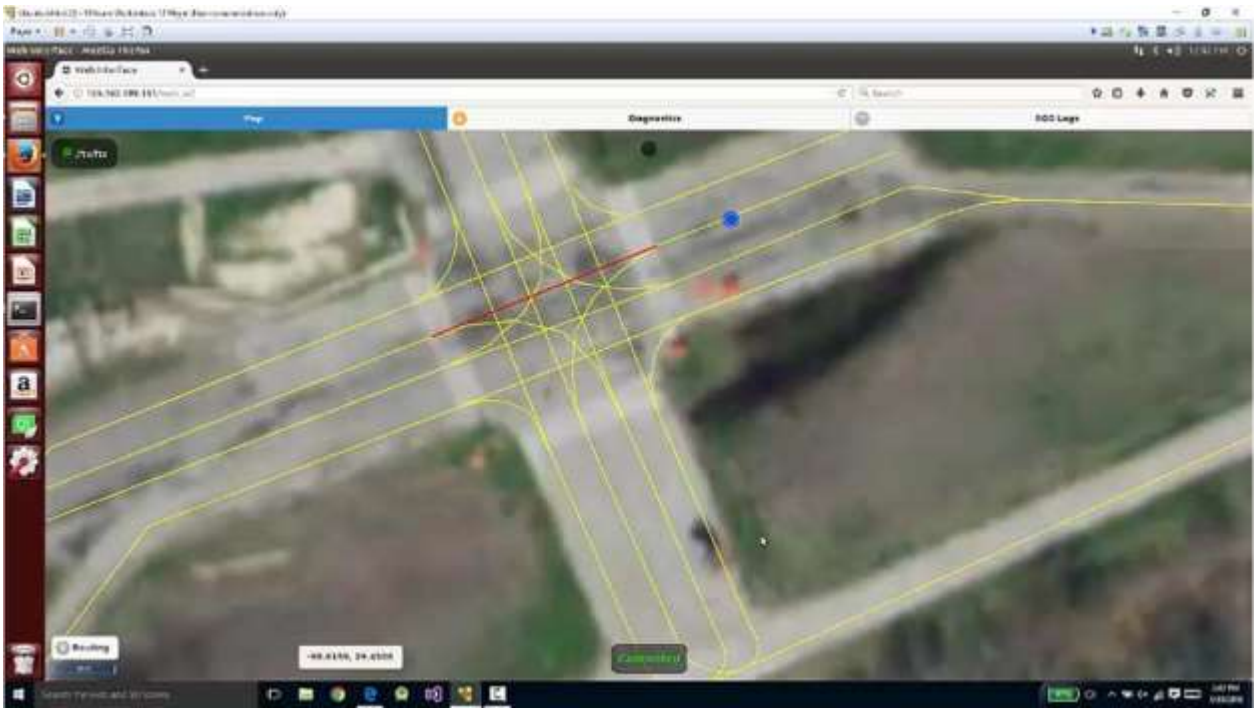


Figure 19.13 Route rejected

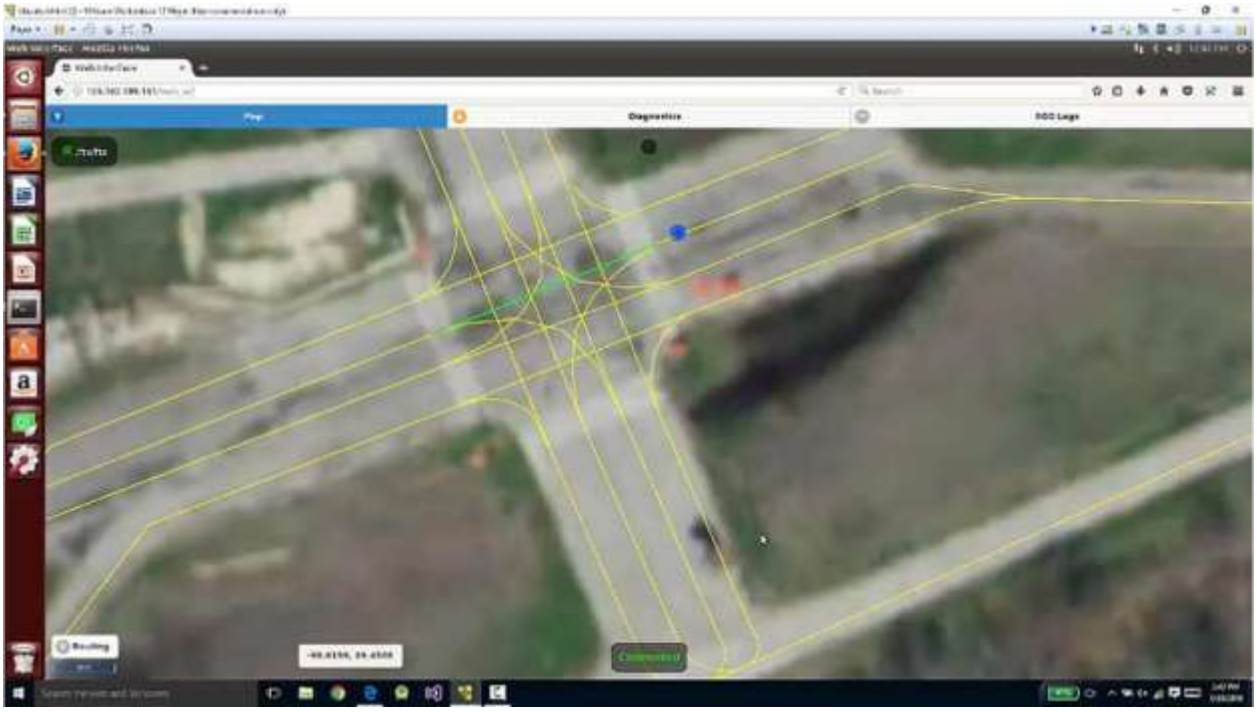


Figure 19.14 Route approved

19.4. Conclusion

The high-level design and architecture was primarily completed for both systems and demonstrated in December 2017 at the SwRI test track.

The WWD safe disable system, including integration into the AV control, was complete and only needed minor changes to resolve communications issues that presented themselves during testing and demonstration. The system was successfully able to monitor vehicle movement nearby and through the configured regions, identify specific vehicles driving the wrong way, alert nearby vehicles of the unsafe situation, and command the wrong-way vehicle to drive to a safe location, which the vehicle was able to do.

The AIM system involved more moving pieces and a more complicated architecture. The system, as demonstrated, included initial integration with the AV control architecture; however, it did not autonomously control the vehicle and instead provided information and guidance to a driver in the vehicle. At the driver's request, the system would automatically calculate a route to a destination and send the route to the backhaul system for evaluation. The backhaul system successfully evaluated the requested route based on currently approved routes from other vehicles and provided the acceptance or rejection to the requesting vehicle, which was then presented back to the driver.

References

American Automobile Association (2016) Driving Costs Hit Six-Year Low, Finds AAA. Retrieved from: https://newsroom.aaa.com/2016/04/driving-costs-hit-six-year-low-finds-aaa-2/

Abraham, H., Lee, C., Brady, S., Fitzgerald, C., Mehler, B., Reimer, B. and Coughlin, J. (2016). Autonomous Vehicles, Trust, and Driving Alternatives: A survey of consumer preferences. MIT AgeLab White Paper (2016-6). AgeLab, Massachusetts Institute of Technology. Retrieved from: http://agelab.mit.edu/files/publications/2016_6_Autonomous_Vehicles_Consumer_Preferences.pdf (July 20, 2017).

Abu-Lebdeh, G. and R. Benekohal (1997). Development of traffic control and queue management procedures for oversaturated arterials. *Transportation Research Record* (1603), 119–127.

Agatz, Niels, Erera, Alan, Savelsbergh, Martin and Wang, Xing. (2011). Dynamic ride-sharing: A simulation study in metro Atlanta. *Procedia Social and Behavioral Sciences* 17: 532-550.

Ahuja, R. K., T. L. Magnanti, and J. B. Orlin (1993, February*). Network Flows: Theory, Algorithms, and Applications* (1 ed.). Prentice Hall.

G. Alesssandretti, A. Broggi, and P. Cerri.Vehicle and guard rail detection using radar and vision data fusion (2007). Intelligent Transportation Systems, IEEE Transctions, vol.8, no. 1, pp. 95-105.

Alexander, L., Jiang, S., Murga, M., and Gonzalez, M. C. (2015). Origin-Destination Trips by Purpose and Time of Day Inferred from Mobile Phone Data. *Transportation Research Part C: Emerging Technologies* 58:240-250.

Alonso-Mora, Javier, Samaranayake, Samitha, Wallar, Alex, Frazzoli, Emilio and Rus, Daniela. (2016). On-demand high-capacity ride-sharing via dynamic trip-vehicle assignment. Proceedings of the National Academy of Sciences of the United States of America 114(3): 462-467.

Amey, A., Attanucci, J. and Mishalani, R. (2014). Real-time Ridesharing: Opportunities and Challenges in Using Mobile Phone Technology to Improve Rideshare Services. *Transportation Research Record* 2217: 103-110. DOI: 10.3141/2217-13.

Anderson, J. M., K. Nidhi, K. D. Stanley, P. Sorensen, C. Samaras, and O. A. Oluwatola. (2014). Autonomous vehicle technology: A guide for policymakers. Rand Corporation. Online at: http://www.rand.org/content/dam/rand/pubs/research_reports/RR400/RR443-2/RAND_RR443-2.pdf

Andreatta, G. and Romeo, L., (1988). Stochastic shortest paths with recourse. *Networks*, 18(3), pp.193-204.

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

Arnott, R., A. De Palma, and R. Lindsey (1993). A structural model of peak-period congestion: A traffic bottleneck with elastic demand. *The American Economic Review*, 161–179.

Arnott, R. (1998). William Vickrey: contributions to public policy. *International Tax and Public Finance*, 5(1), 93-113.

Atlanta Regional Commission, (2012). Activity-Based Travel Model Specifications: Coordinated Travel–Regional Activity Based Modeling Platform (CT-RAMP) for the Atlanta Region. Online at: http://documents.atlantaregional.com/The-Atlanta-Region-s-Plan/RTP/abm-specification-report.pdf

Au, T.-C., N. Shahidi, and P. Stone (2011). Enforcing liveness in autonomous traffic management. In AAAI.

Au, T.-C., S. Zhang, and P. Stone (2015). Autonomous intersection management for semi-autonomous vehicles. The Routledge Handbook of Transportation, 88–104.

Aubin, J., A.M Bayen, & P. Saint-Pierre. (2008). Dirichlet problems for some Hamilton-Jacobi equations with inequality constraints. SIAM *Journal on Control and Optimization* 47.5: 2348-2380.

A. Bachrach, S. Prentice, R. He, and N. Roy (2011). "Range-Robust antonomous navigation in GPS-denied environments", Journal of Field Robotics, vol. 28, no. 5, pp.644-666.

Baiti, P., & Jenssen, H. K. (1998). On the front-tracking algorithm. *Journal of Mathematical Analysis and Applications*, 217(2), 395-404.

Balmer, M., Axhausen, K., & Nagel, K. (2006). Agent-based demand-modeling framework for large-scale microsimulations. *Transportation Research Record* (1985), 125-134.

Bansal, P., K. M. Kockelman, and Y. Wang. (2015). Hybrid Electric Vehicle Ownership and Fuel Economy Across Texas: An Application of Spatial Models. *Transportation Research Record* No. 2495: 53-64.

Bansal, Prateek and Kockelman, Kara. (2016). Are We Ready to Embrace Connected & Self-Driving Vehicles? A Case Study of Texans. *Transportation* 44: 1-35.

Bansal, Prateek and Kockelman, Kara. (2017). Forecasting Americans' Long-Term Adoption of Connected and Autonomous Vehicle Technologies. *Transportation Research Part A* 95: 49-63.

Bar-Gera, H., F. Hellman, and M. Patriksson (2013). Computational precision of traffic equilibria sensitivities in automatic network design and road pricing. *Transportation Research Part B: Methodological* 57, 485– 500.

Bar-Gera, H. (2014). Transportation test problems. Website: http://www.bgu.ac.il/~bargera/tntp/. Accessed April 28, 2009.

Beckmann, M., C. B. McGuire, and C. B. Winsten (1956). Studies in the Economics of *Transportation*. Yale University Press.

Bento, L. C., R. Parafita, and U. Nunes (2012). Intelligent traffic management at intersections supported by v2v and v2i communications. In Intelligent Transportation Systems (ITSC), 2012 15th International IEEE Conference on, pp. 1495–1502. IEEE.

Bento, L. C., R. Parafita, S. Santos, and U. Nunes (2013). Intelligent traffic management at intersections: Legacy mode for vehicles not equipped with v2v and v2i communications. In Intelligent Transportation Systems-(ITSC), 2013 16th International IEEE Conference on, pp. 726–731. IEEE.

Berbeglia, Gerardo, Cordeau, Jean-Francois, and Laporte, Gilbert. (2010). Dynamic pickup and delivery problems. *European Journal of Operational Research* 202(1): 8-15.

Bertsekas, D. P. (1999). Nonlinear programming. Athena scientific Belmont.

Bhat, Suraj. (2016). Quantifying the Potential for Dynamic Ride-Sharing of New York City's Taxicabs (Undergraduate Thesis). Retrieved from: http://orfe.princeton.edu/~alaink/SmartDrivingCars/Papers/Bhat,Suraj_Final_Thesis2016.pdf (June 30, 2017).

Bischoff, J., & Maciejewski, M. (2016). Simulation of city-wide replacement of private cars with autonomous taxis in Berlin. *Procedia Computer Science*, 83, 237-244.

Bischoff, J., Soeffker, N. and Maciejewski, M. (2016). A framework for agent based simulation of demand responsive transport systems. Retrieved from: http://dx.doi.org/10.14279/depositonce-5760 (June 30, 2017).

Bonnefon, J., Shariff, A. and Rahwan, I. (2016). The social dilemma of autonomous vehicles. *Science* 352(6293): 1573-1576.

Boyles, S. D. (2009). Operational, Supply-Side Uncertainty in Transportation Networks: Causes, Effects, and Mitigation Strategies. Ph. D. thesis, The University of Texas at Austin.

Boyles, S. D., K. M. Kockelman, and S. T. Waller (2010). Congestion pricing under operational, supply-side uncertainty. Transportation Research Part C: Emerging Technologies 18(4), 519 – 535.

Boyles, S.D. and Waller, S.T., (2011). Optimal information location for adaptive routing. *Networks and Spatial Economics*, 11(2), pp.233-254.

Boyles, S. D., L. M. Gardner, and H. Bar-Gera (2015). Incorporating departure time choice into highoccupancy/toll (hot) algorithm evaluation. Transportation Research Procedia 9, 90 – 105. Papers selected for Poster Sessions at The 21st International Symposium on Transportation and Traffic Theory Kobe, Japan, 5-7 August, 2015.

M. Braberger, J. Brunner, B. Rinner and H. Schwabach (2004). Real-time video analysis on an embedded smart caera for traffic surveillance. Real-time and Embedded Technology and Applications Symposium, IEEE, pp. 174-181.

Bressan, A. (2000). Hyperbolic systems of conservation laws: the one-dimensional Cauchy problem (Vol. 20). Oxford University Press on Demand.

Bretti, G., Natalini, R., & Piccoli, B. (2006). Numerical approximations of a traffic flow model on networks. NHM, 1(1), 57-84.

Brownell, Chris and Kornhauser, Alain. (2014). A Driverless Alternative: Fleet Size and Cost Requirements for a Statewide Autonomous Taxi Network in New Jersey. *Transportation Research Record* 2416: 73-81.

Buhr, S. (2017). Lyft launches a new self-driving division and will develop its own autonomous ride-hailing technology. Tech Crunch.com (July 21). Retrieved at: https://techcrunch.com/2017/07/21/lyft-launches-a-new-self-driving-division-called-level-5-will-develop-its-own-self-driving-system/

Burns, L. D., Jordan, W. C., & Scarborough, B. A. (2013). Transforming personal mobility. *The Earth Institute*, 431, 432.

Calabrese, F., Di Lorenzo, G., Liu, L., and Ratti, C. 2011. Estimating Origin-Destination Flows using Opportunistically Collected Mobile Phone Location Data from One Million Users in Boston Metropolitan Area. IEEE Pervasive Computing 10(4):36-44.

California Department of Motor Vehicles (DMV) (2017a) Proposed Driverless Testing and Deployment Regulations. Initial Statement of Reasons. March 10, 2017. Accessed at: https://www.dmv.ca.gov/portal/dmv/detail/vr/autonomous/auto.

California Department of Motor Vehicles (2017b). Proposed Driverless Testing and Deployment Regulations. Autonomous Vehicles Express Terms. March 10, 2017. Accessed at: https://www.dmv.ca.gov/portal/dmv/detail/vr/autonomous/auto.

California Department of Motor Vehicles (2018a). February 26, 2018. Order to Adopt Testing of Driverless Autonomous Vehicles. Accessed at: https://www.dmv.ca.gov/portal/wcm/connect/a6ea01e0-072f-4f93-aa6c-e12b844443cc/DriverlessAV_Adopted_Regulatory_Text.pdf?MOD=AJPERES

California Department of Motor Vehicles (2018b). Driverless Testing of Autonomous Vehicles Final Statement of Reasons. February 26, 2018. Accessed At: https://www.dmv.ca.gov/portal/wcm/connect/e11d4dd0-e5ec-453f-8861-41bf8656a69c/DriverlessAV_Final_Statement_of_Reasons.pdf?MOD=AJPERES

CAMPO. (2015). CAMPO 2010 Planning Model Guide, Capital Area Metropolitan Planning Organization (CAMPO), Austin, TX. Online at: http://www.campotexas.org/plans-programs/

E. S. Canepa and C. G. Claudel, "Exact solutions to traffic density estimation problems involving the Lighthill-Whitham-Richards traffic flow model using mixed integer programming," in 2012 15th International IEEE Conference on Intelligent Transportation Systems. IEEE, 2012, pp. 832–839.

E. Canepa, E. Odat, A. Dehwah, M. Mousa, J. Jiang and C. Claudel (2014). "A sensor network architecture for urban traffic state estimation with mixed Eulerian/ Lagrangian sensing based on distributed Computing", Architecture of Computing Systems- ARCS 2014. Springer, pp. 147-158.

Edward S. Canepa, Christian G. Claudel (2017). Networked traffic state estimation involving mixed fixed-mobile sensor data using Hamilton-Jacobi equations. Transportation Research Part B, 104 pp. 686-709.

Carey, M. and A. Srinivasan (1993). Externalities, average and marginal costs, and tolls on congested networks with time-varying flows. *Operations Research* 41(1), 217–231.

Carlino, D., S. D. Boyles, and P. Stone (2013). Auction-based autonomous intersection management. In Intelligent Transportation Systems-(ITSC), 2013 16th International IEEE Conference on, pp. 529–534. IEEE.

Castiglione, J., M. Bradley, and J. Gliebe. (2015). Activity-based travel demand models: a primer. No. SHRP 2 Report S2-C46-RR-1. Online at: https://www.nap.edu/catalog/22357/activity-based-travel-demand-models-a-primer

Chan, N. and Shaheen, S. (2012). Ridesharing in North America: Past, present, and future. Transport Reviews 32(1): 93-112. Retrieved from: http://tsrc.berkeley.edu/sites/default/files/Ridesharing%20in%20North%20America%20Past%20%20Present%20%20and%20Future.pdf (July 20, 2017).

Chapin, D., R. Brodd, G. Cowger, J. Decicco, G. Eads, R. Espino, J. German, D. Greene, J. Greenwald, and L. Hegedus. (2013). Transitions to Alternative Vehicles and Fuels. National Academies Press, Washington, DC. Online at: https://www.nap.edu/catalog/18264/transitions-to-alternative-vehicles-and-fuels

Chen, D., and K. M. Kockelman. (2016a). Management of a Shared Autonomous Electric Vehicle Fleet: Implications of Pricing Schemes. *Transportation Research Record* 2572: 37-46.

Chen, D., K. M. Kockelman, and J. Hanna (2016b). Operations of a Shared, Autonomous, Electric Vehicle (SAEV) Fleet: Implications of Vehicle & Charging Infrastructure Decisions. *Transportation Research Part A: Policy and Practice* 94: 243-254.

Chen, H., B. An, G. Sharon, J. P. Hanna, P. Stone, C. Miao, and Y. C. Soh (2018, February). Dyetc: Dynamic electronic toll collection for traffic congestion alleviation. In Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI-18).

Cheng, D., W. Li, and S. Ishak (2014). Accounting for travel time reliability and trip purpose in an agentbased approach to toll pricing with dynamic feedback control. *Transportation Recor*d 2470, 131–141.

Chiu, Y., J. Bottom, M. Mahut, A. Paz, R. Balakrishna, T. Waller, and J. Hicks (2011). Dynamic traffic assignment: A primer. *Transportation Research E-Circular* (E-C153).

Claudel, C. G., & Bayen, A. M. (2010). Lax–Hopf based incorporation of internal boundary conditions into Hamilton–Jacobi equation. Part I: Theory. Automatic Control, IEEE Transactions on, 55(5), 1142-1157.

Claudel, C. G., & Bayen, A. M. (2010). Lax–Hopf based incorporation of internal boundary conditions into hamilton-jacobi equation. part II: Computational methods. Automatic Control, IEEE Transactions on, 55(5), 1158-1174.

Cohen, Boyd, and Jan Kietzmann (2014). Ride on! Mobility business models for the sharing economy. *Organization & Environment* 27.3: 279-296.

Coifman, B., D. Beymer, P. McLauchlan, and J. Malik (1998). A real-time computer vision system for vehicle tracking and traffic surveillance. *Transportation Research Part C: Emerging Technologies* 6(4), 271–288.

Correia, G. H., & van Arem, B. (2016). Solving the User Optimum Privately Owned Automated Vehicles Assignment Problem (UO-POAVAP): A model to explore the impacts of self-

driving vehicles on urban mobility. *Transportation Research Part B: Methodological*, 87, 64-88.

Cragg, J. (1971). Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods. *Econometrica* 39(5): 829-844.

Cristiani, E., de Fabritiis, C., & Piccoli, B. (2010). A fluid dynamic approach for traffic forecast from mobile sensor data. Communications in Applied and Industrial Mathematics, 1(1), 54-71.

Cyganski, R., Fraedrich, E., & Lenz, B. (2015). Travel time valuation for automated driving: a use-case-driven study. In 94th annual meeting of the transportation research board. Washington, DC: Transportation Research Board.

Dafermos, S. C. and F. T. Sparrow (1969). The traffic assignment problem for a general network. *Journal of Research of the National Bureau of Standards B* 73(2), 91–118.

Daganzo, C. F. (1994a). The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transportation Research Part B: Methodological*, 28(4), 269-287.

Daganzo, C. F. (1994b). The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transportation Research Part B: Methodological* 28(4), 269–287.

Daganzo, C. F. (1995a). The cell transmission model, part II: network traffic. *Transportation Research Part B: Methodological*, 29(2), 79-93.

Daganzo, C. F. (1995b). A finite difference approximation of the kinematic wave model of traffic flow. *Transportation Research Part B* 29(4), 261–276.

Daganzo, C. F. (1999). The lagged cell-transmission model.

Daganzo, C.F. (2005). A variational formulation of kinematic waves: basic theory and complex boundary conditions. *Transportation Research Part B: Methodological* 39.2, 187-196.

Daganzo, C.F. (2006). On the variational theory of traffic flow: well-posedness, duality and applications. UC Berkeley Center for Future Urban Transport: A Volvo Center of Excellence.

Dai, Chengcheng. (2016). Ridehsaring Recommendation: Whether and Where Should I Wait? Proceedings of the International Conference on Web-Age Information Management, Nanchang, China. Retrieved from: https://link.springer.com/content/pdf/10.1007%2F978-3-319-39937-9_12.pdf (July 20, 2017).

Danielis, R. and E. Marcucci (2002). Bottleneck road congestion pricing with a competing railroad service. *Transportation Research Part E: Logistics and Transportation Review* 38(5), 379–388.

De Palma, A., & Lindsey, R. (2011). Traffic congestion pricing methodologies and technologies. *Transportation Research Part C: Emerging Technologies*, 19(6), 1377-1399.

De Palma, A., Lindsey, R., & Quinet, E. (2004). Time-varying road pricing and choice of toll locations. Road pricing: Theory and evidence, *Research in Transportation Economics*, 9, 107-131.

Deloitte. (2014). Global Automotive Consumer Study Exploring consumers' mobility choices and transportation decisions. Retrieved from: https://www.autonews.com/assets/PDF/CA92618116.PDF (July 10, 2017).

Dial, R. B. (2006). A path-based user-equilibrium traffic assignment algorithm that obviates path storage and enumeration. *Transportation Research Part B* 40(10), 917–936.

Dietrich, B. (1969). Uber ein Paradoxon aus der Verkehrsplanung.¨Unternehmensforschung 12, 258–268.

Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik* 1(1), 269–271.

Dresner, K. and P. Stone (2004). Multiagent traffic management: A reservation-based intersection control mechanism. In Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems-Volume 2, pp. 530–537. IEEE Computer Society.

Dresner, K. and P. Stone (2006). Human-usable and emergency vehicle-aware control policies for autonomous intersection management. In Fourth International Workshop on Agents in Traffic and Transportation (ATT), Hakodate, Japan.

Dresner, K., & Stone, P. (2008). A multiagent approach to autonomous intersection management. *Journal of Artificial Intelligence Research*, 31, 591-656.

Dunfan Ye, Daoli Gong, Wei Wang. Application of wireless sensor networks in environmental monitoring. Power Electronics and Intelligent Transportation System (PEITS), 2009 2nd International Conference, pp. 1008-1112.

Ekstrm, J., I. Kristoffersson, and N.-H. Quttineh (2016). Surrogate-based optimization of cordon toll levels in congested traffic networks. *Journal of Advanced Transportation* 50(6), 1008–1033.

El Bsat, S., H. Bou-Ammar, and M. E. Taylor (2017). Scalable multitask policy gradient reinforcement learning. In AAAI, pp. 1847–1853.

Fagnant, D. J., and K. Kockelman. (2015). Preparing a Nation for Autonomous Vehicles: Opportunities, Barriers and Policy Recommendations. *Transportation Research Part A: Policy and Practice* 77: 167-181.

Fagnant, D. J., Kockelman, K. M., & Bansal, P. (2015). Operations of shared autonomous vehicle fleet for austin, texas, market. *Transportation Research Record* (2536), 98-106.

Fagnant, Daniel J. and Kockelman, Kara M. (2016). Dynamic ride-sharing and fleet sizing for a system of shared autonomous vehicles. *Transportation* 45: 1-16.

Fajardo, D., T.-C. Au, S. Waller, P. Stone, and D. Yang (2011). Automated intersection control: Performance of future innovation versus current traffic signal control. *Transportation Research Record* (2259), 223–232

Fangfang Zheng, Henk Van Zuylen (2013). Urban link travel time estimation based on sparse probe vehicle data. Transportation Research Part C, 31, pp. 145-157.

Farahani, R. Z., E. Miandoabchi, W. Y. Szeto, and H. Rashidi (2013). A review of urban transportation network design problems. *European Journal of Operational Research* 229(2), 281–302.

Farhan, Javed and Chen, T. Donna. (2017). Impact of Ridesharing on Operational Efficiency of Shared Autonomous Electric Vehicle Fleet. Under review for publication in *Transportation Research Part C: Emerging Technologies*.

Federal Communications Commission (2017a). Report and Order, July 13, 2017 FCC 17-94A1. Accessed at: http://transition.fcc.gov/Daily_Releases/Daily_Business/2017/db0718/FCC-17-94A1.pdf.

Federal Communications Commission (2017b) FCC. FCC Unlocks New Airwaves for Vehicular Radar Use. July 13, 2017. Accessed at: https://www.fcc.gov/document/fcc-unlocks-new-airwaves-vehicular-radar-use.

Federal Motor Carrier Safety Administration (2017). FMCSA to Hold Public Listening Session on Highly Automated Commercial Vehicles. April 13, 2017. Accessed at: https://www.fmcsa.dot.gov/newsroom/fmcsa-hold-public-listening-session-highly-automated-commercial-vehicles

Federal Trade Commission (2017a). Connected Cars: Privacy Security Issues Related to Connected Automated Vehicles. June 27, 2017. Accessed at: https://www.ftc.gov/news-events/events-calendar/2017/06/connected-cars-privacy-security-issues-related-connected

Federal Trade Commission (2017b) FTC and NHTSA Seek Input on Benefits and Privacy and Security Issues Associated with Current and Future Motor Vehicles. March 20, 2017. Accessed At: https://www.ftc.gov/system/files/attachments/press-releases/ftc-nhtsa-conduct-workshop-june-28-privacy-security-issues-related-connected-automated-vehicles/notice_connected_cars_workshop_with_nhtsa_1.pdf

Karwan Ghazi Fendia, Sarhat Mustafa Adamb, Nick Kokkasb, and Martin Smith (2014). An Approach to Produce a GIS Database for Road Surface Monitoring. ICCEN 2013, APCBEE Procedia 9, pp. 235-240.

Ferrara, A., Sacone, S., & Siri, S. (2015). Event-triggered model predictive schemes for freeway traffic control. *Transportation Research Part C: Emerging Technologies*, 58, 554-567.

FHWA. Document on Traffic Volume Trends. U.S. Department of Transportation, (2017). Retrieved at: https://www.fhwa.dot.gov/policyinformation/travel_monitoring/17maytvt/17maytvt.pdf (July 15, 2017).

Fleetwood, Janet. (2017). Public Health, Ethics, and Autonomous Vehicles. *American Journal of Public Health* 107(4): 532-537.

Florida Department of Transportation, (2013). 2009 National Household Travel Survey – Florida Data Analysis. Retrieved from: http://www.fdot.gov/planning/trends/special/nhts.pdf (June 20, 2017).

Flötteröd, G., & Rohde, J. (2011). Operational macroscopic modeling of complex urban road intersections. *Transportation Research Part B: Methodological*, 45(6), 903-922.

Folsom, T. (2012). Energy and autonomous urban land vehicles. *IEEE Technology and Society Magazine* 2: 28-38.

Ford. (2012). Model T Facts. Online at: https://media.ford.com/content/fordmedia/fna/us/en/news/2013/08/05/model-t-facts.html

Friesz, T. L., D. Bernstein, and N. Kydes (2004). Dynamic congestion pricing in disequilibrium. *Networks and Spatial Economics* 4(2), 181–202.

Y. Fuke and E. Krotkov (1996). "Dead reckoning for a lunar rover on uneven terrain", Robotics and Automation, 1996 IEEE International Conference, vol. 1, IEEE, 1996, pp. 411-416.

Furda, A. and L. Vlacic (2011). Enabling safe autonomous driving in real-world city traffic using multiple criteria decision making. IEEE Intelligent Transportation Systems Magazine 3(1), 4–17.

Gagnier, S. 2013. Car sharing users to reach 12 million by 2020, report says. Online at: http://www.autonews.com/article/20130916/OEM06/130919868/car-sharing-users-to-reach-12-million-by-2020-report-says.

Gajda, J., R. Sroka, M. Stencel, A. Wajda, and T. Zeglen (2001). A vehicle classification based on inductive loop detectors. In Instrumentation and Measurement Technology Conference, 2001. IMTC 2001. Proceedings of the 18th IEEE, Volume 1, pp. 460–464. IEEE.

Gao, S., (2005). Optimal adaptive routing and traffic assignment in stochastic time-dependent networks. Doctoral dissertation, Massachusetts Institute of Technology.

Gardner, L., H. Bar-Gera, and S. D. Boyles (2013). Development and comparison of choice models and tolling schemes for high-occupancy/toll (HOT) facilities. *Transportation Research Part B* 55, 142–153.

Gardner, L., S. D. Boyles, H. Bar-Gera, and K. Tang (2015). Robust tolling schemes for high-occupancy/toll (hot) facilities under variable demand. *Transportation Research Record* 2450, 152–162.

Gardner, L., S. D. Boyles, and S. T. Waller (2011). Quantifying the benefit of responsive pricing and travel information in the stochastic congestion pricing problem. *Transportation Research Part A* 45, 204–218.

Gardner, L., J. Duthie, A. Unnikrishnan, and S. T. Waller (2008). Robust pricing for networks with demand uncertainty. Presented at the 87th Annual Meeting of the Transportation Research Board, Washington, DC.

Gardner, L. M., A. Unnikrishnan, and S. T. Waller (2010). Solution methods for robust pricing of transportation networks under uncertain demand. Transportation Research Part C: Emerging Technologies 18(5), 656 – 667. Applications of Advanced Technologies in Transportation: Selected papers from the 10th {AATT} Conference.

Garavello, M., & Piccoli, B. (2006). Traffic flow on networks (Vol. 1). Springfield: American institute of mathematical sciences.

Gentile, G. (2010). The general link transmission model for dynamic network loading and a comparison with the DUE algorithm. In: Tampère, C.M.J., Viti, F., Immers, L.H. (Eds.),

New Developments in Transport Planning: Advances in Dynamic Traffic Assignment. Edward Elgar, pp. 153–178 .

Global Market Insights, Inc. (2017). Car Sharing Market Size, 2016 – 2024. Online at: https://www.gminsights.com/industry-analysis/carsharing-market?utm_source=globenewswire.com&utm_medium=referral&utm_campaign=Paid_Globnewswire

Godunov, S. K. (1959). A difference method for numerical calculation of discontinuous solutions of the equations of hydrodynamics. *Matematicheskii Sbornik*, 89(3), 271-306.

Goodall, Noah J. (2017). From Trolleys to Risk: Models for Ethical Autonomous Driving. *American Journal of Public Health* 107(4): 496.

Greene, D. L. (2008). Assessment of Fuel Economy Technologies for Light-Duty Vehicles. *Transportation Research Record* 2058. Online at: https://www.nap.edu/catalog/12924/assessment-of-fuel-economy-technologies-for-light-duty-vehicles

Grether, D., Y. Chen, M. Rieser, U. Beuck, and K. Nagel (2008). Emergent effects in multi-agent simulations of road pricing. In Proceedings of the Annual Meeting of the European Regional Science Association (ERSA), pp. 08–08.

Gu, Z., Liu, Z., Cheng, Q., & Saberi, M. (2018). Congestion Pricing Practices and Public Acceptance: A Review of Evidence. *Case Studies on Transport Policy*.

Gucwa, M. (2014). Mobility and energy impacts of automated cars. Proceedings of the Automated Vehicles Symposium, San Francisco.

Guo, R.-Y., H. Yang, H.-J. Huang, and Z. Tan (2015). Link-based day-to-day network traffic dynamics and equilibria. *Transportation Research Part B: Methodological* 71, 248–260.

Guo, R.-Y., H. Yang, H.-J. Huang, and Z. Tan (2016). Day-to-day flow dynamics and congestion control. *Transportation Science* 50(3), 982–997.

F. Gustafsson, F. Gunnarsson, N. Bergman, U. Forssell, J. Jansson, R. Karlsson, and P. J. Nordlund (2017). Particle Filter for Positioning, Navigation and Tracking. Signal Processing, IEEE transactions, pp. 1-11.

van der Gun, J. P., Pel, A. J., & Van Arem, B. (2017). Extending the Link Transmission Model with non-triangular fundamental diagrams and capacity drops. *Transportation Research Part B: Methodological*, 98, 154-178.

Haboucha, C.J., Ishaq, R., Shiftan, Y., (2015). User preferences regarding autonomous vehicles: giving up your private car. Presented at the International Association for Travel Behaviour Research Meeting in Windsor, U.K.

Han, D. and H. Yang (2009). Congestion pricing in the absence of demand functions. *Transportation Research Part E: Logistics and Transportation Review* 45(1), 159 – 171.

Hasch, J., E. Topak, R. Schnabel, T. Zwick, R. Weigel, and C. Waldschmidt (2012). Millimeter-wave technology for automotive radar sensors in the 77 ghz frequency band. IEEE *Transactions on Microwave Theory and Techniques* 60(3), 845–860.

Haurie, A. and P. Marcotte (1985). On the relationship between Nash-Cournot and Wardrop equilibria. *Networks* 15(3), 295–308.

Hawkins, (2018). Waymo's self-driving trucks will start delivering freight in Atlanta. URL: https://www.theverge.com/2018/3/9/17100518/waymo-self-driving-truck-google-atlanta

Hawkings, A.J. (2017). Uber's self-driving cars are now picking up passengers in Arizona. The Verge (Feb. 21). Retrieved at: https://www.theverge.com/2017/2/21/14687346/uber-self-driving-car-arizona-pilot-ducey-california

L. Heng, D. B. Work, and G. X. Gao (2015). "GPS signal authentication from cooperative peers", IEEE Transactions on Intelligent Transportation Systems, vol. 16, no. 4, pp. 1794-1805.

Himpe, W., Corthout, R., & Tampère, M. C. (2016). An efficient iterative link transmission model. *Transportation Research Part B: Methodological*, 92, 170-190.

Hopf, E. (1969). On the right weak solution of the Cauchy problem for a quasilinear equation of first order. *Journal of Mathematics and Mechanics*, 19(6), 483-487.

Hoogendoorn, R., van Arem, B., & Hoogendoorn, S. (2014). Automated driving, traffic flow efficiency and human factors: A literature review. *Transportation Research Record*, 2422, 113–120

Hörl, S., Ruch, C., Becker, F., Frazzoli, E. & Axhausen, K. (2017). Fleet control algorithms for automated mobility: A simulation assessment for Zurich. ETH Zürich Research Collection. Retrieved at: https://www.research-collection.ethz.ch/handle/20.500.11850/175260

Hörl, Sebastian. (2017). Agent-based simulation of autonomous taxi services with dynamic demand responses. Procedia Computer Science 109C: 899-904.

Horni, A., Nagel, K., & Axhausen, K. W. (Eds.). (2016). The multi-agent transport simulation MATSim. London: Ubiquity Press.

Huang, H.-J. (2002). Pricing and logit-based mode choice models of a transit and highway system with elastic demand. *European Journal of Operational Research* 140(3), 562–570.

Immorlica, N., L. E. Li, V. S. Mirrokni, and A. S. Schulz (2009). Coordination mechanisms for selfish scheduling. *Theoretical Computer Science* 410(17), 1589–1598.

Jabari, S. E. (2016). Node modeling for congested urban road networks. *Transportation Research Part B: Methodological*, 91, 229-249.

Jenkins, Ryan. (2016). Autonomous Vehicle Ethics & Law – Toward an Overlapping Consensus. A New America Foundation Report. Retrieved from: https://na-production.s3.amazonaws.com/documents/AV-Ethics-Law.pdf (July 10, 2017).

A. Jimenez, F. Seco, C. Prieto, and J. Guevara (2009). "A comparison of pedestrian Dead-reckoning algorithms using a low-cost MEMS IMU", Intelligent Signal Processing, WISP 2009, IEEE International Symposium, August 2009, pp. 37-42.

Joksimovic, D., M. Bliemer, and P. Bovy (2005a). Optimal toll design problem in dynamic traffic networks with joint route and departure time choice. *Transportation Research Record* (1923), 61–72.

Joksimovic, D., M. C. Bliemer, P. H. Bovy, and Z. Verwater-Lukszo (2005b). Dynamic road pricing for optimizing network performance with heterogeneous users. In Proceedings. 2005 IEEE Networking, Sensing and Control, 2005., pp. 407–412. IEEE.

Kaddoura, I., Bischoff, J. & Nagel, K. (2018). Towards welfare optimal operation of innovative mobility concepts: External cost pricing in a world of shared autonomous vehicles. VSP Working paper, 18-01. Retrieved at http://www.vsp.tu-berlin.de/publications/vspwp/

Kaddoura, I., Kickhöfer, B., Neumann, A., & Tirachini, A. (2015). Optimal public transport pricing: Towards an agent-based marginal social cost approach. *Journal of Transport Economics and Policy* (JTEP), 49(2), 200-218.

Kang, C. (September, 2016). "No Driver? Bring It On. How Pittsburgh Became Uber's Testing Ground" on The New York Times. Retrieved at: https://www.nytimes.com/2016/09/11/technology/no-driver-bring-it-on-how-pittsburgh-became-ubers-testing-ground.html

Kang, C. (September, 2016). "Self-Driving Cars Gain Powerful Ally: The Government" on The New York Times. Retrieved at: https://www.nytimes.com/2016/09/20/technology/self-driving-cars-guidelines.html

Karmarkar, N. (1984). A new polynomial-time algorithm for linear programming. In Proceedings of the Sixteenth Annual ACM Symposium on Theory of Computing, STOC '84, New York, NY, USA, pp. 302– 311. ACM.

Kelly Blue Book. 2016. Future Autonomous Vehicle Driver Study – September 2016. Retrieved from: https://mediaroom.kbb.com/download/Kelley+Blue+Book+Future+Autonomous+Vehicle+Driver+Study+-+FINAL.pdf (July 10, 2017).

Y. K. Ki and D. K. Baik (2006). Vehicle claasification algorithm for single-loop detectors using neural networks. Vehicular Technology, IEEE Transctions, vol.55, no. 6, pp. 1704-1711.

Kickhöfer, B., Grether, D., & Nagel, K. (2011). Income-contingent user preferences in policy evaluation: application and discussion based on multi-agent transport simulations. *Transportation*, 38(6), 849.

Tracy Kijewski-Correa, Martin Haenggi, Panos Antsaklis. Wireless Sensor Networks for Structural Health Monitoring: A Multi-Scale Approach. 17th Analysis and Computation Specialty Conference, St. Louis MO, May, 2006.

Kim, K., Rousseau, G., Freedman, J., & Nicholson, J. (2015, May). The travel impact of autonomous vehicles in metro Atlanta through activity-based modeling. In 15th TRB National Transportation Planning Applications Conference, Atlantic City, May (Vol. 18).

Kockelman, K. M., and T. Li. (2016). Valuing the Safety Benefits of Connected and Automated Vehicle Technologies. In: Transportation Research Board 95th Annual Meeting. Online at: https://www.caee.utexas.edu/prof/kockelman/public_html/TRB16CAVSafety.pdf

Kockelman, K., Sharon, G., Simoni, M., Albert, M., Fritz, H., Hutchinson, R., ... & Pourrahmani, E. (2017). An assessment of autonomous vehicles: traffic impacts and infrastructure needs (No. FHWA/TX-17/0-6847-1). University of Texas at Austin. Center for Transportation Research.

Kohl, N. and P. Stone (2004). Policy gradient reinforcement learning for fast quadrupedal locomotion. In Proceedings of the 2004 IEEE International Conference on Robotics and Automation, ICRA 2004, April 26 - May 1, 2004, New Orleans, LA, USA, pp. 2619–2624.

Korilis, Y. A., A. A. Lazar, and A. Orda (1997). Achieving network optima using stackelberg routing strategies. IEEE/ACM Transactions on Networking (TON) 5(1), 161–173.

Krajzewicz, D., G. Hertkorn, C. Rossel, and P. Wagner (2002). Sumo (simulation of urban mobility)-an opensource traffic simulation. In Proceedings of the 4th middle East Symposium on Simulation and Modelling (MESM20002), pp. 183–187.

Krueger, Rico, Rashidi, Taha, and Rose, John. (2016). Preferences for shared autonomous vehicles. *Transportation Research Part C* (69): 343-355.

LaMondia, J., D. Fagnant, H. Qu, J. Barrett, and K. Kockelman. 2016. Long-Distance Travel Mode Shifts due to Automated Vehicles. *Transportation Research Record*: 2566: 1-10.

Lamotte, R., De Palma, A., & Geroliminis, N. (2016). Sharing the road: the economics of autonomous vehicles.

Lawphongpanich, S., & Yin, Y. (2010). Solving the Pareto-improving toll problem via manifold suboptimization. *Transportation Research Part C: Emerging Technologies*, 18(2), 234-246.

Lax, P. D. (1957). Hyperbolic systems of conservation laws II. *Communications on Pure and Applied Mathematics*, 10(4), 537-566.

Lax, P.D., 1973. Hyperbolic systems of conservation laws and the mathematical theory of shock waves, SIAM.

LeBlanc, L. J., E. K. Morlok, and W. P. Pierskalla (1975). An efficient approach to solving the road network equilibrium traffic assignment problem. *Transportation Research* 9(5), 309–318.

Leclercq, L., Laval, J. A., & Chevallier, E. (2007). The Lagrangian coordinates and what it means for first order traffic flow models. *Transportation and Traffic Theory* 2007. Papers Selected for Presentation at ISTTT17.

Lee C., Ward C., Raue M., D'Ambrosio L., Coughlin J. (2017). Age Differences in Acceptance of Self-Driving Cars: A Survey of Perceptions and Attitudes. Proceedings in the 3rd International Conference on Human Aspects of IT for the Aged Population, Vancouver, Canada. Retrieved from: https://link.springer.com/chapter/10.1007/978-3-319-58530-7_1 (July 20, 2017).

Lee, J. and Park, B., (2012). Development and evaluation of a cooperative vehicle intersection control algorithm under the connected vehicles environment. IEEE Transactions on Intelligent Transportation Systems, 13(1), pp.81-90.

Lenz, B., Fraedrich, E. M., & Cyganski, R. (2016). Riding in an autonomous car: What about the user perspective?.

Levin, M. W. and S. D. Boyles (2015). Intersection auctions and reservation-based control in dynamic traffic assignment. *Transportation Research Record* (2497), 35–44.

Levin, M. W., M. Pool, T. Owens, N. R. Juri, and S. T. Waller (2015). Improving the convergence of simulation-based dynamic traffic assignment methodologies. *Networks and Spatial Economics* 15(3), 655– 676.

Levin, M. W., & Boyles, S. D. (2016). A multiclass cell transmission model for shared human and autonomous vehicle roads. *Transportation Research Part C: Emerging Technologies*, 62, 103-116.

Levin, Michael., Li, Tianxin, Boyles, Stephen, and Kockelman, Kara. (2017). A General Framework for Modeling Shared Autonomous Vehicles with Dynamic Network-Loading and Dynamic Ride-Sharing Application. *Computers, Environment and Urban Systems* 64: 373-383.

Levine, S. and P. Abbeel (2014). Learning neural network policies with guided policy search under unknown dynamics. *Advances in Neural Information Processing Systems*, pp. 1071–1079.

Y. Li, E. Canepa, and C. Claudel, (2014). "Optimal control of scalar conservation laws using linear/quadraticprogramming: Application to transportation networks, IEEE Transactions on Control of Network Systems, vol. 1,no. 1, pp. 28–39.

Lighthill, M. J., & Whitham, G. B. (1955, May). On kinematic waves. II. A theory of traffic flow on long crowded roads. In Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences (Vol. 229, No. 1178, pp. 317-345). The Royal Society.

Lin, D., A. Unnikrishnan, and S. T. Waller (2011). A dual variable approximation based heuristic for dynamic congestion pricing. *Networks and Spatial Economics* 11(2), 271–293.

Litman, T. (1999). Distance-based charges; a practical strategy for more optimal vehicle pricing. Victoria Transport Policy Institute.

Litman, T. (2016). Smart congestion relief: Comprehensive analysis of traffic congestion costs and congestion reduction benefits.

Litman, T. (2017). Autonomous vehicle implementation predictions. Victoria Transport Policy Institute, 28.

Litman, T., & Doherty, E. (2011). Transportation Cost and Benefit Analysis II–Parking Costs. Transportation Cost and Benefit Analysis Techniques, *Estimates and Implications*, 5-4.

Liu, H. X., X. He, and B. He (2009). Method of successive weighted averages (mswa) and self-regulated averaging schemes for solving stochastic user equilibrium problem. *Networks and Spatial Economics* 9(4), 485–503.

Liu, Y. and G.-L. Chang (2011). An arterial signal optimization model for intersections experiencing queue spillback and lane blockage. *Transportation Research Part C: Emerging Technologies* 19(1), 130–144.

Liu, J., A. Khattak, and X. Wang. (2015). The role of alternative fuel vehicles: Using behavioral and sensor data to model hierarchies in travel. *Transportation Research Part C: Emerging Technologies* 55: 379-392.

Liu, J., Kockelman, K. M., Boesch, P. M., & Ciari, F. (2017). Tracking a system of shared autonomous vehicles across the Austin, Texas network using agent-based simulation. *Transportation*, 1-18.

Liu, Z., Wang, S., & Meng, Q. (2014). Optimal joint distance and time toll for cordon-based congestion pricing. *Transportation Research Part B: Methodological*, 69, 81-97.

Lo, H. K. (1999a). A dynamic traffic assignment formulation that encapsulates the cell-transmission model. In 14th international symposium on transportation and traffic theory.

Lo, H. K. (1999b). A novel traffic signal control formulation. *Transportation Research Part A: Policy and Practice*, 33(6), 433-448.

Lo, H. K. and W. Y. Szeto (2005). Road pricing modeling for hyper-congestion. *Transportation Research Part A: Policy and Practice* 39(7), 705–722.

R. Lopes Pereira, J. Trindade, F. Gonçalves, L. Suresh, D. Barbosa, and T. Vazão (2014). A wireless sensor network for monitoring volcano-seismic signals. Natural Hazards and Earth System Sciences, 14, 2014, pp. 3123–3142.

Lou, Y., Y. Yin, and J. A. Laval (2011). Optimal dynamic pricing strategies for high-occupancy/toll lanes. *Transportation Research Part C: Emerging Technologies* 19(1), 64–74.

Loeb, Benjamin, Kockelman, Kara M. (2017). Fleet Performance & Cost Evaluation of a Shared Autonomous Electric Vehicle (SAEV) Fleet: A Case Study for Austin, Texas. Accepted for publication in *Transportation Research Part A – Policy and Practice*.

Loeb, Benjamin, Kockelman, Kara M. and Liu, Jun. (2017). Shared Autonomous Electric Vehicle (SAEV) Operations Across the Austin, Texas Network with a Focus on Charging Infrastructure Decisions. Proceedings of the 96th Annual Meeting of the Transportation Research Board, Washington, D.C. and under review for publication in *Transportation Research Part C: Emerging Technologies*.

Long, J., Gao, Z., Zhao, X., Lian, A., & Orenstein, P. (2011). Urban traffic jam simulation based on the cell transmission model. *Networks and Spatial Economics*, 11(1), 43-64.

Lu, N., Cheng, N., Zhang, N., Shen, X. and Mark, J.W., (2014). Connected vehicles: Solutions and challenges. Connected vehicles: Solutions and challenges. IEEE internet of things journal, 1(4), pp.289-299.

Lukasiewicza, P., K. Karpioa, and A. Orl owskia (2012). The models of personal incomes in USA. In Proceedings of the 5th Symposium on Physics in Economics and Social Sciences, Warsaw, Poland.

M.J. Beckman. (1965). On optimal tolls for highways, tunnels and bridges. *Vehicular Traffic Science*, American Elsevier, New York, pp. 331-341

Ma, S., Zheng, Y. and Wolfson, O. (2013). T-share: A large-scale dynamic taxi ridesharing service. Proceedings of the 2013 IEEE 29th International Conference on Data Engineering (ICDE 2013). Brisbane, Australia.

Ma, Shuo and Wolfson, Ouri. (2013). Analysis and evaluation of the slugging form of ridesharing. Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Orlando, Florida. Retrieved from: http://dl.acm.org/citation.cfm?id=2525365 (July 20, 2017).

Maciejewski, M., Bischoff, J., & Nagel, K. (2016). An assignment-based approach to efficient real-time city-scale taxi dispatching. IEEE Intelligent Systems, 31(1), 68-77.

Maciejewski, M., Bischoff, J., Hörl, S., & Nagel K. (2017). Towards a Testbed for Dynamic Vehicle Routing Algorithms. Cham: Springer International Publishing, pp. 69–79.

Martin, E., and S. Shaheen. (2016). Impacts of car2go on Vehicle Ownership, Modal Shift, Vehicle Miles Traveled, and Greenhouse Gas Emissions: An Analysis of Five North American Cities. Transportation Sustainability Research Center (TSRC), UC Berkeley. Online at: http://innovativemobility.org/wp-content/uploads/2016/07/Impactsofcar2go_FiveCities_2016.pdf

May, A., Shepherd, S., Sumalee, A., & Koh, A. (2008). Design tools for road pricing cordons. Road Congestion Pricing in Europe: Implications for the United States, Cheltenham, UK, 138-155.

Mazaré, P. E., Dehwah, A. H., Claudel, C. G., & Bayen, A. M. (2011). Analytical and grid-free solutions to the Lighthill–Whitham–Richards traffic flow model. *Transportation Research Part B: Methodological*, 45(10), 1727-1748.

P.-E. Mazare, O.-P. Tossavainen, and D. B. Work, Computing travel times from filtered traffic states. Discrete & Continuous Dynamical Systems-Series S, vol. 7, no. 3, 2014.

Merchant, D. K. and G. L. Nemhauser (1978). A model and an algorithm for the dynamic traffic assignment problems. *Transportation Science* 12(3), 183–199.

Milakis, D., Van Arem, B., & Van Wee, B. (2017). Policy and society related implications of automated driving: A review of literature and directions for future research. *Journal of Intelligent Transportation Systems*, 1-25.

Miller-Hooks, E.D. and Mahmassani, H.S., (2000). Least expected time paths in stochastic, time-varying transportation networks. *Transportation Science*, 34(2), pp.198-215.

M. Mousa, M. Abdulaal, S. Boyles, and C. Claudel. "Wireless sensor network-based urban traffic monitoring using Inertial reference data", Distributed Computing in Sensor Systems (DCOSS), 215 International Conference on IEEE, pp. 206-207, 2015.

Mustafa Mousa, Kapil Sharma, Christian G. Claudel. Inertial Measurement Units-Based Probe Vehicles: Automatic Calibration, Trajectory Estimation, and Context Detection (2002). IEEE Transactions on Intelligent Transportation Systems, vol. 50, no. 2, pp. 425-437.

Muoio, D. (September, 2017). "RANKED: The 18 companies most likely to get self-driving cars on the road first" on Business Insider. Retrieved at: http://www.businessinsider.com/the-companies-most-likely-to-get-driverless-cars-on-the-road-first-2017-4/#18-baidu-1

Muñoz, L., Sun, X., Horowitz, R., & Alvarez, L. (2003). Traffic density estimation with the cell transmission model. In American Control Conference, 2003. Proceedings of the 2003 (Vol. 5, pp. 3750-3755). IEEE.

Nagel, K., & Flötteröd G. (2009). Agent-based traffic assignment: going from trips to behavioral travelers. 12th International Conference on Travel Behaviour Research (IATBR), Jaipur

National Archives and Records Administration: Federal Register (2017). Federal Motor Carrier Safety Regulations: Highly Automated Commercial Vehicles; Public Listening Session. April 17, 2017. Accessed at: https://www.federalregister.gov/documents/2017/04/17/2017-07723/federal-motor-carrier-safety-regulations-highly-automated-commercial-vehicles-public-listening

National Conference of State Legislators (NCSL). Website, not dated. Accessed at: http://www.ncsl.org/research/transportation/autonomous-vehicles.aspx.

National Highway Traffic Safety Administration (NHTSA) (2016a). Federal Automated Vehicles Policy: Accelerating the Next Revolution in Roadway Safety. September 2016. Accessed at: https://one.nhtsa.gov/nhtsa/av/pdf/Federal_Automated_Vehicles_Policy.pdf

National Highway Traffic Safety Administration (2016b). Cybersecurity Best Practices for Modern Vehicles. October 2016. Accessed at: https://www.nhtsa.gov/staticfiles/nvs/pdf/812333_CybersecurityForModernVehicles.pdf

National Highway Traffic Safety Administration (2017). Automated Driving Systems: A Vision for Safety. September 12, 2017. Accessed at: https://www.nhtsa.gov/sites/nhtsa.dot.gov/files/documents/13069a-ads2.0_090617_v9a_tag.pdf

NCHRP. Travel Demand Forecasting: Parameters and Techniques. VOL. 716. Transportation Research Board, (2012). Online at: https://www.nap.edu/catalog/14665/travel-demand-forecasting-parameters-and-techniques

Newell, G. F. (1993). A simplified theory of kinematic waves in highway traffic, part I: General theory. *Transportation Research Part B: Methodological*, 27(4), 281-287.

Newell, G. F. (1993). A simplified theory of kinematic waves in highway traffic, Part II: Queueing at freeway bottlenecks. *Transportation Research Part B: Methodological*, 27(4), 289-303.

NHTSA, (2013). Preliminary Statement of Policy Concerning Automated Vehicles Preliminary Statement of Policy Concerning Automated Vehicles.

NHTSA (2015). Document on Critical Reasons for Crashes Investigated in the National Motor Vehicle Crash Causation Survey. Publication DOT-HS-812-115, U.S. Department of Transportation.

L. Ojeda and J. Borenstein (2007). "Personal dead0rechoning system for GPS-denied environments." Safety, Security and Rescue Robotics, pp. 1-6.

Outwater, M., Bradley, M., Ferdous, N., Trevino, S. and Lin, H., (2015). Foundational Knowledge to Support a Long-Distance Passenger Travel Demand Modeling Framework: Implementation Report. FHWA, DTFH61-10-R-00036.

Papageorgiou, M., H. Hadj-Salem, and J.-M. Blosseville (1991). Alinea: A local feedback control law for on-ramp metering. *Transportation Research Record* (1320), 58–64.

Patel, R., M. W. Levin, and S. D. Boyles. (2016) Effects of autonomous vehicle behavior on arterial and freeway networks. *Transportation Research Record* 2561, 9-17.

Patriksson, M. (1994). The Traffic Assignment Problem — Models and Methods. Utrecht, Netherlands: VSP.

Paul, B., K. Kockelman, and S. Musti. (2011). Evolution of the light-duty vehicle fleet: anticipating adoption of plug-In hybrid electric vehicles and greenhouse gas emissions across the US fleet. *Transportation Research Record* 2252: 107-117.

Peeta, S., & Ziliaskopoulos, A. K. (2001). Foundations of dynamic traffic assignment: The past, the present and the future. *Networks and Spatial Economics*, 1(3), 233-265.

Peters, J. and S. Schaal (2006). Policy gradient methods for robotics. In 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2006, October 9-15, 2006, Beijing, China, pp. 2219– 2225.

Pigou, A. C. (1920a). The Economics of Welfare. London: Macmillan and Co.

Pigou, A. C. (1920b). The Economics of Welfare. Palgrave Macmillan.

Pigou, A.C. (1924). The economics of welfare. Transaction Publishers

Polychronopoulos, G.H. and Tsitsiklis, J.N., (1993). Stochastic shortest path problems with recourse.

PUMS (Public Use Microdata Sample) (2011-2015) United State Census Bureau: American Community Survey. Retrieved from: https://www.census.gov/programs-surveys/acs/data/pums.html (June 10, 2017).

Quarles, N., and Kockelman, K. (2017). Americans' Plans for Acquiring and Using Electric, Shared and Self-Driving Vehicles. Under review for publication in Transportation Research Record, and available at www.caee.utexas.edu/prof/kockelman/public_html/TRB18surveyEVAV.pdf

Raadsen, M. P., Bliemer, M. C., & Bell, M. G. (2016). An efficient and exact event-based algorithm for solving simplified first order dynamic network loading problems in continuous time. *Transportation Research Part B: Methodological*, 92, 191-210.

Rambha, T. and S. D. Boyles (2016). Dynamic pricing in discrete time stochastic day-to-day route choice models. *Transportation Research Part B: Methodological* 92, Part A, 104 – 118. Special issue Day-to-Day Dynamics in Transportation Networks.

Rambha, T., Boyles, S.D. and Waller, S.T., (2016). Adaptive transit routing in stochastic time-dependent networks. *Transportation Science*, 50(3), pp.1043-1059.

Rambha, T., S. D. Boyles, A. Unnikrishnan, and P. Stone (2017). Marginal cost pricing for system optimal traffic assignment with recourse under supply-side uncertainty. In Review in *Transportation Research Part B: Methodological*.

Rau, P., M. Yanagisawa, and W. G. Najm. (2015). Target Crash Population of Automated Vehicles. In: 24th International Technical Conference on the Enhanced Safety of

Vehicles (ESV). Online at: https://www-esv.nhtsa.dot.gov/proceedings/24/files/Session%2021%20Written.pdf

Regele, R. (2008). Using ontology-based traffic models for more efficient decision making of autonomous vehicles. In Fourth International Conference on Autonomic and Autonomous Systems (ICAS'08), pp. 94–99. IEEE.

Reiter, M. S., and K. M. Kockelman. (2017). Emissions and Exposure Costs of Electric Versus Conventional Vehicles: A Case Study for Texas. *International Journal of Sustainable Transportation* 11.7: 486-492.

Richards, P. I. (1956). Shock waves on the highway. Operations research, 4(1), 42-51.

Rodier, C., Alemi, F. and Smith, D. (2016). Dynamic Ridesharing: An Exploration of the Potential for Reduction in Vehicle Miles Traveled. Proceedings of the 95th Annual Meeting of the Transportation Research Board, Washington, D.C.

Rosenthal, R. W. (1973). A class of games possessing pure-strategy Nash equilibria. *International Journal of Game Theory* 2(1), 65–67.

Roughgarden, T. (2004). Stackelberg scheduling strategies. *SIAM Journal on Computing* 33(2), 332–350.

Roughgarden, T. and E. Tardos (2002). How bad is selfish routing? *Journal of the ACM* (JACM) 49(2), 236–259.

Russell, J. (March, 2017). "China's Didi Chuxing opens U.S. lab to develop AI and self-driving car tech" on Tech Crunch. Retrieved at: https://techcrunch.com/2017/03/08/didi-us-research-lab/

San Francisco Municipal Transportation Agency. 2015. Travel Decisions Survey 2015 – Summary Report, San Francisco. Retrieved from: https://www.sfmta.com/sites/default/files/reports/2016/Travel%20Decision%20Survey%202015%20Report_Accessible.pdf (June 30, 2017).

Santi, Paolo, Resta, Giovanni, Szell, Michael, Sobolevsky, Stanislav, Strogatz, Steven and Ratti, Carlo. (2014). Quantifying the benefits of vehicle pooling with shareability networks. Proceedings of the National Academy of Sciences of the United States of America 111(37): 13290-13294.

Schaller, B. (2010). New York Citys congestion pricing experience and implications for road pricing acceptance in the United States. *Transport Policy* 17(4), 266–273.

Schakel, W. J., Van Arem, B., & Netten, B. D. (2010, September). Effects of cooperative adaptive cruise control on traffic flow stability. In Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on (pp. 759-764). IEEE.

Scheltes, A., & de Almeida Correia, G. H. (2017). Exploring the use of automated vehicles as last mile connection of train trips through an agent-based simulation model: An application to Delft, Netherlands. International Journal of Transportation Science and Technology, 6(1), 28-41.

Schoettle, B. and Sivak, M. (2014). A Survey of Public Opinion About Autonomous and Self-Driving Vehicles in the U.S., the U.K., and Australia. University of Michigan, Technical

Report No. UMTRI-2014-21. Retrieved from: https://deepblue.lib.umich.edu/handle/2027.42/108384 (June 30, 2017).

Schoettle, B. and Sivak, M. (2016). Motorists Preferences for Different Levels of Vehicle Automation – 2016. University of Michigan, Technical Report No. SWT-2016-8. Retrieved from: http://umich.edu/~umtriswt/PDF/SWT-2016-8.pdf (July 20, 2017).

Schoettle, B., and M. Sivak. (2015). A Preliminary Analysis of Real-World Crashes Involving Self-Driving Vehicles. Online at: http://umich.edu/~umtriswt/PDF/UMTRI-2015-34.pdf

Schreiter, T., Smits, I. E. S., van Lint, H., & Hoogendoorn, S. (2010, November). The cell transmission model with capacity drop. In Proceedings of the 11th International Congress of the Research School Transportation, Infrastructure and Logistics (TRAIL).

Shaheen, S., Coehn, A., Zohdy, I., and Kock B. (2016a) Smartphone Applications to Influence Travel Choices: Practices and Policies. Publication FHWA-HOP-16-023. FHWA, U.S. Department of Transportation, 2016a. Retrieved at: https://ops.fhwa.dot.gov/publications/fhwahop16023/fhwahop16023.pdf (January 24, 2018).

Shaheen, S., Cohen, A., and Zohdy, I. (2016b) Shared Mobility: Current Practices and Guiding Principles. Publication FHWA-HOP-16-022. FHWA, U.S. Department of Transportation, 2016b. Retrieved at: https://ops.fhwa.dot.gov/publications/fhwahop16022/fhwahop16022.pdf (January 24, 2018).

Sharon, G., M. Albert, T. Rambha, S. Boyles, and P. Stone (2018, February). Traffic optimization for a mixture of self-interested and compliant agents. In Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI-18).

Sharon, G., J. P. Hanna, M. W. Levin, T. Rambha, M. Albert, S. D. Boyles, and P. Stone (2017a, May). Real-time adaptive tolling scheme for optimized social welfare in traffic networks. In Proceedings of the 16th International Conference on Autonomous Agents and Multiagent Systems (AAMAS-17).

Sharon, G., Levin, M. W., Hanna, J. P., Rambha, T., Boyles, S. D., & Stone, P. (2017b). Network-wide adaptive tolling for connected and automated vehicles. *Transportation Research Part C: Emerging Technologies*, 84, 142-157.

Shladover, S. E., Su, D., & Lu, X.-Y. (2012). Impacts of cooperative adaptive cruise control on freeway traffic flow. *Transportation Research Record*, 2324, 63–70.

Shoup, D. C. (2006). Cruising for parking. *Transport Policy*, 13(6), 479-486.

Simoni, M., A. Pel, R. Waraich, and S. Hoogendoorn (2015). Marginal cost congestion pricing based on the network fundamental diagram. *Transportation Research Part C: Emerging Technologies* 56, 221 – 238.

Small, K. A., Verhoef, E. T., & Lindsey, R. (2007). The economics of urban transportation. Routledge.

Smith, M. J. (1984). The existence of a time-dependent equilibrium distribution of arrivals at a single bottleneck. *Transportation Science* 18(4), 385–394.

Sommer, K. (2013). Continental Mobility Study 2013. Continental AG. Retrieved from: https://www.continental-corporation.com/resource/blob/7380/6cddc571cd3d3b5cacd279fe0d1a00c1/mobistud-2013-dl-data.pdf (June 30, 2017).

Spieser, K., Treleaven, K., Zhang, R., Frazzoli, E., Morton, D. and Pavone, M. (2014). Toward a systematic approach to the design and evaluation of automated mobility-on-demand systems: A case study in Singapore. Road Vehicle Automation 229-245.

Srivastava, A., & Geroliminis, N. (2013). Empirical observations of capacity drop in freeway merges with ramp control and integration in a first-order model. *Transportation Research Part C: Emerging Technologies*, 30, 161-177.

StataCorp. (2015). Stata Statistical Software: Release 14. College Station, TX: StataCorp LP.

Sumalee, A., Zhong, R. X., Pan, T. L., & Szeto, W. Y. (2011). Stochastic cell transmission model (SCTM): A stochastic dynamic traffic model for traffic state surveillance and assignment. *Transportation Research Part B: Methodological*, 45(3), 507-533.

Szeto, W. (2008). Enhanced lagged cell-transmission model for dynamic traffic assignment. *Transportation Research Record*, (2085), 76-85.

Talebpour, A., & Mahmassani, H. S. (2016). Influence of connected and autonomous vehicles on traffic flow stability and throughput. *Transportation Research Part C: Emerging Technologies*, 71, 143-163.

Tampère, C. M., Corthout, R., Cattrysse, D., & Immers, L. H. (2011). A generic class of first order node models for dynamic macroscopic simulation of traffic flows. *Transportation Research Part B: Methodological*, 45(1), 289-309.

Tan, Z., H. Yang, and R.-Y. Guo (2015). Dynamic congestion pricing with day-to-day flow evolution and user heterogeneity. *Transportation Research Part C: Emerging Technologies* 61, 87 – 105.

Tao, C. C., (2007). Dynamic taxi-sharing service using intelligent transportation system technologies. Wireless Communications, Networking and Mobile Computing 3209-3212. DOI: 10.1109/WICOM.2007.795.

Tientrakool, P., Ho, Y. C., & Maxemchuk, N. F. (2011, September). Highway capacity benefits from using vehicle-to-vehicle communication and sensors for collision avoidance. In Vehicular Technology Conference (VTC Fall), 2011 IEEE (pp. 1-5). IEEE.

Tirachini, A., D. A. Hensher, & J. M. Rose (2014): 'Multimodal pricing and optimal design of urban public transport: the interplay between traffic congestion and bus crowding', *Transportation Research Part B,* 61, 33–54

Truck Report Forum, 2013. The Real Cost of Trucking – Per Mile Operating Cost of a Commercial Truck. URL: https://www.thetruckersreport.com/infographics/cost-of-trucking/

Tsekeris, T. and S. Voß (2009). Design and evaluation of road pricing: state-of-the-art and methodological advances. NETNOMICS: *Economic Research and Electronic Networking* 10(1), 5–52.

United States Congress: House of Representatives Committee on Energy and Commerce Subcommittee on Digital Commerce and Consumer Protection (2017). HR – Staff Draft on Highly Automated Vehicle Testing and Deployment. July 17, 2017. URL: http://docs.house.gov/meetings/IF/IF17/20170719/106309/HMKP-115-IF17-20170719-SD002.pdf

United States Congress: Senate Committee on Commerce, Science, & Transportation. (2017, June 13) Senators Release Bipartisan Principles for Self-Driving Vehicles Legislation. URL: https://www.commerce.senate.gov/public/index.cfm/pressreleases?ID=3319E728-1F3A-4316-8F7E-8CAF66000F8C.

United States Congress (USC). Current Legislation Search. Accessed at: https://www.congress.gov/.

United States Congress (2017a). S. 1885. November 28, 2017. (To support the development of highly automated vehicle safety technologies. Accessed at: https://www.congress.gov/bill/115th-congress/senate-bill/1885?q=%7B%22search%22%3A%5B%22automated+vehicle%22%5D%7D&r=4

United States Congress (2017b). HR 3388. September 7, 2017. Safely Ensuring Lives Future Deployment and Research In Vehicle Evolution Act or the "SELF DRIVE Act. Accessed at: https://www.congress.gov/bill/115th-congress/house-bill/3388?q=%7B%22search%22%3A%5B%22autonomous+vehicle%22%5D%7D&r=5

United States Congress (2017c). HR 3412. July 28, 2017. Let NHTSA Enforce Automated Vehicle Driving Regulations Act" or the LEAD'R Act. Accessed at: https://www.congress.gov/bill/115th-congress/house-bill/3412/text?q=%7B%22search%22%3A%5B%22autonomous+vehicle%22%5D%7D&r=2

United States Congress (2017d). HR 3401. July 28, 2017. Amend chapter 301 of Subtitle VI of Title 49 United States Code. Accessed at: https://www.congress.gov/bill/115th-congress/house-bill/3401?q=%7B%22search%22%3A%5B%22automated+vehicle%22%5D%7D&r=1

United States Congress (2017e). H.R. 3416, July 28, 2017. Establish in the National Highway Traffic Safety Administration a Rural and Mountainous Advisory Council. Accessed at: https://www.congress.gov/bill/115th-congress/house-bill/3416?q=%7B%22search%22%3A%5B%22automated+vehicle%22%5D%7D&r=3

United States Congress (2017f).H.R. 3404. To provide for the establishment in the National Highway Traffic Safety Administration of a Highly Automated Vehicle Advisory Council. July 28, 2017. Accessed at: https://www.congress.gov/bill/115th-congress/house-bill/3404?q=%7B%22search%22%3A%5B%22automated+vehicle%22%5D%7D&r=5

United States Congress (2017g).H.R. 3407. July 28, 2017. To amend chapter 301 of subtitle VI of title 49, United States Code, to require a cybersecurity plan for highly automated vehicles. Accessed at https://www.congress.gov/bill/115th-congress/house-bill/3407?q=%7B%22search%22%3A%5B%22automated+vehicle%22%5D%7D&r=6 :

United States Congress (2017h).H.R. 3430. July 28, 2017. To establish in the National Highway Traffic Safety Administration a Highly Automated Vehicle Information Sharing Advisory Council. Accessed at: https://www.congress.gov/bill/115th-congress/house-bill/3430?q=%7B%22search%22%3A%5B%22automated+vehicle%22%5D%7D&r=7

United States Congress (2017i).H.R. 3405. July 28, 2017. To amend title 49, United States Code, to expand the exemption from the motor vehicle safety standards for testing or evaluation purposes to cover manufacturers of highly automated vehicles and automated driving system components. Accessed at: https://www.congress.gov/bill/115th-congress/house-bill/3405?q=%7B%22search%22%3A%5B%22automated+vehicle%22%5D%7D&r=8

United States Congress (2017j).H.R. 3411. To establish in the National Highway Traffic Safety Administration an Automated Driving System Cybersecurity Advisory Council to make recommendations regarding cybersecurity for the testing, deployment, and updating of automated driving systems. July 28, 2017. Accessed at: https://www.congress.gov/bill/115th-congress/house-bill/3411?q=%7B%22search%22%3A%5B%22automated+vehicle%22%5D%7D&r=10

United States Congress (2017k).H.R. 3414. July 27, 2017. To establish in the National Highway Traffic Safety Administration a Disability Mobility Advisory Council to make recommendations regarding advancing mobility access for the disabled community with respect to the deployment of automated driving systems. Accessed at: https://www.congress.gov/bill/115th-congress/house-bill/3414?q=%7B%22search%22%3A%5B%22automated+vehicle%22%5D%7D&r=11

United States Congress (2017l).H.R. 3413.July 27, 217. To establish in the National Highway Traffic Safety Administration an Advisory Council on Improving Mobility Access for Underserved Populations and Senior Citizens (ACCESS Act). Accessed at: https://www.congress.gov/bill/115th-congress/house-bill/3413?q=%7B%22search%22%3A%5B%22automated+vehicle%22%5D%7D&r=13

United States Congress (2017m). H.R. 3408. July 28, 217. To amend section 30113 of title 49, United States Code to establish new exemptions for motor vehicle safety standards (AMEND Act). Accessed at: https://www.congress.gov/bill/115th-congress/house-bill/3408/text?q=%7B%22search%22%3A%5B%22automated+vehicle%22%5D%7D&r=18

Unnikrishnan, A. and Waller, S.T., (2009). User equilibrium with recourse. *Networks and Spatial Economics*, 9(4), p.575.

United States Department of Transportation (USDOT) (2009) Intelligent Transportation Systems Benefits, Costs, Deployment, and Lessons Learned. Washington, D.C: Department of Transportation, Research and Innovative Technology Administration, Intelligent Transportation Systems Joint Program Office.

US Department of Transportation, (2011). The value of travel time savings: departmental guidance for conducting economic evaluations revision 2. Retrieved at: https://www.transportation.gov/sites/dot.dev/files/docs/vot_guidance_092811c.pdf

van den Berg, V. and E. T. Verhoef (2011). Winning or losing from dynamic bottleneck congestion pricing?: The distributional effects of road pricing with heterogeneity in values of time and schedule delay. *Journal of Public Economics* 95(7), 983–992.

van den Berg, V. A., & Verhoef, E. T. (2016). Autonomous cars and dynamic bottleneck congestion: The effects on capacity, value of time and preference heterogeneity. *Transportation Research Part B: Methodological*, 94, 43-60.

VanMiddlesworth, M., K. Dresner, and P. Stone (2008). Replacing the stop sign: Unmanaged intersection control for autonomous vehicles. In Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 3, pp. 1413–1416. International Foundation for Autonomous Agents and Multiagent Systems.

Vasirani, M. and S. Ossowski (2009). A market-inspired approach to reservation-based urban road traffic management. In Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems-Volume 1, pp. 617–624. International Foundation for Autonomous Agents and Multiagent Systems.

Verhoef, E. T. (2002). Second-best congestion pricing in general networks. Heuristic algorithms for finding second-best optimal toll levels and toll points. *Transportation Research Part B: Methodological*, 36(8), 707-729.

Verhoef, E. T. (2003). Inside the queue:: hypercongestion and road pricing in a continuous time–continuous place model of traffic congestion. *Journal of Urban Economics* 54(3), 531–565.

Vickrey, W. (1997). Public economics: selected papers by William Vickrey. Cambridge University Press.

Vickrey, W. S. (1963). Pricing in urban and suburban transport. *The American Economic Review*, 53(2), 452-465. Chicago

Vittorio Astarita, Maria Vittoria Caruso, Guido Danieli, Demetrio Carmine Festa (2012). "A Mobile Application for Road Surface Quality Control." *Social and Behavioral Scieneces*, Volume 54, pp. 1135-1144.

Vujanic, A., and Unkefer, H. (2011). Embedded software consumer pulse survey. Accenture Research. Retrieved from: https://newsroom.accenture.com/content/1101/files/EmbeddedSoftwareOverall.pdf (June 30, 2017).

Wadud, Z., MacKenzie, D., & Leiby, P. (2016). Help or hindrance? The travel, energy and carbon impacts of highly automated vehicles. *Transportation Research Part A: Policy and Practic*e, 86, 1-18.

Wael Al Rahal Al Orabia, Sawsan Abdul Rahmana, May El Barachib, Azzam Mourada (2016). Towards On Demand Road Condition Monitoring Using Mobile Phone Sensing as a Service. The 7th International Conference on Ambient Systems, Networks and Technologies (ANT 2016), pp. 345–352.

van Wageningen-Kessels, F. (2016). Framework to Assess Multiclass Continuum Traffic Flow Models. *Transportation Research Record*, (2553), 150-160.

Wakabayashi, (2018). Uber's Self-Driving Trucks Hit the Highway, but Not Local Roads. URL: https://www.nytimes.com/2018/03/06/technology/uber-self-driving-trucks.html

Waller, S.T. and Ziliaskopoulos, A.K., (2002). On the online shortest path problem with limited arc cost dependencies. *Networks*, 40(4), pp.216-227.

Waller, S. T., K. M. Kockelman, S. V. Ukkusuri, and S. Boyles (2006). Examining the benefit of accounting for traffic dynamics in congestion pricing applications. In Transportation Research Board 85th Annual Meeting, Number 06-2770.

Walters, A.A. (1961). The theory and measurement of private and social cost of highway congestion. *Econometrica: Journal of the Econometric Society*: 676-99.

S. Wan and E. Foxlin (2001). "Improved pedestrian navigation based on drift-reduced MEMS IMU chip", Proceedings of the 2010 Iternational Technical Meeting of the Institute of Navigation, 2001, pp. 220-229.

R. Wang and D. B. Work (2014). "Interactive multiple modelensemble Kalman filter for traffic estimation and incidentdetection," in 17th International IEEE Conference on Intelligent Transportation Systems (ITSC). IEEE, 2014, pp. 804–809.

Wardrop, J. (1952). Some theoretical aspects of road traffic research. Proceedings of the Institute of Civil Engineers, Part II, 325–378.

Washington State. Office of Governor. (2017). Executive Order 17-02, Autonomous Vehicle Testing and Technology in Washington State and Autonomous Vehicle Work Group. URL: http://www.governor.wa.gov/office-governor/official-actions/executive-orders

Webster, F. V. (1958). Traffic signal settings. Technical report.

S. Weiss, D. Scaramuzza and R. Sierwart (2011). "Monocular-SLAM-based navigation for autonomous mocro helicopters in GPS-denied environments". Journal of Field Robotics, vol. 28, no. 6, pp.854-874.

Wie, B. (2007). Dynamic stackelberg equilibrium congestion pricing. *Transportation Research Part C: Emerging Technologies* 15(3), 154–174.

Wie, B. and R. L. Tobin (1998). Dynamic congestion pricing models for general traffic networks. *Transportation Research Part B: Methodological* 32(5), 313–327.

Winter, K., Cats, O., van Arem, B. & Martens, K. (2017). Impact of relocation strategies for a fleet of shared automated vehicles on service efficiency, effectiveness and externalities. 2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), Naples, 2017, pp. 844-849. doi: 10.1109/MTITS.2017.8005630

Work, D. B., Tossavainen, O. P., Blandin, S., Bayen, A. M., Iwuchukwu, T., & Tracton, K. (2008, December). An ensemble Kalman filtering approach to highway traffic estimation using GPS enabled mobile devices. In Decision and Control, 2008. CDC 2008. 47th IEEE Conference on (pp. 5062-5068). IEEE.

Yang, F., Y. Yin, and J. Lu (2007). Steepest descent day-to-day dynamic toll. *Transportation Research Record* 2039, 83–90.

Yang, H., Q. Meng, and D. Lee (2004). Trial-and-error implementation of marginal-cost pricing on networks in the absence of demand functions. *Transportation Research Part B: Methodological* 38(6), 477–493.

Yang, H., W. Xu, B. sheng He, and Q. Meng (2010). Road pricing for congestion control with unknown demand and cost functions. *Transportation Research Part C: Emerging Technologies* 18(2), 157 – 175.

Yang, H. and X. Zhang (2008). Existence of anonymous link tolls for system optimum on networks with mixed equilibrium behaviors. *Transportation Research Part B: Methodological* 42(2), 99–112.

Yang, H., X. Zhang, and Q. Meng (2007). Stackelberg games and multiple equilibrium behaviors on networks. *Transportation Research Part B: Methodological* 41(8), 841–861.

Yang, Q. and H. N. Koutsopoulos (1996). A microscopic traffic simulator for evaluation of dynamic traffic management systems. *Transportation Research Part C: Emerging Technologies* 4(3), 113–129.

Yap, M. D., Correia, G., & Van Arem, B. (2016). Preferences of travellers for using automated vehicles as last mile public transport of multimodal train trips. *Transportation Research Part A: Policy and Practice*, 94, 1-16.

Ye, H., H. Yang, and Z. Tan (2015). Learning marginal-cost pricing via a trial-and-error procedure with dayto-day flow dynamics. *Transportation Research Part B: Methodological* 81, Part 3, 794 – 807. {ISTTT} 21 for the year 2015SI: {ISTTT21}.

Ye, D., Gong, D., Wang, W. Application of wireless sensor networks in environmental monitoring. Power Electronics and Intelligent Transportation System (PEITS), 2009 2nd International Conference, pp. 1008-1112.

Yin, Y. and Y. Lou (2009). Dynamic tolling strategies for managed lanes. Journal of Transportation Engineering 135(2), 45–52.

Stan Young (2007). Real-Time Traffic Operations Data Using Vehicle Probe Technology. Proceedings of the 2007 Mid-Continent Transportation Research Symposium, Ames, Iowa.

Yperman, I.S. , (2007). The Link Transmission Model for Dynamic Network Loading. KU Leuven dissertation.

Yperman, I.S. , Logghe, S , Tampère, C.M.J. , Immers, L.H. (2006). The multi-commodity Link Transmission Model for dynamic network loading. Transportation Research Board Annual Meeting, 85.

Zhang, X., & Yang, H. (2004). The optimal cordon-based network congestion pricing problem. *Transportation Research Part B: Methodological*, 38(6), 517-537.

Zhang, G., Y. Wang, H. Wei, and P. Yi (2008). A feedback-based dynamic tolling algorithm for highoccupancy toll lane operations. *Transportation Research Record* (2065), 54–63.

Zheng, N., R. A. Waraich, K. W. Axhausen, and N. Geroliminis (2012). A dynamic cordon pricing scheme combining the macroscopic fundamental diagram and an agent-based traffic model. *Transportation Research Part A: Policy and Practice* 46(8), 1291–1303.

Zheng, F., and Henk Van Zuylen (2013). Urban link travel time estimation based on sparse probe vehicle data. Transportation Research Part C, 31, pp. 145-157.

Zhao, Y. and Kockelman, K., (2017). Anticipating the Regional Impacts of Connected and Automated Vehicle Travel in Austin, Texas. International Journal of Sustainable Transportation.

Zhou, B., M. Bliemer, H. Yang, and J. He (2015). A trial-and-error congestion pricing scheme for networks with elastic demand and link capacity constraints. *Transportation Research Part B: Methodological* 72, 77–92.

Zhu, M., X. Li, H. Huang, L. Kong, M. Li, and M.-Y. Wu (2009). Licp: A look-ahead intersection control policy with intelligent vehicles. In Mobile Adhoc and Sensor Systems, 2009. MASS'09. IEEE 6th International Conference on, pp. 633–638. IEEE.

Appendix 1: Formulation of Boundary and Internal Conditions based on Triangular Fundamental Diagram

Definition of Initial, Upstream, Downstream, and Internal Conditions

The initial condition can be expressed as a piecewise linear function, with each linear piece defined by:

$$c_{ini_i}(x) = \begin{cases} -k_i x + b_i & : x_i \le x \le x_{i+1} \\ +\infty & : otherwise \end{cases} \quad (1)$$

With the above definition, the initial condition can be written as $c_{ini} = \min_i N_{ini_i}$

Similarly, the upstream boundary condition is assumed to be piecewise linear, with each piece defined by:

$$c_{up_j}(t) = \begin{cases} q_j t + d_j & : t_j \le t \le t_{j+1} \\ +\infty & : otherwise \end{cases} \quad (2)$$

With this definition, the upstream boundary condition can be written as $c_{up} = \min_j N_{up_j}$

The downstream boundary condition is also assumed to be a piecewise linear function, with each piece defined by:

$$c_{down_j}(t) = \begin{cases} p_j t + c_j & : t_j \le t \le t_{j+1} \\ +\infty & : otherwise \end{cases} \quad (3)$$

This enables us to define the downstream boundary condition function as $c_{down} = \min_j N_{down_j}$,

One of the major results of Mazaré et al. (2011) is that the solutions associated with each linear piece of the initial, upstream, downstream, and internal boundary conditions can be computed analytically as follows:

Solution to a Linear Initial Condition

If $0 \le k_i \le k_c$, the initial condition imposes a free-flow state.

$$N_{c_{ini}}(x,t) = \begin{cases} k_i(tv_f - x) + b_i & : x_i + tv_f \le x \le x_{i+1} + tv_f \\ k_c(tv_f - x) + b_i + x_i(k_c - k_i) & : x_i + tw \le x \le x_{i+1} + tv_f \end{cases} \quad (4)$$

else, if $k_c, \leq k_i \leq k_j$, the initial condition imposes a congested state

$$N_{c_{ini}{}^i}(x,t) = \begin{cases} k_i(tw-x) - tk_j w + b_i & : x_i + tw \leq x \leq x_{i+1} + tw \\ k_c(tw-x) - tk_j w + x_{i+1}(k_c - k_i) + b_i & : x_{i+1} + tw \leq x \leq x_{i+1} + tv_f \end{cases} \quad (5)$$

Solution to a Linear Upstream Boundary Condition

For an upstream boundary condition N_{up} defined as: ${N_{up}}^j(t) = q_j t + d_j$ with $d_j = -q_j t + \sum_{l=0}^{j-1}(t_{l+1} - t_l)\, {q_j}^l$, the solution component can be expressed as:

$$N_{c_{up}{}^j}(x,t) = \begin{cases} d_j + q_j\left(t - \dfrac{x - x_0}{v_f}\right) & : x_0 + v_f(t - t_{j+1}) \leq x \leq x_0 + v_f(t - t_j) \\ d_j + q_j t_{j+1} + k_c\left((t - t_{j+1})v_f - (x - x_0)\right) & : x_0 \leq x \leq x_0 + v_f(t - t_{j+1}) \end{cases} \quad (6)$$

Solution to a Linear Downstream Boundary Condition

For a downstream boundary condition ${N_{down}}^j$, defined as ${N_{down}}^j(t) = p_j t + b_j$ with $b_j = -p_j t + N_{ini}^{(n-1)}(x_n) + \sum_{l=0}^{j-1}(t_{l+1} - t_l)\, {q_j}^l$, the solution component can be expressed as:

$${N_{down}}^j(x,t) = \begin{cases} b_j + p_j t - \left(\dfrac{p_j}{w} + k_j\right)(x_n - x) & : x_n + w(t - t_j) \leq x \leq x_n + w(t - t_{j+1}) \\ b_j + p_j t_{j+1} + k_c\left((t - t_{j+1})v_f + x_n - x\right) & : x_n + w(t - t_j) \leq x \leq x_n \end{cases} \quad (7)$$