Are We There Yet? A Proposal for an Autonomous Taxi System in New Jersey and a Preliminary Foundation for Empty Vehicle Routing

Jaison J. Zachariah
Advisor: Dr. Andrew L. Papanicolaou

Submitted in partial fulfillment of the requirements for the degree of Bachelor of Science in Engineering
Department of Operations Research and Financial Engineering
Princeton University

June 2013
I hereby declare that I am the sole author of this thesis.

I authorize Princeton University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

______________________________
Jaison J. Zachariah

I further authorize Princeton University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

______________________________
Jaison J. Zachariah
Abstract

Vehicles are ubiquitous in society. They engender mobility, allow commerce to flourish, and increase the utility of all. However, inefficient systems cost people safety, money, and mobility. Fully automated vehicles, which have had massive recent technological success, offer solutions that address the issues with traditional human operated vehicles.

This thesis seeks to take advantage of the benefits of autonomous vehicles by implementing an autonomous taxi system in New Jersey. First, this thesis outlines the design of a system of autonomous taxis (aTaxis) that permits ridesharing. The system will then be simulated; a fleet of aTaxis will be assembled and a synthesized dataset that contains specific travel behavior of all residents of New Jersey will be used to inform the system. The performance of this system under different parameters will be assessed for state of New Jersey. In particular, congestion reduction due to ridesharing will be examined. A framework for optimal empty vehicle routing will then be explored to gain even more efficiencies from the system.
Acknowledgements

I would like to thank Professor Kornhauser for providing me with the raw dataset that would allow me to run this simulation. It took quite some time and effort to get it right, but we finally did it. Your help along the way was invaluable. The way you approach this subject reveals your passion; it was a pleasure to work with you on this research.

I would also like to thank the 2011 ORF 467 class and Talal Mufti for enhancing the Trip Synthesizer that created the novel dataset. Jinka, your identification and correction of problems in the synthesizing process was crucial to my results. Thanks to the 2012 ORF 467 class who also helped monumentally in parsing the data (and stumbling on errors) and setting up the project. Good luck to the 2013 ORF 467 class. I hope you can use this thesis as a foundation for your research.

I would like to thank Michael Bino who put up with my incessant pestering to get the database and other technical items squared away. Also, I would like to thank my thesis fairies / goddesses – Amy Oh, Bells, Kacie, Mel, and Tory – for helping me “keep on keepin’ on.” Mikey, Moff, Todd, and Fox, thanks for keeping me sane. Thank you to all my proofreaders. You guys did a phenomenal job!

Finally, Dr. Papanicolaou, thank you for your guidance and advice along the way. I know this research is not your primary area of expertise, but nonetheless you did a superb job in advising a project for which I had passion and excitement. This research is the future. Best of luck Down Under and in all future ventures.
This thesis is dedicated to all those who gave me the support to get through this enriching academic venture. Mom, Dad, Chris, Mani, Siju, Family, Sprint Boys, and Friends, Thanks for believing in me and putting up with “the thesis.” I hope this product justifies the reduced engagements with you all. You will all be riding in Autonomous Vehicles in the near future!
Man as a toolmaker has the ability to make a tool to amplify an inherent ability that he has.

-Steve Jobs
# Contents

Abstract ................................................................. iii
Acknowledgements ....................................................... iv
List of Figures .............................................................. ix

1 Introduction .............................................................. 1
1.1 Overview .............................................................. 1
1.2 Importance of Autonomous Vehicles .......................... 2
  1.2.1 Safety Considerations ....................................... 3
  1.2.2 Congestion and Mobility Considerations ............... 4
  1.2.3 Environmental Considerations ................................ 4
1.3 Existing Research and Literature ............................. 5
  1.3.1 Personal Rapid Transit ...................................... 5
  1.3.2 Travel Demand Data ......................................... 6
  1.3.3 Autonomous Taxi System Feasibility Analysis .......... 9
1.4 Goals ................................................................. 10

2 Dataset and Autonomous Taxi System Design ..................... 12
2.1 Trip Synthesizer ................................................... 12
  2.1.1 Trip Synthesizer Foundations .......................... 13
  2.1.2 Multi-Modal Adjustments .................................. 15
2.2 Autonomous Taxi System Design .................................. 17
  2.2.1 Autonomous Taxi Stand Grid ................................ 17
  2.2.2 Trip Definitions .......................................... 19
  2.2.3 Rideshare Methodology ................................... 19
2.3 Metrics for Analysis ............................................... 23
  2.3.1 Volume Assessment ....................................... 23
  2.3.2 Rideshare Assessment .................................. 23

3 Simulation Implementation ........................................... 25
3.1 Modules .......................................................... 25
  3.1.1 aTaxi.java .................................................. 25
  3.1.2 DestinationQueue.java .................................... 26
  3.1.3 DepartureQueue.java .................................... 27
  3.1.4 RideShare.java ........................................... 27
  3.1.5 RideShareAnalysis.java ................................ 28
List of Figures

1.1 The Frequency of Transportation Related Incidences in the United States. . . . . . . 3
1.2 Example of a PRT System. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 6
1.3 Comparison of Tour and Trip Concepts. . . . . . . . . . . . . . . . . . . . . . . . . . 7
1.4 General Schematic for Gravity Model. . . . . . . . . . . . . . . . . . . . . . . . . . 8
1.5 Depiction of Cost Savings in Ann Arbor for Various Modes of Transportation. . . . 10

2.1 Comparison of Synthesizer Output and 2010 Census Data - Population and Gender. 14
2.2 Comparison of Synthesizer Output and 2010 Census Data - Age and Gender. . . . . 14
2.3 List of All Tour Types . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 15
2.4 Multi-Modal Adjustments: Utilizing New Jersey Transit Trains . . . . . . . . . . . . 16
2.5 Grid System - New Jersey. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 18
2.6 Zoom into Grid System - Mercer County. . . . . . . . . . . . . . . . . . . . . . . . . 19
2.7 The Average Vehicle Occupancy vs. Departure Delay for Essex County. . . . . . . . 20
2.8 The Average Vehicle Occupancy vs. Common Destinations for Essex County. . . . 21

3.1 Graphical View of Data Structure Relationships. . . . . . . . . . . . . . . . . . . . . 26
3.2 Example aTaxi Tour. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 32
3.3 Tabular Comparison of Cape May AVO for Different Permutation Levels. . . . . . . 33
3.4 Graphical Comparison of Cape May AVO for Different Permutation Levels. . . . . 34

4.1 Mercer County in New Jersey. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 37
4.2 Mercer County Statistics . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 38
4.3 Mercer County Cumulative Trip Distribution by Time . . . . . . . . . . . . . . . . . 39
4.4 Mercer County Cumulative Trip Distribution by Highest Volume Pixels . . . . . . 39
4.5 Mercer County Top 100 Trip Producing Pixels. . . . . . . . . . . . . . . . . . . . . . 40
4.6 Mercer County Highest Volume Pixel. . . . . . . . . . . . . . . . . . . . . . . . . . . 41
4.7 U.S. Average Vehicle Occupancy Reported by National Household Travel Survey. . . 42
4.8 Mercer County AVO vs CD. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 43
4.9 Mercer County AVO vs DD. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 43
4.10 Mercer County MSR vs CD. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 45
4.11 Mercer County MSR vs DD. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 45
4.12 NJ Statewide Statistics. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 47
4.13 NJ Cumulative Trip Distribution by Time . . . . . . . . . . . . . . . . . . . . . . . . 48
4.14 NJ Cumulative Trip Distribution by Distance . . . . . . . . . . . . . . . . . . . . . . 49
4.15 NJ Cumulative Trip Distribution by Highest Volume Pixels . . . . . . . . . . . . . 49
Chapter 1

Introduction

1.1 Overview

Autonomous vehicle research and technology have been progressing at a very rapid pace over the past decade. From the Google car to major car manufacturers, many prominent players in technology and transportation have invested significant amounts of time and effort to make full automation a reality. Autonomous functions in vehicles have reached many milestones; over the past decade, features such as automated parking, lane detection and control, and radar powered distance control have automated functions that improve the driving experience. However, fully automated (or driverless) vehicles are on the verge of becoming a commercial reality. For instance, the Google Car fleet has successfully navigated 300,000 fully autonomous miles of the streets and highways of Mountain View, California [25].

It is very possible that these vehicles will become more common in five to seven years. According to Diana Samuels [22], a technology expert and contributor to the Silicon Valley Business Journal, the push toward automation is fueled by the consumers themselves so the utilization of these autonomous vehicles will be realized as soon as it becomes commercially and technologically feasible. Since an autonomous vehicle is not dependent on an operator, it no longer needs to remain stationary when that operator is not using it. It can continue to do useful work by simply transporting another rider. Thus, instead of a system of ownership of vehicles, commercialization of autonomous vehicles might create an environment where autonomous vehicles might be shared by the community. In this sense, these autonomous vehicles serve more of the function of autonomous taxis as they shuttle individuals to destinations and then seek more passengers to transport. This thesis seeks to lay the groundwork for such an autonomous taxi system and begin to assess its properties and address its logistics.

Example

Andrew is a typical vehicle owner in the state of New Jersey. Every day, his vehicle costs him money through a combination of maintenance, gasoline, insurance, parking and depreciation. He spends an hour each day commuting to and from work and makes other trips during his day. His schedule is hectic and his time is valuable.

A new mobility service implements a coordinated autonomous vehicle infrastructure in New
Jersey. Desiring a ride, Andrew can now simply request one using his smartphone. The vehicle will arrive at the pick-up point and Andrew can comfortably ride in the car without worrying about the actual operation of the vehicle. At the destination, Andrew exits the autonomous vehicle which is then routed to the next pick-up point.

In this example, Andrew is offered a safe, convenient, and time efficient ride in the autonomous vehicle for a fraction of the cost of owning and operating his personal vehicle. This autonomous taxi system in New Jersey enhances his mobility experience and provides the essential transportation service without the inconvenience, dangers, and costs of driving.

Benefits of an Autonomous Taxi System with Ridesharing

The autonomous taxi concept can be developed further by considering ridesharing opportunities. Just as travelers who are destined to the same or nearby locations can share a taxicab, individuals with similar location should be able to share an autonomous taxi. If there exists a significant rideshare potential, then such a system could have lasting impacts on the way we travel. This would also have a significant impact on reducing the number of vehicles on the road and reducing congestion.

This sentiment is precisely what the first part of this thesis seeks to explore. Using a dataset of the simulated daily tours of every individual in New Jersey on a typical weekday, an aTaxi system will be set up in the state of New Jersey. Simulations can assess the rideshare potential and feasibility of the system. Once the potential is ascertained, methods to make the system more efficient can be explored. For instance, the intelligent routing of unused or empty cars will allow maximum utility to be extracted from the system. The latter portion of this thesis lays the groundwork to formulate a solution to that problem.

1.2 Importance of Autonomous Vehicles

According to renowned economist Joseph Schumpeter (1883 – 1950):

\[\ldots\text{In capitalist reality\ldots, it is not [price] competition which counts but the competition from the new commodity, the new technology\ldotscompetition which commands a decisive cost or quality advantage and which strikes not at the margins of the profits and the outputs of the existing firms but at their foundations and their very lives.}\]

The steam powered boat, the Bessemer steel making process, and the transistor all upended industries and impacted the course of human history. Autonomous vehicles have the capacity to change the landscape of transportation as it exists now. Transportation affects almost all aspects of life and is linked to many large industries such as insurance, healthcare, shipping, and defense. Autonomous vehicles can fundamentally change the way many industries operate today.

The core offerings of autonomous vehicles in an aTaxi system are:

- improved safety,
- reduced congestion and greater mobility, and
- minimized environmental impact and energy costs.
1.2.1 Safety Considerations

In terms of safety, highway accidents far outnumber the accidents on all other modes of transportation combined (See Figure 1.1). According to the United States Department of Transportation, there were 2.2 million injuries and 33,808 deaths in the U.S. caused by highway accidents in 2009 [28]. These numbers do not even include accidents in local traffic where pedestrians and the stop and go nature of traffic can contribute toward a large number of incidences. Moreover, according to Bob Goos, chairman of the International Organization for Road Accident Prevention, “road danger is a man-made crisis, with human error accounting for over 90 percent of accidents” [18].

Globally, there are over 1 million road traffic fatalities per year. Countries with poorer infrastructure and higher densities of vehicle use have very high rates of incidences. For instance, China has over 68,000 deaths\(^1\) due to traffic accidents per year and; it is the leading cause of death for people under the age of 45 [1]. Smart automated vehicle systems can address this issue and save many lives.

The human component of driving is what poses the greatest risk for accident. Autonomous vehicles remove the need for an operator and thereby have the potential to substantially reduce the accident numbers. With smart software wired to state-of-the-art hardware that processes swats of data instantaneously and allows the vehicle to react quicker than humans, the autonomous vehicle truly offers a great advantage over human operated vehicles.

---

\(^1\)This is an underreported number. According to Keith Bradsher [1], chief Hong Kong correspondent for the New York Times, Chinese authorities have a history of manipulating incidence reports.
1.2.2 Congestion and Mobility Considerations

Autonomous vehicles have the technological capacity to drive better than the average human driver. This is because autonomous vehicles do not engage in behaviors that cause traffic delays. For instance, sporadic and unnecessary braking by humans cause traffic waves, a phenomenon which describes how congestion travels and amplifies backward in a wavelike form. Autonomous vehicles will only brake when necessary and thereby reduce traffic waves. Moreover, traffic disturbances would also be dealt with more efficiently. For instance, if an accident shuts down a lane, autonomous vehicle systems can efficiently coordinate the merging onto the open lane and begin to shift vehicles toward that open lane some distance away from the accident.

Moreover, external systems and information can be used to construct the optimal path to the destination in real-time, which would make the travel even more efficient. These systems can avoid the areas where road demand outweighs road capacity. Currently, a traveler in New Jersey spends about 48 hours stuck in highway traffic every year [29]. This figure is averaged over all travelers, so individuals who commute more will experience even greater time spent in traffic. With a smart autonomous vehicle system, factors that add to congestion will permit a more efficient system and liberate up to two days a year in travelers’ lives.

Now, take for example a fleet of these autonomous vehicles that offers taxi services that permitted rideshares. This would create a system that reduces traffic congestion and minimizes delays since relatively few vehicles on the road could service the existing demand of a region. Hence, there is better capital utilization due to the reduced quantity of vehicles. In addition, there is a greater capacity utilization since each of these cars is more productive. Researchers at Columbia University recently conducted a study that displayed the effects on capital that autonomous taxi systems could have. They discovered that with the institution of a comprehensive autonomous vehicle system, the taxicab fleet of Manhattan could be reduced by up to 80% [3].

This automated taxi system would offer a greater mobility experience as well. This includes a high quality experience during the travel. Without the need of an operator, riders can utilize the time spent traveling for leisure or work. In addition, autonomous taxis would be nonstop and direct. Unlike trains which have a set number of stops where people load or unload, autonomous vehicles take one directly to the destination. This aTaxi system is also demand responsive. That is, there is no set schedule. Either an individual comes to an aTaxiStand with waiting aTaxis or an individual simply requests one and waits for a prompt arrival. Finally, it is area wide. Unlike a train or other public transport system which travels across a fixed and restricted guideway, autonomous taxis have the capability of delivering passengers wherever a regular car could drive.

1.2.3 Environmental Considerations

Autonomous taxis also reduce the environmental impact of driving by operating the vehicle in a more efficient way than humans. With shared ridership comes the potential overall decrease in the number of vehicles on the road from better capital utilization. This will allow for faster moving traffic and overall better riding experience due to less congestion. With these efficiencies realized, the lower level of emissions could improve environmental conditions.

In addition, humans drive very inefficiently. Autonomous vehicles will eliminate these suboptimal behaviors such as the rapid acceleration, excessive braking, and aggressive driving. Autonomous
vehicles can be designed to operate at the optimal performance level to maximize fuel economy. Moreover, with the rise of autonomous vehicles, basic vehicle design will likely change. With this alteration in paradigm and design, basic items needed in human operated vehicles can be eliminated. These actions could include optimizing the form of the vehicle and making the vehicle lighter, which could lead to more efficient transportation.

1.3 Existing Research and Literature

The idea of adopting automated units as a primary mode of transportation is not novel. In 1939, the World Fair presented the “World of Tomorrow” in which the most popular attraction was Futurama, an exhibit by General Motors that presented the vision of society in 1960. This included a futuristic vision of transportation, “a network of thousands of scale miles of ‘magic motorways’ along which 16,000 tiny autos and trucks zipped among cities and farms without delay or congestion” [19].

In 1977, the Tsukuba Mechanical Engineering Lab created the first autonomous vehicle that “that functioned by following white street markers and was able to reach speeds of up to 20 mph on a dedicated test course” [7]. Later that year, Elbert Sawyer was granted the patent for the Personal Rapid Transit Unit (PRT) the first feasible guided transportation unit that could be expanded at a larger scale [23].

In 1995, a pioneer of driverless cars, Ernst Dickmanns applied his computer vision research to autonomous vehicles. In his paper, *Vehicles Capable of Dynamic Vision*, he outlined the fundamental theoretical and computational methods that allow vision to guide autonomous vehicles [4]. He successfully applied this research and drove 1,000 miles autonomously round trip between Munich and Denmark with the Bundeswehr University Munich’s Mercedes S-Class.

1.3.1 Personal Rapid Transit

The vision of converting to an automated, taxi-like mode of transportation has also existed for many years. One autonomous taxi-like system that still has a particularly strong following is the Personal Rapid Transit system (see Figure 1.2). This system relies on a “dedicated guideway” [2] as the conduit for movement. Although extensive research has been conducted on its feasibility and it has been successful in some implementations such as Morgantown, West Virginia [10], PRTs have failed to obtain widespread adoption.

The reason why the PRT has not progressed is its dependence on an extensive dedicated guideway infrastructure required to interconnect all of the stations. While PRT stations can be readily integrated in facilities that would embrace the accessibility it offered and residential areas that tend to have open space that make ideal station locations, the guideways pose a significant challenge. It is expensive to place them underground, considered unsafe to place on ground level, and societally unacceptable to place them above ground. Who would want a PRT outside their window?

However, autonomous vehicles offer the opportunity to address both shortcomings of PRT and have the opportunity to deliver mobility all throughout New Jersey. A system consisting of autonomous vehicles operating from designated stands located throughout New Jersey where aTaxis would travel between stands using New Jersey’s existing roads and highways. Such a system can readily begin to operate effectively with a single pair of stands since they are developed to work with
existing infrastructure and human operated vehicles. Additional aTaxi stands can then be located wherever customer demand is sufficient.

As the system expands, the mobility afforded grows more quickly. Much later, a point of diminishing return will be reached as the stands are located in areas with less consumer demand. However, this point would be close enough to a 100% service rate that the remaining mobility can be offered at very small additional cost. Importantly, the system has the opportunity to grow naturally from minimal beginnings to serve much of the region.

1.3.2 Travel Demand Data

This thesis seeks to simulate such an autonomous taxi system. However, before the system of aTaxis can be set up, travel demand data is needed to determine where trips are produced and where they are destined. The system and the travel demand data are the two distinct and important components required for the analysis. The travel demand data should provide information about the trips in the area of interest with temporal and spatial specificity; that is, it should indicate exactly where and exactly when people are making trips. Once this data is obtained, it can be used to feed the aTaxi system simulation. From here, the performance of the theoretical aTaxi system can be assessed.

Tour and Activity Based Travel Models

Trip forecasting is an intricate process that has been researched thoroughly. One methodology that has met some success in accurately predicting trips has been tour and activity based travel models
A tour model is defined by a sequence of trips that starts and ends at home. A tour can be thought of as a generalization of primary travel purpose; for instance, an individual who leaves Home to go to Work is considered taking a Work tour. Tour purposes are typically Home, Work, School, or Other. A trip, on the other hand, is one leg of the entire tour. This means that trip ends are specific homes, workplaces, schools, or other locations that make up the series of start and end points during travel. See Figure 1.3 for an illustration of these concepts.

![Figure 1.3: Left: Example of Trips. Right: Example of Tours.](image)

An activity based model treats travel as a derived demand; that is, it assumes individuals travel with the desire to participate in activities at the trip end rather than the desire to simply be in a vehicle. Activity-based models are capable of modeling the tradeoff between participating in an activity as a stop on the way to another activity and making a new tour for that activity. This tradeoff can be a function of the desirability of the activity’s location. In these models, individuals living in accessible regions have a greater likelihood of making additional tours, whereas “persons in inaccessible locations will seek chain trips into more complex tours” [5].

One particular version of a tour and activity based model is the gravity model (see Figure 1.4). Essentially this divides a region according to its Travel Area Zone (TAZ), which is the type of land it is (residential, commercial, etc.). Associated with each of these TAZ is a production and attraction vector that records the number of trips that the TAZ produces and the number of trips the TAZ attracts. These values are determined by the attributes and land type of the TAZ. Trips are divided into type (Home to Work, Home to School, etc.) and a matrix indicating flow of trips from one TAZ to another is created.

The gravity model assumes that trips produced at an origin and attracted to a destination are directly proportional to the total trip productions at the source and the attraction at destination. It includes a friction or disutility factor \((F)\) that marks the impedance of making the trip. These impedances will alter how attractive a particular zone is relative to any point in the city. This typically is some function of distance since longer trips are less apt to be taken.

To determine the number of trips:

\[
T_{ij} = \frac{A_j F_{ij} K_{ij}}{\sum_{all	ext{zones}} A_k F_{ij} K_{ik}} \cdot P_i
\]

Where:

\(i\) = origin zone
\(j\) = destination zone
\(T_{ij}\) = trips produced at i and attracted at j
Figure 1.4: General Schematic for Gravity Model.

\[ P_i = \text{total trip production at } i \]
\[ A_j = \text{total trip attraction at } j \]
\[ F_{ij} = \text{a calibration term for interchange } ij \text{ (e.g. } \frac{1}{\text{Distance}}) \]
\[ K_{ij} = \text{a socio economic adjustment factor for interchange } ij \]

This model will be applied across all trip types to determine the tour behavior for each type of trip. This will produce an origin to destination trip matrix that links a TAZ to a TAZ.

Unfortunately, this method lacks robustness since it requires accurate calibration of the tunable parameters, \( K_{ij} \) and \( F_{ij} \). According to studies conducted by researchers at the University of Manitoba, the tunable parameters over a 20 period were rather unstable and inconsistent from one prediction time frame to the next. Hence, they are “not appropriate for use in predicting origin-destination matrices” [6].

Additionally, the demand analysis using this method is insufficient for simulation purposes since it is not precise enough spatially to properly evaluate the walk disutility associated with accessing the aTaxi station locations. For instance, the size of a TAZ can expand large areas of land so characteristics must be generalized for that entire region. Moreover, there is no time component kept during the analysis and this temporal imprecision prevents the gathering of adequate departure and rideshare information.

**A New Approach to Demand Modeling**

Other demand models that forecast trip behaviors exist. These include the dynamic network model, land use models, traffic models, and a number of others. Refer to Donnelly et al. for a detailed analysis of forecasting measures.

The fundamental problem that exists in all these models is that time and space specificity is
not attained. In order to overcome these deficiencies, this thesis will utilize a novel approach to travel demand modeling — using an individual trip synthesizer. With this synthesizer, a tour can be created at an individual traveler’s level and attributes such as time and location of departure can be assigned to each trip of this individual. In this way, the trips of every individual in a given area can be simulated. The individual trip synthesizer that generated the entire trip set for travelers in New Jersey will be discussed in Section 2.1.

1.3.3 Autonomous Taxi System Feasibility Analysis

Researchers at the Earth Institute at Columbia University in conjunction with major industry players such as General Motors, Volvo Group and Verizon Wireless conducted a thorough analysis of the financial and technological feasibility of autonomous taxi systems [3]. They identified six major technologies that would provide “mobility experiences at a radically lower cost.” These include:

- The “Mobility Internet”: cloud based coordination of real-time spatial and temporal data of vehicles and travelers,
- Self Driving Cars: vehicles operated without human intervention,
- Shared Vehicles: vehicles used by multiple people throughout the day,
- Specific-Purpose Vehicle Designs: vehicles tailored to type of mobility and number of passengers,
- Advanced Propulsion Systems: systems that utilize alternate sources of energy to power vehicles, and
- Smart Infrastructure: infrastructure that uses sensors and communication to utilize resources more efficiently.

At the intersection of these ideas, a system of reduced congestion and optimal mobility can be reached. The researchers also conducted case studies on various types of cities in the United States to measure the impact that autonomous an autonomous taxi system could have on mobility. These cases illustrate the transformative power of autonomous vehicles and the extent of congestion reduction and cost savings across all types of mobility infrastructures across the United States and the world.

Case Study: Ann Arbor, Michigan

Ann Arbor was selected as a representative medium sized city. With a population of 285,000 making 740,000 trips a day in 200,000 personally owned vehicles, it made an excellent candidate for analysis.

The researchers obtained travel data for Ann Arbor and used queueing and network models to determine the performance of shared, driverless vehicles. The fleet size that adequately covered most demand with acceptable wait times was determined and a cost analysis was conducted. The cost of providing the automated service was compared against the current cost of personal car ownership. The analysis focused on the urban area of Ann Arbor where 120,000 car serviced 528,000 trips per day. Results from the automated vehicle simulations indicated that the same number of trips could be serviced by just under 20,000 vehicles — reducing the number of vehicles in urban areas...
by 6x. Moreover, the simulations identified major cost savings potential (up to 10x) from switching to personally owned vehicles to an autonomous system (See Figure 1.5).

Additionally, profit analysis was done for a hypothetical autonomous taxi service provider. An automated mobility service that provides “customers with a service comparable to car ownership with better utilization of their time” priced at $7 per day and utilized by 100,000 residents of Ann Arbor could result in a profit of $500,000 a day.

Case Study: New York, New York

Manhattan has a population of 1.6 million people living in 23 square miles of land. Very few residents own cars and the primary mode of transportation is walking or public transportation (subway, taxicabs). Researchers studied Manhattan to determine the impact of densely populated city centers with limited vehicular mobility. In particular, the study compared the cost of the autonomous fleet to taxicabs within Manhattan since that remained the primary competition.

Researchers aggregated taxicab trip data for Manhattan and used queueing and network models to assess capability of the autonomous system to meet demand. A fleet size was determined under acceptable parameters (such as waiting times and coverage). Once the fleet size was determined, a cost analysis was conducted.

Researchers discovered that the 13,000 taxicabs and 40,000 for-hire vehicles account for 470,000 trips per day in the five boroughs with 88% occurring within Manhattan. The simulations indicated that a fleet of 9,000 autonomous vehicles could meet the demand with shorter wait times and greater capacity utilization. In terms of expenses, researchers concluded that the cost of a taxicab service is approximately $4 per trip-mile. The autonomous vehicle system would reduce this cost to $.40 due to fewer taxi units, reduced labor costs, and fewer empty vehicles.

1.4 Goals

This thesis is split into two primary sections. The first seeks to delve deeper into the congestion reduction properties of an autonomous taxi system. There are many methods of reducing congestion
with autonomous vehicles. Autonomous vehicles can be optimized to avoid traffic slowing behaviors, coordinated to maximize efficiency, and externally informed to find best routes. However, this thesis will address the congestion reduction that can be realized as a result of the fewer vehicles needed in a smart aTaxi system in which ridesharing is realized and utilized. This will maximize capital and capacity utility.

Chapter 2 will explain in detail the trip synthesizer and the aTaxi system that will be instituted. Chapter 3 will cover the implementation of the simulation. Chapter 4 will show the results of the simulations at both a county and statewide level.

In this preliminary simulation, enough vehicles exist to satisfy all the demand. Rideshare potential will be assessed and the reduction in congestion will be assessed by noting the average vehicle occupancy and the miles saved by utilizing this system. In the second portion of the thesis, realistic limitations are imposed and the system is “closed” to utilize a set number of cars that take multiple tours. Since aTaxis are now a scarce resource, routing is crucial to achieve the most impact for the lowest cost. The foundations for routing empty vehicles will be developed in Chapter 5.
Chapter 2

Dataset and Autonomous Taxi System Design

The assessment of the rideshare potential and congestion reduction properties of an autonomous taxi system requires two distinct items: demand data and aTaxi simulation framework. More precisely, the demand data will feed into the aTaxi simulation framework that is created. Spatially and temporally specific data is crucial for a precise simulation. As discussed in Section 1.3.2, traditional models of travel forecasting lack this.

In order to overcome the deficiencies of the traditional model, a simulation of the entire tour of individuals on a given day is necessary. This is precisely what Talal Mufti produces in his master’s thesis [16]. With the dataset in hand, the system of autonomous taxis and the rules for ridesharing can then be explored.

2.1 Trip Synthesizer

The framework of the autonomous taxi system simulation requires trip data with exact time and locations to be delineated. The spatial granularity will determine the precise aTaxiStands locations that travelers in the system will utilize while the temporal granularity will determine exact times of departure. Together, these elements can be used to assess the rideshare potential of a demand dependent system.

In 2011, the Transportations Systems Planning and Analysis class in the Operations Research and Financial Engineering department of Princeton University began to address the demand data shortcomings of conventional demand modeling methods. The class extended the tour and activity based travel models (Section 1.3.2) and initiated the creation of an individual activity-based trip synthesizer to create higher resolution information. In particular, the class laid the foundation for synthesizing individual trip travel demand for all 8,791,894 individuals in New Jersey — a novel approach to travel demand modeling and trip forecasting. This was subsequently enhanced by research conducted by Talal Mufti [2012] and Jingkang Gao [2013].
2.1.1 Trip Synthesizer Foundations

In order to generate the travel demand in New Jersey, the relevant sources of trips were identified. The primary source of the trips were the 21 counties within New Jersey. For accuracy and completeness, trips made by out-of-state residents working in New Jersey were also synthesized. These regions include major hubs (such as Philadelphia, New York City, Rockland County, Bucks County) and more generalized sources (North, South, West).

Mufti outlines his general approach to creating of the dataset:

1. create a population of individuals whose characteristics in aggregate resemble that of New Jersey,
2. assign workplaces and schools,
3. assign tours (activity patterns) and assign the trips that constitute these tours,
4. assign arrival and departure times.

Generation of Populace

The first step of the synthesizer is based on population and household demographics from the 2010 Decennial Census. The utilization of census data to inform the characteristics of the population is an incredibly precise method of simulating the actual population of New Jersey. Information is known down to census block level which is the “smallest geographic unit used by the United States Census Bureau for tabulation of 100-percent data” [26]. The incredible detail of the characteristics of the census blocks can be used to inform the features of the regions in the simulated system. In particular, distributions for factors such as age and salary can be created from census blocks and synthesized residents of this census block will have characteristics drawn from this distribution. After an individual is created, a traveler type of student, worker, or other is assigned based on age and regional attributes. This will be used in subsequent steps to create tours for the individual. In this way, the trip synthesizer generates demographic characteristics for each of the 8,791,894 individuals living in 118,654 census blocks that comprise New Jersey.

The accuracy with which the populace is generated is shown in Figures 2.1 and 2.2. A side by side comparison of census data and synthesizer data shows essentially mirror images. Producing data that matches the actual distribution is exactly the characteristic that is sought after in the trip synthesizer.

Assignment of Anchors

Now that the population has been created, trip characteristics can be assigned. As mentioned in Section 1.3.2, trips are the individual legs of a person’s entire travel itinerary while tours describe the sequence of trips or the general itinerary itself. This terminology will be utilized throughout this thesis.

In designing a simulation of demand forecast, the primary difficulty is predicting where an individual would go. Certain things, such as whether to send a traveler to lunch, are surprisingly difficult to predict. Other things, such as sending an individual to work or school, can be predicted with great certainty since they are common among all individuals in a population. The trips to
more certain types of destinations (such as workplaces and schools) are termed “rigid activities” or “anchors” [30]. These anchors are then assigned. The synthesizer scans the list of all employers and schools in New Jersey and assigns an employer or school to every individual that is identified to make these types of trips. These assignments are made stochastically based on an individual’s attributes, region characteristics from census data, and distance from home.

**Assignment of Tours and Activity Patterns**

Each individual’s demographic signature is used to draw the individual’s daily travel tour behavior from appropriate distributions. In addition, the specific name and address of each establishment visited during a trip is identified by selecting from distributions that are determined by an individual’s attributes, region characteristics from census data, and distance from home. There are 17 types of
tours an individual in the simulation can take (See Figure 2.3). Once a tour is set, the specific trip ends are selected and their location information is recorded.

**Assignment of Temporal Attributes**

The final task of the synthesizer is to assign arrival and departure times, in seconds from midnight, for each trip. For each individual with a specific tour type, the synthesizer checks the types of the nodes (school, workplace, other) involved in the trip as well as other attributes such as the location of the trip within the tour. Arrival time distributions and duration distributions for each type of trip are then used to randomly select precise departure times for each leg of the trip in a tour. The arrival times are computed based on the distance traveled and constant speed of travel which is designated at 30mph. With this portion of the synthesizer completed, both the desired spatial and temporal specificities of the dataset are achieved.

### 2.1.2 Multi-Modal Adjustments

Although Talal’s trip synthesizer generated valid trip data, additional adjustments were made by Jingkang Gao and this author to produce a trip set that reflected the true patterns of travel in the state of New Jersey. A key deficiency in the original model was that it only permitted a unimodal form of transportation. That is, it assumed the only mode of transportation from origin to destination was the aTaxi. This is typically reasonable; however, practicality and efficiency dictate

---

The simulation of the aTaxi system will use Manhattan distances since existing roadways are not used in the system. The use of Manhattan distances will be a better approximation than using Euclidean direct distances since it can partially account for circuitry of existing roadways.
that for longer trips or trips along which robust travel modes already exist, a multi-modal form of transportation is more accurate.

For instance, commuters traveling to New York City in the morning for work currently utilize a bimodal form of transportation. These workers take their personal vehicles to the train station and then ride the train into New York. Similarly, an aTaxi system should reflect the use of train stations and other highly utilized existing modes of transportation.

To reflect this in the synthesizer, trips involving a leg in either New York City or Philadelphia were adjusted to utilize the existing New Jersey Transit train system. The method of approach was to split the trip into two legs — the train ride leg and the aTaxi leg. For trips that had an original destination as New York City or Philadelphia, an aTaxi is taken to the nearest train station. NJ Transit then transports the traveler to New York City or Philadelphia. Similarly, if a trip originated in Philadelphia or New York City, NJ Transit takes the traveler to the nearest train station to the destination. The aTaxi then transports the traveler from the train station to the original destination. Figure 2.4 graphically displays this adjustment. Only the legs involving the aTaxi rides are preserved in the dataset; NJ Transit rides complement the aTaxi system but do not inform them so they are eliminated.

An approximated train schedule was used to adjust the arrival and departure times for each traveler in the simulation. Using the frequency of trains, the speed at which they travel, and the distance traveled, each multi-modal transportation user was given an updated travel schedule that more accurately matched NJ state characteristics. Refer to Gao’s [2013] work for more details on the underlying assumptions and process.

With this final adjustment in place, the synthesizer produces output that details individuals, their attributes, and extensive information about their trips. The outputs are categorized by the
county or out-of-state region of residence. In this way, 28 output files hold comprehensive trip information for every traveler in New Jersey.

2.2 Autonomous Taxi System Design

With a dataset that details the 8.7 million individuals in New Jersey and their 32 million trips, an aTaxi system can be designed. Such a system should be thought of as horizontal elevator:

1. travelers arrive at a station,

2. depending on the parameters of the ridesharing, travelers enter an existing aTaxi going to the same or nearby destination OR enter a new aTaxi,
   
   (a) if a new aTaxi is entered, the aTaxi then waits some departure delay time to see whether another traveler can share that ride, or
   
   (b) if an existing aTaxi is entered, the aTaxi continues waiting the remaining departure delay time that was initiated when the first passenger entered that aTaxi

3. the aTaxi then departs to the destination after waiting the departure delay that was specified.

Under this process an examination of the rideshare potential and the congestion reduction potential of an autonomous taxi system can be conducted.

2.2.1 Autonomous Taxi Stand Grid

The first step in developing an aTaxi system involves the placement of aTaxiStands, docking sites to which travelers go in order to access an aTaxi. A necessary feature of the system that provides it viability is accessibility. Regardless of their location, travelers should be able to simply walk a short distance to the nearest aTaxiStand and wait for an aTaxi to take them to their destination. A stipulation that all trip demand must be serviced will be imposed; hence, aTaxiStands must be pervasive.

Taking accessibility and full servicing into account, the simplest approach of determining the locations of the aTaxis will be to create an array of aTaxiStand locations that covers the entire state. Each of these aTaxistands will have as many aTaxis as is needed to service all tours originating from that cell. For simplicity, the state was pixelated into square pixels, 0.5 miles on a side (See Figures 2.5 and 2.6).

A coordinate system was then formed in order to reference these pixels; the origin is placed at the intersection of the southernmost and westernmost points of New Jersey. This location was set at (-75.6°E, 38.9°N). Since the units of geographic measurements are longitude and latitude, a conversion system to Cartesian coordinates was then developed to permit access to the pixels via integerized pointers.

Translation of the origin to a point south and west of the New Jersey boundary (-75.6°E, 38.9°N) allows the integer value of any point in the new coordinate system defined as:

\[
X_{Pixel} = \text{Int}(108.907 \times (\text{longitude} + 75.6)) \\
Y_{Pixel} = \text{Int}(138.2 \times (\text{latitude} - 38.9))
\]
where 108.907 and 138.2 convert longitude and latitude units into half mile units.

These coordinates now address some pixel in the array of New Jersey. By locating an aTaxiStand at the center of each pixel, any trip within the pixel will be assumed to be served by the nearest aTaxiStand, which is likely to be the one located at the center of the pixel. Thus, a simple coordinate transformation and integerization converts any latitude and longitude into a pointer to a unique aTaxiStand pixel. For example, a trip end at \((40.050^\circ N, -75.050^\circ E)\) points to the pixel centered at \((158.5, 59.5)\) on the grid. The integer pair \((158, 59)\) can then be used as a pointer to reference any data associated with this pixel. In this way, \((158, 59)\) is the pointer to the aTaxiStand serving all trip ends in the pixel with:

\[-75.0586 \ldots ^\circ < \text{longitude} < -75.0494 \ldots ^\circ\]
\[+40.0433 \ldots ^\circ < \text{latitude} < +40.0505 \ldots ^\circ\]
2.2.2 Trip Definitions

Not all trips that are in the dataset will be serviced by an aTaxi. For instance, using an aTaxi for a trip in which the destination is very close to the origin is not practical. In reality, this trip probably will not occur due to the limited value in waiting for an aTaxi as opposed to walking to the nearby destination. In this analysis, the assumption will be that individuals destined to a location within one pixel of their departure will not take an aTaxi.

A basic categorization of the types of trips yields three primary classes: intrapixel, walk, and aTaxi trips. Intrapixel trips are trips that have a source and destination at the same pixel. Walk trips are the trips whose destination pixel is one pixel away from the source in all directions (N, NE, E, SE, S, SW, W, NW). These trips are NOT serviced by aTaxis due to the short length of the trip and the wait disutility associated with aTaxis. The aTaxi trips are all other trips that have a destination that is more than one pixel away from the origin pixel.

2.2.3 Rideshare Methodology

In this section will present an overview of the rideshare process. Refer to Sections 3.1 and 3.2 for details about implementation of this methodology for the simulation.

The simulations that are conducted will have two parameters that can be altered to assess rideshare properties. This is the departure delay and the circuitry. The departure delay is simply the amount of time an aTaxi will wait at the aTaxiStand after the first traveler has entered the vehicle. One subtlety arises when a traveler arrives at an aTaxiStand at the exact time that an aTaxi is set to
Figure 2.7: The Average Vehicle Occupancy vs. Departure Delay for Essex County.

depart. In this case, the traveler will still be eligible for sharing that ride and the rideshare analysis will be done. As the departure delay increases, there is an increase in shared ridership as shown in Figure 2.7. This figure displays the increase of the average vehicle occupancy of aTaxis departing Essex County as the departure delay increases under two different rideshare environments (CD = 1 and CD = 2).

Circuitry deals with the number of destinations that a vehicle is permitted to visit. A common destination (CD) value of 0 indicates absolutely no ride sharing, a CD value of 1 indicates ridesharing to up to one destination, a CD value of 2 indicates ridesharing to up to two destinations, and so on. As the number of common destinations increase, the rideshare potential should also increase (See Figure 2.8) since there are more opportunities for travelers to enter existing aTaxis instead of entering new ones. The figure confirms that the rideshare potential increases as the maximum number of destinations for each aTaxi increases.

Let us distinguish a traveler as a person who arrives at an aTaxiStand seeking a ride and a passenger as an individual already in an aTaxi waiting to depart. In order for a new traveler to share a ride in an existing aTaxi, the insertion of the new traveler’s destination can not and should not violate certain circuitry rules. In particular, it can not cause the new aTaxi tour to deviate significantly from the direct trip from origin to destination for either the traveler or passengers. If large deviations occurred and it took significantly longer to get to a destination via the aTaxi service, the system will not be utilized by travelers; people would simply use personal cars to make the direct trips to destinations. Hence, certain circuitry criteria must exist to permit rideshare while minimizing the disutility associated with it.

To motivate this criteria, consider the mechanics of the aTaxi system. After the insertion of a traveler’s destination, a new order of destination visits is produced and each passenger has a new
distance to the destination based on the aTaxi path. The simple rule to minimize disutility of the rideshare is that after the insertion of a traveler’s destination in the current queue of destinations, the new distance from origin to destination following the aTaxi path cannot be greater than the circuity threshold percentage of the distance of the direct trip from origin to destination for any party involved. It was determined that 20% is a practical estimate of acceptable circuity for rideshare; hence, to share a ride, an additional trip in an aTaxi cannot increase the distance of any direct trip by more than 20%.

This idea is a bit convoluted but very simple to explain through an example. Take a new traveler that just arrives at an aTaxiStand and is going to the exact same location as a current passenger in an existing aTaxi. This new traveler will share in the existing ride for all analysis where CD > 0 since it is going to a location that is already in the destination queue of that aTaxi.

Now, say a new traveler (Traveler n with destination N) arrives at an aTaxiStand with one waiting aTaxi holding one passenger (Passenger p with destination P). Passenger p’s destination is different from the destination of Traveler n; however, if CD > 2, Traveler n could still share in this ride since the aTaxi can make trips to up to two distinct locations. In order to share the ride, the trip from origin to P to N OR the trip from origin to N to P must not eclipse the distance between origin to N AND origin to P by more than 20%. In order for the trip to be shared in this example, the conditions that must be satisfied are:

\[
\begin{align*}
\min(Distance_{origin-P-N}, Distance_{origin-N-P}) & \leq 1.2 \times Distance_{origin-P} \\
\min(Distance_{origin-P-N}, Distance_{origin-N-P}) & \leq 1.2 \times Distance_{origin-N}
\end{align*}
\]

Although time can be considered an important determining factor for a rideshare, it will produce the exact same results as using distance since constant speed was used for all distance calculations.
Notice in the analysis that both permutations of the trip had to be checked in order to determine whether rides are shared. That is, origin-N-P and origin-P-N must both be computed and checked. The situation could arise in which neither route satisfies the circuity conditions, only one satisfies the circuity conditions, or both do. In the first situation, the trip is not shared. In the second case, the route that satisfies the conditions will be selected.

The interesting case is when both satisfy the circuity condition. In this case, the route that is shorter will be the one that is selected. This simple rule will determine the sharing of rides: when a unique potential destination is introduced by a traveler arriving to an aTaxi (and the CD threshold has not been reached for the aTaxi), if the permutation that generates the shortest cumulative distance traveled by the aTaxi does not satisfy the circuity conditions, then no permutation will. If this shortest permutation does satisfy all the conditions, then the permutation is accepted as the new route and the aTaxi is shared. The beauty of this process is that it can be readily applied to all simulations for all values of CD. The implementation of this heuristic will be explored in Section 3.2.2.

A robust analysis will check all possible orders of destination visits, identify the order that produces the minimum distance aTaxi trip, and then determine whether ridesharing is possible given that order. However, this problem becomes intractable as the number of common destinations increases since this is a computationally intense calculation. Approximations to deal with this situation will be seen in Section 3.2.2.

The departure process for the aTaxis will follow the horizontal elevator process outlined in the beginning of this chapter. First, assume there are infinitely many empty aTaxis at an aTaxiStand. A traveler will arrive at an aTaxiStand at some time, \( oTime \). All the vehicles that were set to depart prior to \( oTime \) will already have departed. There are then several cases to be considered:

1. there are no existing aTaxis with passengers
2. there are existing aTaxis with passengers
   (a) there is rideshare available in some aTaxi
   (b) there is no rideshare available in any aTaxi

In the first case, there are no existing aTaxis with passengers in the aTaxiStand; hence, the traveler will enter an empty aTaxi. There is no lag time between arrival at the station and the entry into an aTaxi; this traveler will enter the aTaxi at \( oTime \). The taxi will then wait the departure delay in order to wait for more potential passengers. If \( \text{common destination} = 0 \), then there will never be any ridesharing so the maximum occupancy for every aTaxi will be 1. If \( \text{common destination} > 0 \), infinite rideshares are permitted but an aTaxi can depart only to up to \( \text{common destination} \) distinct locations. At \( oTime + \text{departure delay} \), this vehicle departs.

In the second case, there are existing aTaxis with passengers in the aTaxiStand that have some departure time equal to the \( oTime + \text{departure delay} \). Therefore, a check of each aTaxi must be made to see whether the traveler’s destination will permit a rideshare. The search will start at the top of the aTaxiStand since the head of the queue contains the aTaxis that have the earliest departure time and it is in the traveler’s interest to depart as early as possible. For each aTaxi, if any of the existing destinations of the passengers within that aTaxi match the traveler’s destination, the ride is shared without it counting towards the \( \text{common destination} \) limit. If no current passenger is heading to the
same location, and the common destination limit is reached, this aTaxi cannot share an additional ride and the next aTaxi is analyzed. If the common destination limit has not been reached, then the circuitry analysis is performed to see whether, after including this traveler’s destination, some permutation of the destinations will satisfy all the circuitry conditions (See Equation 2.1). If no permutation satisfies the circuitry conditions, the ride cannot be shared and the next aTaxi in the queue is analyzed.

As soon as the first rideshare opportunity is found, the traveler will enter this vehicle and share the ride. No changes to the depart time will be made as only the oTime of the first entrant into an aTaxi determines that value. In the case that the entire aTaxi queue has been searched and no rideshare opportunity is found, the traveler will enter a new aTaxi. This traveler will enter this new aTaxi at oTime and wait the departure delay in order to share rides with more potential passengers. The aTaxi departs at depart time.

2.3 Metrics for Analysis

With the trips now available and the aTaxiStands located, analysis can be done for

- trip volumes
- rideshare potential

for different levels of stratification:

- pixel
- county
- state

2.3.1 Volume Assessment

For the trips originating in each county, the trips can be sorted by origination pixel and time in order to analyze the origination data over time for each pixel. This allows origination volume information to be gathered and analyzed. For instance, the pixels with the highest number of potential taxi trips for a given time period can be determined by counting the number of trips coming from the \((y, x)\) coordinate. This can also be done with destination pixels and time to ascertain the total travel behavior of a particular area over time.

2.3.2 Rideshare Assessment

From here, the travel data can be used to simulate the aTaxis and actually make the trips to determine rideshare potential. In this simulation, there are aTaxis at each pixel that wait for passengers to arrive according to departure data in the trip synthesizer.

The first major metric that will be examined is the simple average vehicle occupancy which was seen in Figures 2.7 and 2.8. This is the average number of travelers in each aTaxi:

\[
Simple\ Average\ Vehicle\ Occupancy = \frac{Total\ Number\ of\ Travelers}{Total\ Number\ of\ aTaxis}
\]
This will provide a rough indication of the levels of ridesharing. For instance, if the simple AVO is close to 1, then there is an indication that very little ridesharing occurred since there are almost as many travelers as vehicles departed. On the other hand, higher simple AVO values indicate a higher level of ridesharing since fewer vehicles were needed to satisfy the same level of demand. “Simple AVO” will also be referred to as “AVO” in this thesis.

Although this gives us a fair indication of the shared ridership, a complete analysis would include some measure of distance. In particular, it is valuable to ascertain the number of miles saved as a result of the share.

\[
\text{Miles Saved Ratio} = \frac{\text{Miles Traveled}_{\text{No Share}}}{\text{Miles Traveled}_{\text{Shared}}}
\]

Intuitively, this provides another indication of the value of sharing rides. This ratio compares the effect of the share on the mileage of vehicles. If this ratio is close to 1, then there were very limited shared rides or the sharing of rides did not effectively reduce the total vehicle miles traveled. Conversely, if this ratio is high, there could have been a greater number of shared rides or larger reduction in total vehicle miles traveled.

This measure can also be thought of as a weighted average vehicle occupancy. The weights in this case are the distances traveled. For instance, in simple AVO, the number of travelers is divided by the number of cars. In MSR, the number of miles traveled by the travelers separately is divided by the number of miles the car actually traveled.

The MSR value will typically be lower than the simple AVO. This can be explained by the “pairing phenomenon” in which a long trip is paired with a short trip. A more optimal system would share rides based on how close existing destinations are. Currently, the aTaxi queue at an aTaxiStand is checked starting from the head. As soon as the first rideshare opportunity is found, the ride is shared. This gives rise to situations where many long trips are paired with short trips since this fits the rideshare criteria more readily. A more robust analysis would analyze ALL aTaxis and pair short trips with short trips and long trips with long trips. This would reduce the vehicle miles traveled in the system.

The values of these metrics will vary across different parameter sets (CD and DD). Hence, these metrics provide a way to compare the different aTaxi system environments. In addition, these values can be compared to existing national, state, or regional values to assess the impact that such a system could have in practice. Together, the analysis of the AVO with the Miles Saved Ratio will provide a more holistic assessment of the effects of ridesharing.
Chapter 3

Simulation Implementation

Thus far, a framework for an autonomous taxi system has been set forth. However, simulation of the actual aTaxis is necessary to glean useful information about the properties and effectiveness of the proposed system. This chapter will explain the software that runs the simulations and indicate how to execute it. Over 1,000 lines of rideshare code, data structures, and helper functions are available on https://github.com/jayzach99/Rideshare.git. In addition, there are sample datasets with which analysis can be repeated. This chapter also provides the underlying assumptions and approximations of the rideshare logic along with their justification.

The simulation consists of a set of modules that are executed under different parameters; these will be discussed in the following section. It is also important to note that the dataset that is currently available only contains the trip information for individuals in New Jersey. However, this code can easily be adapted to any type of trip data given that it is in the appropriate format. For instance, a simulation of the entire United States can be done with a dataset that outlines every trip taken by travelers.

3.1 Modules

The main simulation code is written in Java. Java lends itself well to simulations due to the encapsulation provided by its object oriented properties and the robustness of its automatic memory management. This allows the focus to be on logic and development of the modules. The subsections below include a brief description of the modules. Refer to Figure 3.1 to see a visual representation of the connections of the data structures. For more thorough explanations, refer to the readme and the comments in the modules.

3.1.1 aTaxi.java

This class creates an autonomous taxi object. The aTaxi holds the depart_time, which is the origination time of the traveler’s trip plus the set Departure_Delay; the DestinationQueue object, which lists all the destinations; cumDistance, which denotes the total distance to be traveled by that car based on trips in its DestinationQueue; and next, which is the pointer to the next aTaxi in the queue.
This aTaxi object can be used fairly generally. However, the functionality to be part of a list (queue, stack, etc.) exists since one aTaxi can point to another aTaxi via the next variable. This is logically feasible in the implementation of the simulation since a real aTaxi enters a queue at an aTaxiStand and then departs. Likewise, this aTaxi enters a queue in the DepartureQueue.

Note that to get destination information, we must interface the DestinationQueue object. This interfacing functionality is built into the aTaxi object. That is, if any destination information needs to be pulled or any actions taken on the DestinationQueue, it can be done through the functions available in aTaxi.

The aTaxi contains a method public void insertandReorderTrip that inserts a new destination into this aTaxi and then reorders the trip according to the argument newOrder. If a ridesharing opportunity is discovered in a particular aTaxi and the destination order needs to be adjusted, this method provides that functionality.

- Execution: Not Executed.
- Dependencies: DestinationQueue.java

### 3.1.2 DestinationQueue.java

A DestinationQueue is an object that holds the destination information of a particular aTaxi. It is effectively a private class for an aTaxi that has been extracted to its own module. A new DestinationQueue is instantiated with every new aTaxi object created and the aTaxi module will make accesses to DestinationQueue to get the destination and passenger information.
The DestinationQueue object contains the num_destinations, which is the number of destinations that an aTaxi has; num_passengersInCar, which is the number of travelers in the aTaxi (more than one person can go to a particular destination so there can be more travelers in the car than destinations); Node first; and the Node last.

A Node is a private class that holds 1) an array item which contains a destination pixel pair and the number of travelers going to that pixel and 2) a Node next which is the pointer to the next Node in the list. Essentially, the DestinationQueue object has a list of Nodes each of which contain information about a specific travel destination and has a pointer to the next destination. The order of the Nodes is the order in which the aTaxi will visit the destinations so this is frequently changed by calls in higher level classes when rideshare possibility is found.

- Execution: Not Executed.
- Dependencies: No Dependencies.

3.1.3 DepartureQueue.java

A DepartureQueue is an object that essentially represents the aTaxiStand. It is used as the pointer to the first aTaxi (which can subsequently access all the aTaxis at a certain pixel) and to hold metadata about the aTaxi data. The DepartureQueue object holds the first and last aTaxis in the queue in the variables first and last. It also holds the number of aTaxis in the DepartureQueue in number_vehicles.

In order to utilize this object, an aTaxi object must be created and then inserted into the queue via the enqueue method. The DepartureQueue implements Iterable<aTaxi> so accesses to the aTaxis can be made by iterating the DepartureQueue. This is possible because each aTaxi points to the next aTaxi in the queue.

- Execution: Not Executed.
- Dependencies: aTaxi.java

3.1.4 RideShare.java

The RideShare class is the brains behind the ridesharing analysis. It contains two public static methods that will check for all rideshare possibilities when utilized in the appropriate order.

The first method checkRideshareSame is used to identify whether the current aTaxi is already visiting the destination that our traveler has. If so, it will return a boolean as the first item in an array indicating that there is a rideshare. If not, it will return a boolean as the first item in an array and the subsequent elements will be populated by the y coordinate and x coordinate of the destination pixels. This array method of returning is effective since there are essentially two unique items packaged into one array and returned to the calling class. The first return value, which is the first element in the array, is a boolean of the rideshare indication. The other return, which are the remaining elements of the array, is the series of pixel coordinates. The secondary return will be used to send destination information to other ridesharing code.

The second method, checkRideshareCircuitry, is used to determine whether adding another destination to the current set of destinations in an aTaxi's DestinationQueue would provide a
rideshare opportunity. The destinations present in the vehicle, the new destination of the traveler, and the all destination orders that should be checked are passed into the function as arguments. This function then proceeds through every order that was passed to it in order to find the shortest path order that fits the circuitry criteria. If none meets the criteria, there is no ridesharing opportunity and a boolean is returned as the first element in an array indicating this. If an order does match the circuitry criteria, then there is a rideshare opportunity. A boolean indicating this is returned as the first element of an array and the subsequent elements of the array indicate the order of the destination that makes this possible. This will later be used to rearrange the Nodes of the DestinationQueue.

- Execution: Not Executed.
- Dependencies: DestinationQueue.java, HelperFunctions.java

### 3.1.5 RideShareAnalysis.java

RideShareAnalysis is the module that stitches all the pieces together and controls the flow of the simulation. The fundamental framework is outlined in Section 2.2. Specifically, the ridesharing methodology and logic are detailed in Section 2.2.3.

This class takes in four arguments.

1. The filename with trips data in it. This must have 11 columns of integers separated by commas and a newline character at the end of each line. Refer to 3.3.1 for more details on input files.

2. The departure delay time in minutes.

3. The common destination that indicates the maximum number of locations that one aTaxi can visit.

4. The number of rows in the file with trip data in it. This initialization is necessary to read in data. The values are available in a summary table within SummaryTable2.xlsx. This can be found on the github.

There are two streams of output. The module outputs results using the standard output stream and also outputs the current pixel of analysis in the standard error stream. The former should be redirected to a file and the latter can simply remain directed to the console. The standard error stream was primarily used for debugging purposes but is quite useful when determining the progression of the code.

RideShareAnalysis then goes through pixel by pixel and manages the aTaxi departures along with ridesharing analysis. In order to do this pixel by pixel, the data is sorted by the appropriate columns (See Section 3.3.1) to ensure that trips originating from the same pixel are contiguous in memory.

The module then creates a DepartureQueue (which represents the aTaxiStand) and essentially follows the rideshare methodology outlined in Section 2.2.3. The pseudocode runs as follows:

- Depart all aTaxi’s whose depart_time is strictly below current time increment.

- If the DepartureQueue is empty, add new aTaxi.
• Else create boolean “shared”, set to false, and see whether rideshare opportunities exist.
  
  – if CD != 0
    
    * For each aTaxi
      
      · if destination is found in the aTaxi’s DestinationQueue, set “shared” to true and break out of loop.
      
      · elseif number_destinations < CD, do circuitry analysis to see whether there is rideshare opportunity. If criteria is met, set “shared” to true.

• If boolean “shared” is still false, then the DepartureQueue is not empty but there is no shared ridership potential. Create a new aTaxi.

First RideShareAnalysis departs all aTaxis that have a depart_time strictly below the current time increment. If the DepartureQueue is empty, then a new aTaxi is added for that traveler. Upon the creation of a new aTaxi, the depart_time (Departure_Delay plus current time increment) is set. The traveler’s destination information is added to the aTaxi as well.

If the DepartureQueue is not empty, the RideShareAnalysis module checks for shared ridership. If CD = 0, there can be none so a new aTaxi is created. If CD > 1, checks must be made on the existing aTaxis to see whether there is a rideshare opportunity. The aTaxi at the head (the earliest to depart) is checked first. First, a check is done to determine whether any destination within the aTaxi matches the traveler’s destination. If so, the ride is shared. If not and the DestinationQueue of the aTaxi has as many num_destinations as CD permits, the ride cannot be shared and a new aTaxi is created. However, if num_destinations is less than CD, rideshare opportunities through circuitry are checked. If a rideshare opportunity exists by adding the new traveler’s destination as an additional leg of the aTaxi tour, the trip is inserted into the DestinationQueue and the DestinationQueue is reordered to reflect the optimal sequence of trips that minimizes distance while permitting that rideshare. If rideshare opportunities do not exist, then a new aTaxi is created.

If the number of trips in a particular pixel are exhausted and there are still aTaxis in the DepartureQueue, depart them all. RideShareAnalysis then proceeds to the next pixel until all pixels are analyzed.

• Execution1: java (-Xmx500m) RideShareAnalysis [input_filename] [DD] [CD] [number_rows_in_file] > [output_filename].csv

• Dependencies: DepartureQueue.java, aTaxi.java, RideShare.java, ManageFiles.java, Permutation.java, HelperFunctions.java

3.1.6 Other Modules

There are some other key modules that are necessary for the functioning of this program. These must be present for execution of the primary modules.

---

1 Arguments in parenthesis are optional. Arguments in the square brackets are required but must be supplied. Do not include paranthesis or brackets when executing. The [number_rows_in_file] argument can be found in SummaryTable2.xlsx on the github.
Permutation.java

Permutation.java returns all orders that need to be checked for rideshare opportunities. The first function, getOrders, returns all permutations that need to be checked. There is also the getLinearOrders function that returns a subset of orders that can be significantly smaller than the number of orders that getOrders. This getLinearOrders function is the implementation of the circuity approximation that will be discussed in Section 3.2.2.

- Execution: Not Executed.
- Dependencies: StringBuilder.java (within java.util)

HelperFunctions.java

HelperFunctions.java is simply a repository of functions that are called repeatedly in other classes. The primary public static function is getManhatDist which returns the Manhattan distance between two pixels.

- Execution: Not Executed.
- Dependencies: No Dependencies.

ManageFiles.java

ManageFiles.java manages the input. The input is a trips CSV file with 11 columns of integers (See Section 3.3.1). This module essentially parses the input, reorders by the appropriate columns, and stores the trips into array for use by caller class. There is also the unused functionality of writing to CSV. This is from previous iterations of the, but it is fully functional.

- Execution: Not Executed.
- Dependencies: Arrays.java (within java.util), BufferedReader.java, StringTokenizer.java, FileReader.java, BufferedWriter.java (within java.io)

3.2 Design Decisions

The performance and framework of the simulation necessitated some early fundamental decisions that shaped how the development was conducted. Two prime examples of this can be found in dealing with data structure decisions and the complexity of the circuity analysis.

3.2.1 Time versus Space

A general approach to the software was creating a scalable infrastructure that would adapt well to large data sets. The fundamental computational complexity tradeoff between storage space and computational time [24] exists in this simulation and one significant implementation decision that this tradeoff affects is the data structures used to store data. Effectively there are two data structure options: arrays and lists.
One of the key implementation decisions was to use lists. In particular, queues are utilized since they adhere to the First In First Out behavior also seen in aTaxiStands. Although arrays have constant access time among other advantages, they are deficient for this simulation since a size must be set during initialization. Therefore, to ensure that the simulation functions for all datasets and to accommodate the worst case scenario, an arbitrarily large array must be set. Doing this across all the potential trips and pixels could exhaust memory rather quickly and cause lagged speeds in processing. Although dynamically resizing arrays are also an option, the complexity introduced by their utilization is not appealing. Elements of a queue, on the other hand, are created as they are needed so there is no waste in space. As the dataset scales, this saving of space becomes more significant than the extra time needed to do certain computations on the queue.

3.2.2 Circuity Approximation

As mentioned in Section 2.2.3, this aTaxi system simulation permits rideshare analysis under different parameters. This includes specifying the number of destinations to which an aTaxi can travel with its passengers (or permitting circuity). A CD or common destination value of 0 indicates no ride share and a common destination value of \( n \) indicates a ridesharing to up to \( n \) distinct locations.

When checking for rideshare, every order must be assessed to see whether circuity conditions hold. For lower values of this common destination parameter, identifying every permutation of destinations is plausible. However, as the number of locations to which a vehicle can travel increases, the number of orders that need to be checked increases extremely fast. This essentially boils down to the Traveling Salesman Problem. In the Traveling Salesman Problem, a salesman must find the optimal route across a set of cities by only visiting each city only once. In computer science, TSP is in a category problems classified as NP-complete. In NP-complete problems, a solution can be verified rather quickly but attaining the solution is computationally difficult [9].

Brute force approaches yield a \( O(N!) \) runtime which can dominate the computation in problems. For instance, say \( CD = 15 \) and a traveler is attempting to share a ride in a vehicle that holds 14 people going to 14 different locations. In order to determine which path yields the shortest path there must be \( 15! \approx 1.3 \) trillion orders checked in order to establish whether there is rideshare potential. This computation will dominate the runtime and make the simulation intractable. In practice, computations become very lagged at even relatively low CD values such as 8 or 9.

The common approach to the TSP problem is approximation with techniques such as simulated annealing, minimum spanning trees, and heuristic measures [11]. However, many of these methods have tunable parameters or need a large number of iterations to be accurate; this requires extra time and space to compute. Since the TSP problem will be repeated for every rideshare check, these approximations will not provide the performance that is sought. Instead, a holistic perspective was taken to identify the fundamental characteristics of the aTaxi framework that would allow an even simpler approach to the approximation.

First core features of the system were listed. In this framework, a traveler seeking a rideshare in an aTaxi never replaces a passenger already in the aTaxi. This aligns well with a real system since once an individual enters an aTaxi, it would be a hassle to reassign that individual. Even if it is optimal to replace and reassign an existing passenger, this will not happen. Also, when making rideshare checks, if the circuity threshold criteria is violated for the traveler or any current passenger, then that traveler will not share in the ride. Moreover, the current order of the trips is
the shortest path permutation. Because the shortest path is always the current order, it is likely that the structure of the current set of trips will hold as the number of destinations increase.

As more destinations are added to the aTaxi, checking the set of all permutations will consume time. In order to mitigate this, a subset of orders in which there is the highest likelihood of rideshare must be selected. Looking at example tours within the framework, it can be posited that the structure of the aTaxi trip visits to the destinations should change less as the number of destinations increases. As more rides are added (for which CD criteria are met), a complete reordering will not be frequent since it will be more likely to violate a circuity criteria for a current passenger. Hence, instead of checking all permutations, it is more strategic to identify and utilize permutations that preserve the relative order of the trips and thereby preserve the tour structure. In this case, after many trips have been added to the aTaxi, an additional trip can only be added if it can be interpolated into the current tour’s existing path.

For instance, Figure 3.2 shows a sample situation. Point A is the origin. Points B and C are existing destinations in the aTaxi. The shortest existing path permutation was found to be ABC. Point D is the new traveler’s destination. If we were allowing all permutations to be checked to see whether traveler D could share in this ride, the six orders, ABCD, ABDC, ACBD, ACDB, ADBC, and ADCB must be analyzed for ridesharing opportunities. However, some orders, such as ACBD, seem completely nonsensical while other orders, such as ABDC, seem completely feasible. If the restriction of preserving relative orders is placed, then the set of orders that needs to be checked is reduced to ABCD, ADBC, ABDC. These also correspond to the trip orders that are visually likely to engage in ridesharing. Although checking three additional orders is not computationally costly in this small case, it could become very costly at larger values of CD. For instance, at CD = 10, this approximation checks 10 orders rather than 3.5 million orders.

In order to assess the validity of this approximation, trials were run varying certain parameters. First, a permutation threshold (LO) is set; this is the value up to which all permutations are used. Any value higher than LO will be done in a linear (or relative order preserving) fashion. Then AVO analysis across different CD values were done. The expectation is to see significant deviation from the actual AVO figures where the LO value is significantly lower than the CD value. In other words, in cases in which a greater number of order preserving permutations is used, we expect to deviate from the correct solution (which going through all permutations guarantees to find). The magnitude of this deviation will allow us to assess whether the approximation is valid; small deviation between
the approximate and actual is favorable. Figure 3.3 shows that this deviation very small. Take CD = 8 for instance. At LO = 0, AVO is 2.024. The true ride share potential at CD = 8 is captured when LO = 8 and is 2.033. This is less than a 0.5% deviation! Figure 3.4 shows just how good the approximation is.

The higher that LO is set, the better the approximation will be. After reaching a certain threshold, relative orders of the destinations in an aTaxi will only minutely change. Hence, once we get to a CD value where performance begins to lag, we can take permutations that preserve relative orders for any higher value of CD instead of checking all permutations.

In practice, a DD over 5 and a CD over 5 would probably not be feasible for autonomous cars. However, if this simulation were to include other autonomous vehicles such as vans or buses, the functionality to perform analysis at higher CD levels is necessary. Hence, this CD parameter should be generalized to be arbitrarily large. For the simulations in this thesis, the permutation threshold is set at 7.

### 3.3 Running the Simulation

This section will detail how to run the simulations. In particular, the input files, the execution, and the output files will be explained.

#### 3.3.1 Input Files

The input files hold the trip data of each traveler in New Jersey. Although this input file is derived from the output of the trip synthesizer, the direct trip synthesizer output cannot be fed into the simulation. The output from the synthesizer produces excessive detail about each individual, their attributes, and their trip details for the simulation. The aTaxi framework only needs trips times and coordinates to determine the rideshare potential. Hence, the extraneous data must be parsed out of the synthesizer generated dataset. In addition, numerous painstaking adjustments must be made to clean up the data. Among those include removing extreme outliers, eliminating individuals that did not make any trips, checking boundaries to ensure each pixel is mapped only to one county, and correcting select distributions. This is processing, space, and time intensive so automation is key. The code to do this can be found on the github under “Database Items.”

Once the data is cleaned, the input files are completed. There is one master file that contains all

![AVO (DD = 2) for Varying CD and Permutation Thresholds](image)

Figure 3.3: Tabular Comparison of AVO for Different Permutation Thresholds. Note that AVO for a particular CD value at LO = 0 (where only relative order preserving orders are allowed) and at LO = 8 (where all permutations are checked for the first 8 riders) do not vary significantly.
Figure 3.4: Graphical Comparison of AVO for Different Permutation Thresholds. At low CDs, the AVO values for low permutation thresholds do not deviate at all from the true AVO. As the CD value gets larger, the deviation only slightly increases. Even at CD = 8, the difference between LO = 0 and LO = 8 is barely noticeable. This indicates that the order preserving permutation selection will be a very good approximation.

trips in New Jersey and there are smaller files in which the trips are separated by county of origin. All input files contain eleven integer columns with NO headers:

1. Pointer: The pointer back to the row in the original trips synthesizer output to which this trip correlates. When used in conjunction with the ABC, it produces a unique pointer to one of the 8.7 million travelers

2. ABC: The home county of the individuals

3. OCounty: The county from which this trip leg originates

4. OY: The y coordinate of the trip origination pixel

5. OX: The x coordinate of the trip origination pixel

6. OTime: The time the traveler desires to engage in a trip; Alternatively, the arrival time of individuals to a TaxiStand

7. DCounty: The county to which this trip leg is destined

8. DY: The y coordinate of the trip destination pixel

9. DX: The x coordinate of the trip destination pixel

10. DTime: The arrival time of traveler

11. Manhat_Dist: The Manhattan distance between origin and destination
Note that an individual’s home county (ABC) is a distinct concept from the county from which his trips legs originate (oCounty). Take an individual who lives in Mercer County, commutes to Bergen County and returns to Mercer County in the evening. There are two legs to this tour; the first trip leg will have ABC = Mercer and OCounty = Mercer while the second leg will have ABC = Mercer and OCounty = Bergen. The ABC is preserved and carried in this dataset because the Pointer and ABC together point to a unique traveler from the synthesizer.

3.3.2 Execution

In order to execute the rideshare simulation, all the compiled java Class files of the files listed in Section 3.3.1 must be in one directory. The file (input_filename) that contains the trip data that will be analyzed must be in this directory as well. For very large datasets (e.g. statewide data), use -Xmx500m to request more memory. The DD and CD values correspond to the departure delay and common destination values. These are the parameters that are toggled to assess system performance. The number_rows_in_file can be obtained from the SummaryTable2.xlsx spreadsheet and corresponds to the number of trips in the file. Finally, the output should be redirected to a file so that it can be analyzed. The output will be in CSV format. To run the simulation, run in the command prompt:

```
java (-Xmx500m) RideShareAnalysis [input_filename] [DD] [CD] [number_rows_in_file]
> [output_filename].csv
```

3.3.3 Outputs

There are two active streams of output. The standard output stream, which contains the raw analysis output, will be redirected to [output_filename].csv. The standard error output is directed to the console and will display a running count of the number of pixels that have been assessed. The output file contains the relevant departure and rideshare information. The number of columns in the output file varies with the value of CD. The 8 CD-independent columns found in every output file is:

1. OY: The y coordinate of the origination pixel
2. OX: The x coordinate of the origination pixel
3. DepartTime: The time the aTaxi departs from aTaxiStand
4. DepartOccupancy: The number of passengers in vehicle at depart time
5. NumberDestinations: The number of destinations that this aTaxi will visit
6. CumDistance: The cumulative distance that this aTaxi will travel along all trip legs
7. DESTY_1: The y coordinate of first destination pixel
8. DESTX_1: The x coordinate of first destination pixel

Since all aTaxi tours must have one origin and at least one destination, these columns will never be null. The CD-dependent columns correspond to the coordinates of the destination pixels when
additional destinations are permitted. These are only active when CD > 1. For instance, when CD = 3, 4 additional columns become activated corresponding to DESTY\_2, DESTX\_2, DESTY\_3, and DESTX\_3. However, CD = 3 does not necessitate that each aTaxi have 3 destinations; it simply permits up to 3. Hence, these additional columns can have null values if the number of destinations in an aTaxi does not add up to CD.
Chapter 4

Simulations, Results, and Analysis

The previous chapters discussed the dataset, the framework for an aTaxiStand system across New Jersey, and the implementation of the simulation that can be used to assess the ridesharing opportunities. A case study applying these methods to Mercer County can be conducted to analyze the dataset and examine the simulation. These methods can then be applied to the entire state of New Jersey so statewide properties can be ascertained. Trip data in the analysis only deals with the aTaxi trips since the goal is to assess properties of the aTaxi system. Walk and intrapixel trips are excluded since they do not inform the aTaxi system at all. The definition of the types of trips can be found in Section 2.2.2.

4.1 Case Study: Mercer County

![Map of New Jersey with Mercer County highlighted]

Figure 4.1: Mercer County in New Jersey.
4.1.1 Mercer County Overview

Mercer County, the “capital county,” is situated halfway in between New York and Philadelphia and is home to 366,513 people. Due to its excellent infrastructure and roadways, key corridors along the New Jersey Turnpike are some of the most robust commerce regions in New Jersey. In addition, Mercer County holds a highly skilled labor force and houses well known educational institutions such as Princeton University and Rider University. Mercer County has many parks, public golf courses, and other recreational sites as well [14].

Mercer County’s eclectic mix of residential, industrial, commerce, and recreational regions makes it a prime county to analyze. Figure 4.2 presents some high level statistics about Mercer county.

<table>
<thead>
<tr>
<th>Features of Mercer County</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population</td>
</tr>
<tr>
<td>Land Area (sq. miles)</td>
</tr>
<tr>
<td>Total Originating Trips</td>
</tr>
<tr>
<td>Trip Producing Pixels</td>
</tr>
<tr>
<td>Trip Producing Land Area (sq. miles)</td>
</tr>
<tr>
<td>Open Space</td>
</tr>
</tbody>
</table>

Figure 4.2: Mercer County Statistics

4.1.2 Mercer County Volume Analysis

Before the rideshare analysis is done, the dataset should be analyzed to obtain a better sense of the trips patterns and volume characteristics of the county at hand. The simulation produced 1,314,115 trips originating from Mercer County. It is likely that the city centers produce most of the trips since all but 5.1% of the land is used for industrial, commercial, and residential purposes.

To obtain a sense of time and location of the trips, cumulative trip distributions based on certain factors should be assessed. Specifically, the number of trips over time and the number of trips occurring by the order of highest ranked volume pixels should be observed. This will indicate how traveling is done over time and the traveling occurs. These are the most insightful distributions since they can be used to inform the beginnings of an aTaxi system. For instance, the distribution by time could indicate what times of the day aTaxi services would be most useful and the highest ranked volume pixel distribution could indicate which aTaxiStands should be opened at the initial phases of the system development.

Figure 4.3 shows the cumulative distribution of trips in Mercer county over the 86,400 seconds of the day. The cumulative distribution graph displays the total percentage of trips (y-axis) that have been completed by some time (x-axis). The slope of the cumulative distribution graph indicates the rate at which the volume of trips originate at a time period. For instance, in Figure 4.3, there is a jump at about 25,000 seconds (about 7am) to 28,000 seconds (about 8am). This corresponds to the morning rush hour of school and work trips. This suggests that an aTaxi service during the 6am to 9am rush period could have congestion reducing benefits since travel demand is high during that time period which, in turn, increases the likelihood of a rideshare arising. Smaller successive jumps can be seen in the period between 58,000 seconds (about 4pm) and 70,000 seconds (about 7pm).
These trips corresponds to the waves of travelers who are leaving school and work during different times of afternoon and evening.

It is also useful to understand the sources of the trips. There are 852 trip producing pixels in
Mercer County. Figure 4.4 depicts the cumulative distribution of trips by rank of the highest volume pixel. That is, pixels are ordered by the number of trips they produce and placed on the x axis. The cumulative percentage of trips is placed along the y axis.

This graph suggests that the trip originations in Mercer County are very dense since relatively few pixels are responsible for most of the trip production in the county. About 50% of trips in Mercer County come from the top 72 (~8.5%) trip producing pixels and about 90% come from the top 40% pixels. Figure 4.5 provides a visualization of the trips that sheds light on this finding. The figure shows a skyscraper plot of the 100 highest trip origination volume pixels. The height of the skyscraper bar corresponds to the number of trips generated by the pixel that it covers. It is apparent that the two hubs for trips are the bustling city centers of Trenton and Princeton. Set along I-95, these are commercial, educational, and residential hubs that should have many trip originations.

Figure 4.5: Mercer County Top 100 Trip Producing Pixels. Note that the area surrounding Princeton and Trenton are the regions with the most active pixels.

Figure 4.6 shows the pixel region that produces the most trips in Mercer County. The pixels cover part of Princeton Borough and originate 50,000 trips combined. The production of a great quantity of trips occurs due to the activities and sites found in this region. Princeton University, which has about 7,000 students and 6,000 faculty and staff [21], Princeton High School, and Princeton Public Library all contribute to the number of trip originations. Moreover, the Princeton “Dinky” Station
produces a substantial number of trips since it is the closest station to Princeton residential areas. Other features of these pixels that explain the high volumes include Palmer Stadium, Jadwin Gym, and the other Princeton University facilities as well as the numerous recreational and occupational sites along Nassau Street. The density of all these factors within the pixel region is the reason that it is the highest origination volume pixel region in Mercer County.

Figure 4.6: Mercer County Highest Volume Pixel. This encompasses Princeton University and the nearby town.

4.1.3 Mercer County Rideshare Analysis

With a better sense of the data and trip patterns, rideshare analysis can be conducted. Essentially, congestion reduction will occur if there are fewer vehicles on the road. It has been posited that one way to reduce the number of vehicles on the road is by sharing rides because the capacity of a vehicle is better utilized. As mentioned in Section 2.3.2, there are two major metrics for assessing rideshare potential and feasibility — average vehicle occupancy and the miles saved. The definitions will be presented again in this section.
Average vehicle occupancy is the number of individuals per car departed in an defined area such as a state or county. In other words:

$$\text{Simple Average Vehicle Occupancy} = \frac{\text{Total Number of Travelers}}{\text{Total Number of aTaxis}}$$

The average vehicle occupancy essentially gives a measure of the capacity utilization of a car. This permits comparison of different rideshare environments (different DD and CD values). In checking AVO, the higher values should be found when the departure delay and common destination parameters are increased.

To put AVO values into realistic perspective, it should be compared against current averages. The National Household Travel Survey publishes a country wide assessment that reports average vehicle occupancy as well as many other statistics [30]. Figure 4.7 specifies the average vehicle occupancies for different purposes since 1977. The value used to compare against the simulation output will be “All Purposes” since the synthesizer simulates all different types of rides. In addition, since the 2010 Census informed the synthesizer, the 2009 value will be used. Hence, 1.67 passengers per car will be the AVO figure that the simulation must eclipse to successfully reduce the current average number of vehicles on the road.

The miles saved ratio will ascertain the value of ridesharing on reducing vehicle miles traveled under different parameter sets. It compares the miles traveled by a system with no ridesharing (CD = 0) to the miles traveled by all the aTaxis in the system when ridesharing is permitted (CD > 0). The miles saved ratio is calculated as:

$$\text{Miles Saved Ratio} = \frac{\text{Miles Traveled}_{\text{No Share}}}{\text{Miles Traveled}_{\text{Shared}}}$$

The analysis henceforth will only assess practical parameters, particularly the range from 0 to 5 for both DD and CD. These values will impose a realistic measures to the simulation. Waiting longer than 5 minutes to depart will impose great disutility to impatient passengers. Moreover, passengers are less likely to use the system if there are more than 5 locations that could potentially be visited; in this case, another public mode of transportation would provide the same utility.

Finally, much of the insights presented in the case study will also apply to the state of New Jersey and will not be repeated in the New Jersey analysis (Section 4.2.3).
Mercer County Average Vehicle Occupancy

Simulations for Mercer County were conducted after extracting the trips originating at the pixels belonging to Mercer County. In particular, the system was simulated for CD = 0, 1, 2, 3, 4, 5 and DD = 0, 1, 2, 3, 4, 5. Figures 4.8 and 4.9 graphically display the results of the simulation. For actual values, see Figures 7.1 and 7.2 in the Appendix.

The graphs show that as CD and DD increase, the AVO values increase. This matches the expectation since waiting longer for a rideshare or allowing a vehicle to visit more than one location could only allow more ridesharing to occur. It is also apparent that CD = 0 produces an AVO of 1 regardless of the DD. This was a system design and validates the functionality of the simulation.
The largest increases in AVO due to an increase in DD occur when DD goes from 0 to 1. This should be expected because much of the rideshare pairing occurs between long trips and short trips (the “pairing phenomenon”) since these tour types are more likely to satisfy circuity conditions. Since trips arrive very frequently, a short trip can almost immediately be paired with a longer trip and much ridesharing can be captured early. After the first minute, a greater number of passengers with varied destinations arrive to a greater number of waiting vehicles; the later minutes will allow more of the unfilled spaces in aTaxis with similar destinations to be filled. However, the “easy” pairing of a long trip and short trip will already have been exhausted by this time. This is also indicated by the diminishing return on ridesharing as DD increases. This shows that the 1st minute is the most productive with ridesharing rates waning as departure delay is increased.

The largest increase in AVO due to CD increases occurs as CD goes from 1 to 2. This is logical because many short trips pair up with long trips due to the circuity conditions. Take, for instance, two trips, one with a destination close to the origin and one with a destination far from the origin. If a rideshare analysis were being conducted, it would likely pass the circuity test — the aTaxi will make the short trip first and then head to the second passenger’s destination. The roundabout caused by the shorter trip would be so relatively short compared to the second traveler’s direct trip that the ride would be shared. In this way, the highest jump can be expected as the number of distinction locations increases from 1 to 2.

These AVO values compared to those of the state of New Jersey (See Figure 7.3 in the Appendix) are generally lower across all parameter sets. For instance, the AVO at (DD, CD) = (1,3) is only 1.86 in Mercer County and 2.06 in New Jersey. This is a surprising outcome since Mercer County’s location near Philadelphia and its characteristics would suggest that it would have greater ridesharing. However, as seen in the volume analysis, most trips are concentrated around the cities of Princeton and Trenton. Although rideshare is very possible here, trips from other areas are more sporadic and leads to fewer riders per vehicle. This will lower the AVO. Regions with more of a dispersion of trip volumes will generally see higher AVOs. Since trips are less densely distributed throughout the state of New Jersey than in Mercer County itself, there is greater AVO over the entire state.

The system’s AVO does very well in comparison to the national AVO figures under the appropriate parameters. Hence, instituting a simple system with parameters as low as CD = 2 and DD = 2 will eclipse the average national AVO of 1.67. However, in order to really be of true value, the AVO needs to be well above the national value. AVO values of 2 and above indicate a 20% increase over national figures. This can be achieved at reasonable (DD, CD) levels such as (2,3) or (3,2). See Figure 7.1 in the appendix for the full range of values.

Mercer County Miles Saved Ratio

The miles saved ratio graphs (Figures 4.10 and 4.11) have a striking resemblance to the shapes of the AVO graphs. This is expected since the miles saved ratio can be viewed as a weighted AVO in which the weights are simply the distances. For instance, in the calculation, instead of dividing the number of travelers by the total number of aTaxi as in simple AVO, the number of miles that each individual traveler demanded is divided by the actual miles driven by the aTaxi in MSR.

An MSR value of 2 indicates that the total miles traveled by the aTaxi is half the amount that would have been traveled in an environment where each individual rides in a separate vehicle. Just
as AVO indicates higher capacity utilization, high values of MSR indicate significant vehicle mile savings associated with ridesharing. It is also interesting to note that MSR values are smaller than AVO values. This is partially explained by the pairing phenomenon, the suboptimal occurrence in which long trips are paired with short trips in the rideshare simulation. A more robust system would pair long trips with long trips and short trips with short trip to minimize total vehicle miles.

Compared to New Jersey as a whole, Mercer County has slightly higher MSR values. This means that due to rideshare, fewer vehicle miles were driven in Mercer County compared to NJ as a whole. This is an interesting result because the simple AVO value was actually lower than NJ state average. This indicates that although fewer people share rides in Mercer County compared to NJ state, the
shared rides actually go farther distances compared to NJ rides. This is a significant finding because it displays the true efficiency of the rideshare system.

Even at reasonable (DD, CD) pairs, trip miles can be halved. For instance at (2, 3), the system produces a MSR ratio of 2.14. The maximum MSR achieved in our feasible parameter range is 3.07 at (5, 5). In this situation, the vehicle miles traveled can be cut by two-thirds! See Figure 7.2 in the Appendix for a complete table of values.

**Assessment**

There is significant rideshare potential in Mercer County along with the opportunity to reduce vehicle miles traveled. This analysis of Mercer County can begin to suggest optimal parameter sets for aTaxiStand operations. Since both metrics experience diminishing returns as the values increase, it is important to select parameters (CD and DD) that extract as much value as possible where there are high rates (indicated by the slope of the line) and then judiciously proceed in tuning the parameters once the diminishing of the rates begins to occur. The parameter set should also produce values that exceed the national values in order make the system more feasible.

DD values seem to experience diminishing returns after about 1 minute and CD values diminish after 2 destinations. The parameter set (DD, CD) = (1, 2) produces an MSR of 1.63 and an AVO value of 1.77 which eclipses the national AVO of 1.67. This reveals that there are fewer vehicles on the road and significant miles saved due to the rideshare. Hence, (DD, CD) = (1, 2) makes a good baseline parameter set. System improvements can be made from this baseline by appropriately tuning the parameters while avoiding actions that would impose disutility to the users of the system.
4.2 New Jersey Results

4.2.1 New Jersey Overview

New Jersey’s features make it an excellent state on which a simulation such as this can be tested and implemented. For instance, New Jersey is geographically varied with a mixture of open space and utilized space. New Jersey consists of mountainous regions in the north, coastal cities along the shoreline, and pine barrens in the southeast. Open space constitutes about 25% of the state. Hence, this simulation can be checked for robustness across all major types of geography. Moreover, New Jersey’s location and infrastructure makes it quite accessible. It lies between New York and Philadelphia and proximity to these city centers induces greater traveling throughout New Jersey by both its residents and out-of-state residents. The expansiveness of the roadways such as I-95, Route 1, and the Garden State Parkway along provides accessibility throughout the state. This allows more land to be utilized efficiently and gives rise to prime residential and recreational areas throughout the state. New Jersey is economically, educationally, and recreationally active; this activity translates into trips.

New Jersey’s qualities are also favorable in terms of running the simulation. The sample size of 8.7 million people producing 30 million aTaxi trips across 21,586 pixels is large enough to reduce noise in the synthesizer output but small enough to run quickly and efficiently. The following sections will present volume analysis and rideshare potential analysis for the state of NJ.

<table>
<thead>
<tr>
<th>Features of New Jersey</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Population</td>
<td>8,791,894</td>
</tr>
<tr>
<td>Land Area (sq. miles)</td>
<td>7,354</td>
</tr>
<tr>
<td>Total Originating Trips</td>
<td>30,220,968</td>
</tr>
<tr>
<td>Trip Producing Pixels</td>
<td>21,586</td>
</tr>
<tr>
<td>Trip Producing Land Area (sq. miles)</td>
<td>5,397</td>
</tr>
<tr>
<td>Open Space</td>
<td>26.6%</td>
</tr>
<tr>
<td>Shortest Trip Length (miles)</td>
<td>1</td>
</tr>
<tr>
<td>Longest Trip Length (miles)</td>
<td>202.5</td>
</tr>
<tr>
<td>Median Trip Length (miles)</td>
<td>12.5</td>
</tr>
<tr>
<td>Average Trip Length (miles)</td>
<td>16.9</td>
</tr>
</tbody>
</table>

Figure 4.12: NJ Statewide Statistics.

4.2.2 New Jersey Volume Analysis

In the simulation, New Jersey has 30,220,968 trips that will be serviced by the aTaxi system. These trips occur over 21,586 pixels of the grid over the 86,400 seconds in a day. The cumulative distributions represented in Figures 4.13, 4.15, and 4.14 offer the most insight about New Jersey travel behavior. These travel behaviors can then be used to inform the beginnings of an aTaxi system.

Figure 4.13 displays the cumulative trip distribution over time. Starting at midnight, about half the trips are completed in the first 16 hours. Hence, the later hours generate a higher concentration of trips. The first 25,000 seconds (7 hours) from midnight does not generate many trips. However, 70% of all trips are taken in the 12 hours spanning 7am and 7pm. After 7pm, there is a decline in the number of trips originations.
Some behaviors that were seen in the Mercer County case study (Section 4.1) are seen statewide in New Jersey. For instance, there is a concentrated morning rush as well as a more disaggregated period of return trips home. Moreover, the regular spikes produced by train arrivals are present statewide.

![Cumulative Distribution by Time](image)

Figure 4.13: NJ Cumulative Trip Distribution by Time.

Another interesting distribution to assess is the total trips distribution by distance. The trip distribution by distance from origin can reveal trip length characteristics for traveler’s in New Jersey. Figure 4.14 displays this distribution. The shortest trip is 1 mile and the longest trips is 202.5 miles. The travel lengths are skewed to the right; that is, the average length of a trip in the simulation is 16.9 miles whereas the median length of a trip is 12.5. The median is a better indicator for our system needs since it is less sensitive to the relatively few long trips that cause the skew. The shortest trip is 1 mile because our system assumes all shorter trips will be conducted by walk or bike.

On a typical weekday in New Jersey, the synthesizer suggests that about half the destinations are within 12.5 miles of the origination, 75% of the destinations are within 22 miles of origination and 90% of destinations are within 37 miles or origination. The values produced by the synthesizer are a little higher than national averages; according to the United States Department of Transportation, the median distance traveled by a vehicle in 2009 was 10.1 miles [27]. Part of this discrepancy can be accounted for by the walk or intrapixel trips that are accounted for in the U.S. Department of Transportation Study but that was ignored in our analysis. Moreover, the simulation uses Manhattan distances which can lead to over-approximation of distance and produce a higher median distance value.

Figure 4.15 shows the trip distribution by the highest ranked pixels in terms of volume. As in the Mercer County case study, this suggests a very dense trip distribution; that is, few pixels capture
Figure 4.14: NJ Cumulative Trip Distribution by Distance.

Figure 4.15: Cumulative Distribution of NJ Trips by Highest Volume Pixels. The x-axis contains the top pixels in rank order based on trips produced. This shows that most trips are generated from few pixels. The top 32% highest volume pixels produce 90% of all the trips.
Figure 4.16: NJ Top 1,000 Trip Producing Pixels. Areas of high volume include I-95 corridor between Philadelphia and New York, regions outside of New York and Philadelphia, beaches and resorts along the shoreline, and Bucks County.

a high proportion of the trips. It can be seen that there are diminishing returns as more pixels are served with aTaxiStands. This information is incredibly useful since it indicates that simply starting with a few aTaxiStands can serve a disproportionately high demand. For instance, to serve 75% of the total demand, just 3,610 aTaxiStands are needed; however, to serve 100% of the demand 21,586 aTaxiStands would need to be opened.

Figure 4.16 depicts the top 1,000 trip producing pixels in the state of New Jersey. As expected, the regions that produce the most trips include the areas around Philadelphia and New York, and the I-95 corridor including Trenton, Princeton, and Newark. Note that the out-of-state skyscrapers are quite high as well. For instance, Bucks County, Rockland, West, and South regions are in the top 1,000 trip producing pixels. The map is a bit misleading because trips originating from out-of-state were mapped to a single pixel; thus, the entire spatial distribution over the land area is not quite captured for the out-of-state originated trips. Nonetheless, the volume information is captured.

The highest trip producing pixel is displayed in Figure 4.17. This corresponds to Newark Liberty International Airport in Newark. This single pixel produced 195,952 trips. This is logical since an airport should be expected to have numerous patrons on a given day. Moreover, the pixel contains
the Newark Liberty Airport Amtrak and extensive parking infrastructure which leads to more trips produced.

4.2.3 New Jersey Rideshare Analysis

The same process that was applied to the Mercer County case will be applied to the state of New Jersey. Just as in the Mercer County case study, the average vehicle occupancy and miles saved ratio should be identified for the state across different parameter sets. Refer to Section 2.3.2 for details about the properties of each metric and Section 4.1.3 for details on applications of the properties in this system.

New Jersey Average Vehicle Occupancy

Just as in the Mercer County case, the largest increases in AVO due to DD increases occur when DD goes from 0 to 1. This again is explained partially by the pairing phenomenon. As soon as a delay is instituted, a short trip can pair with a long trip very easily and provide a rideshare environment. The graph also reveals diminishing returns over the longer delays. The first minute is the most productive and the rate of AVO increase diminishes from there. The largest increase in AVO due to CD increases occurs as CD goes from 1 to 2. This is logical because due to the circuity conditions, many short trips pair up with long trips. Figures 4.18 and 4.19 graphically display the results of the simulation. For actual values, see Figures 7.3 and 7.4 in the Appendix.

It is interesting to note that in Figure 4.18 at DD = 0, various CD parameters have AVOs noticeably above 1. This indicates that two travelers arrived at the aTaxiStand at the same time, immediately (since there is no departure delay) determined that rideshare was feasible between them, and then shared that ride. This occurs with noticeable frequency due to the New Jersey train station adjustments. According to the simulation, the train makes arrivals at set schedules so all
the individuals who depart at a given stop, will exit the train at the same time. This time will then become all of their origin times for the next leg of the trip. Hence, they are all eligible for ride sharing at the same exact moment. Although doing this with a 0 second delay is not entirely realistic, this feature captures the ridesharing potential at sites such as train stations where there is a mass of travelers seeking rides immediately. This is exactly where ridesharing is useful.

Similar to the potential of ridesharing at certain pixels such as a train station, rideshare potential can differ temporally. In the assessment of the trip volumes, the trip distribution by time (Figure 4.13) indicated that certain periods of the day produced more trips. This includes the morning rush (7AM to 9AM) and evening travel back for students and workers (4PM to 7:30PM). Trips that
occurred between 25,000 and 32,400 seconds after midnight as well as trips that were taken between 58,000 and 70,000 seconds after midnight were extracted and analyzed. Figures 7.5 through 7.10 in the appendix provide the graphical and tabular results of the simulation.

The trips that occur during the morning period and the evening period do exhibit higher rideshare potential across all parameter sets. For instance, the AVO of (DD, CD) = (1, 3) of NJ all day is 2.06, yet in the morning it is 2.16, and in the evening it is 2.31. Another interesting behavior is that the CD = 1 is very high for the morning hours. This indicates that many people are going to the same destination since a shared ride is only permitted to 1 distinct location. Currently, these individuals drive separately in the morning. Taking advantage of this shared AM ridership could significantly reduce the congestion in the mornings. See Figures 7.5 and 7.6 for more detail.

Overall, New Jersey’s AVO under different parameters is quite competitive with the national average. Creating a statewide system with parameters as low as CD = 2 and DD = 1 will eclipse the average national AVO. With higher, yet realistic parameters even more ridesharing can be realized.

**New Jersey Miles Saved Ratio**

As noted in the case study, the miles saved ratio graphs (Figures 4.20 and 4.21) have a resemblance to the shapes of the AVO graphs since the MSR is just a weighted AVO. Also, the MSR is lower than the AVO due to the pairing phenomenon as well as the structure of the rides.

At moderate parameters, such as (DD, CD) = (2, 3), there is significant amount of miles saved due to the sharing that is permitted. For example, at (5, 5) the value almost reaches 3. These values indicate a substantial number of miles can be taken off the road which leads to less congestion and overall greater mobility. The key is in tuning the parameters low enough to induce travelers to utilize the system but high enough to realize the congestion reduction potential.
Assessment

There is significant rideshare potential in New Jersey along with the opportunity to reduce vehicle miles traveled. Moreover, the simulation indicates that rideshare opportunities vary spatially and temporally. For instance, shared rides opportunities are abundant at different locations throughout New Jersey. Certain pixels such as train stations will have a huge potential for ride share due to regular influx of a mass of people seeking to travel. In addition, shared ride opportunities are not static throughout the day. Greater rideshare potential and vehicle miles saved potential exist during certain portions of the day. This delivers important insight because it can inform the beginnings of an aTaxi system. Serving high potential areas at high potential times will allow the system to reduce congestion at heavily trafficked regions during high volume time. Taking advantage of the spatial and temporal variations in ridesharing is key in allowing this system to be even further optimized and utilized.
Chapter 5

Empty Vehicle Routing

In the previous chapter, rideshare potential was assessed and the advantages of using a system of aTaxis that permits shared rides were seen. In this system, however, there is a fundamental assumption that sufficient aTaxis exist to supply travel demand. Essentially, each aTaxiStand held an infinite number of aTaxis that could service all the demand. In a real system, however, restrictions will have to be placed on the number of vehicles within that system. The most impact on congestion reduction will occur when there are a fewer number of vehicles that are intelligently coordinated. One method of coordination that can yield congestion reduction is the intelligent routing of the vehicles.

The methodology presented in the previous work is valuable since it provides insight into the rideshare potential and aTaxi vehicle demand. That is, with an unlimited measurement budget (running as many simulations with tweaks to parameters as needed) and unlimited resources (unlimited number of cars at each station) the true rideshare potential and the fleet sizes required to satisfy demand across pixels, regions, counties or state are captured.

However, in a practical setting, neither having an infinite set of cars nor retiring a car after a single tour is realistic. In a practical implementation of this system, there will be a finite number of stations and cars that take multiple tours over the course of the day. Running the current rideshare simulation without any limits will give us the upper limit of the expected value (since the synthesizer is simply a set of realizations of random variables drawn from distributions) of the total number of vehicles that is needed. However, once we start reducing the number of cars to realistic implementable levels, we will not have full coverage with our current routing heuristic. Hence, we will have to alter the problem to a closed problem where we have a finite number of cars but utilize these cars for more than one tour.

Take for instance, the toy case where we have only 2 travelers and 2 pixels (i.e. aTaxiStands) separated by 2 minutes of travel time. There is no Departure Delay or Circuity in this example so cars leave as soon as a passenger enters and can only have one destination. The first passenger arrives at Pixel 1 with the destination of Pixel 2 at 12:00:45PM. The passenger boards and arrives at Pixel 2 at 12:02:45PM. The car then retires. Then, 11 seconds later, another passenger arrives at Pixel 2 to go to Pixel 1. A vehicle from the infinite pool of aTaxis at Pixel 2 will be boarded at 12:02:56 PM and this vehicle will send the passenger to Pixel 1. Here, 2 aTaxis were used when only one aTaxi actually was needed. The obvious way to adjust for this is to have the empty car enter a
queue of a limited number of cars and be reused to serve demand.

5.1 Empty Vehicle Routing Formulation

With this idea of utilizing vehicles for multiple tours, the simulation results provide an interesting insight. At any given period of time, there is a number of vehicles arriving at the pixel and a number of vehicle departing the pixel. In the closed system, we can assume that the arriving vehicles provides the supply that meets the demand for departing vehicles within a reasonable time period. The ideal situation results when the number of arriving vehicles and departing vehicles are equivalent. The interesting problem arises when there is a difference between the number of arriving vehicles and the number of departing vehicles.

Figures 5.1 and 5.2 shows the aTaxi activity over the course of a day at the Princeton pixel that was seen in Section 4.1.2. In particular, this graph shows the cumulative number (and distribution) of arriving and departing vehicles at the pixel over time. The difference between the number of vehicles arriving to the pixel at a certain time and the number of vehicles departing from the pixel at that same time is the excess supply of aTaxis. This excess supply of aTaxis constitute the “empty” vehicles since these aTaxis will not be utilized once they arrive at the pixel due to the dearth of demand. The graph shows that Princeton has a pretty even departure and arrival flow until about 28,000 seconds (about 8am). Then, the morning rush brings many arrivals. The difference between the arrivals and departures do not even out until the evening when every begins to returns home.

In a closed system, when more cars arrive at a pixel than is needed, there is an excess in supply of vehicles. Like in the Princeton pixel example, these cars in the system simply remain there without being utilized during the time period. Conversely, when fewer cars arrive at a node are needed, there is unmet demand that needs to be satisfied. More cars are needed to meet all demand. Hence, the key approach is to develop a routing scheme that eliminates this inefficiency from the imbalance of supply and demand. In particular, vehicles in pixels which have excess supply should be routed to the pixels that have projected unmet demand in some future time period.

To have a clearer understanding, it is important to note that the difference (delta) between arriving and departing vehicles at a node at any time can be extracted from the simulation. A positive delta indicates excess supply of aTaxis and a negative delta indicates unmet demand. At a particular time period, \( t \), nodes which have a positive delta have vehicles that are not doing useful work once all demand at that node has been met. Hence, instead allowing these vehicles to remain there during that time period, they will be routed to pixels that have a negative delta in a future time period. At time \( t \), empty vehicles at “supply” nodes should be routed to “demand” nodes that will have unmet demand at a future time period. Of course, if there is an excess demand at the same pixel in the next time period, it would be optimal to keep the vehicle in that pixel. This will be reflected in the problem formulation.

The problem that is described is in a class of network flow optimization problems called the Transportation or Transshipment problem. In the following section, the fundamental framework for solving the problem will be presented.
5.2 The Transportation Problem

The transportation or transshipment problem seeks to assign supply to demand while minimizing the cost of moving the material. A particular property of the transportation problem is that the set of all nodes $N$ can be partitioned into two unique sets $S$ and $D$ where $N = S \cup D$ and $S \cap D = \emptyset$. The
nodes $S$ are the source (supply) nodes and the nodes $D$ are the sink (demand, destination) nodes. This division of nodes forms a particular type of graph called the bipartite graph in which every connection links a node in $S$ to a node in $D$. Figure 5.3 displays a bipartite graph. One stipulation for the problem formulation is that the supply must be nonnegative at each supply node and the demand must be nonnegative at each demand node [31].

![Figure 5.3: Example of Bipartite Graph for Transportation Problems.](image)

A general transportation problem can be concisely written as:

$$
\text{minimize} \quad \sum_{i \in S} \sum_{j \in D} c_{ij} x_{ij} \\
\text{subject to} \quad \sum_{j \in D} x_{ij} = s_i \quad i \in S \\
\sum_{i \in S} x_{ij} = d_j \quad j \in D \\
x_{ij} \geq 0 \quad i \in S, j \in D
$$

Where:
- $i$ = index of particular supply node
- $j$ = index of particular demand node
- $S$ = set of all supply nodes
- $D$ = set of all demand nodes
- $s_i$ = total supply at node $i$
- $d_j$ = total demand at node $j$
- $x_{ij}$ = the decision, what quantity to send from $i$ to $j$

The objective function is the component of the problem that is minimized. In the transportation problem, the cost of transportation is typically minimized. In addition, an optimization problem is subject to constraints. In this problem, all demand must be met, all supply must be exhausted, and vehicles can only travel from supply node to demand node. Another more subtle situation arises if the total supply values and the total demand values are not the same. In this case, a dummy node
can be inserted that either provides the additional supply or demand.

This framework can be extended to the routing problem at hand. By making assumptions and adapting the basic model, a preliminary routing solution to reduce congestion can be obtained.

5.3 The Routing Framework

The preliminary transportation model presented in Section 5.2 cannot be directly adapted to the problem at hand. The primary difficulty is that the aTaxi routing problem has an added dimension of time that prevents it from fitting into the simple transportation problem formulation. A decision (the vector \( x \)) must be made at time \( t \) for the routing of excess supply at \( t \) to meet the demand at future time periods. Although the future time period does not necessarily have to be just \( t + 1 \), restricting the possible set of demands to only those demands 1 time period in the future offers great simplification. This can be described as the myopic approach to the time dimensioned (or time dependent) transportation problem. With this simplification, the problem can then be reduced a series of one period problems where the supply and decisions at time \( t \) will meet demand at time \( t + 1 \). The one period problems can then be reformulated with incremented time and solved successively to get an overall routing strategy.

There are several assumptions and details that need to be clarified. So far, the discussion has referred to time in units of “periods.” A period length must be large enough so that it captures sufficient vehicle movement over the time period but not so large that the granularity that is sought is lost. A logical choice might be an hour since cars can be assumed to transport one person each hour. Moreover, an hour gives a rerouted vehicle sufficient time to be able to reach a larger range of pixels.

Another fundamental and unrealistic assumption that is made in this preliminary routing assessment is that all nodes are reachable from any node within an hour. That is, there will be no restriction to the length a vehicle can travel to reach the node exhibiting a demand. Although this is not entirely feasible, it permits a simplification that allows simpler formulation of the problem and still provides much insight. Moreover, distance is the cost of the transportation that is being minimized so it can be expected that a majority of distances of rerouting trips should be short and viable in practice.

Every pixel that has empty aTaxis at time \( t \) will be the supply for demand at time period \( t + 1 \). Indexing the variables by time allow us to frame the problem according to the myopic period iteration model:

\[
\begin{align*}
\text{minimize} & \quad \sum_{i \in S} \sum_{j \in D} c_{ij} x_{ij}^t \\
\text{subject to} & \quad \sum_{j \in D} x_{ij}^t = s_i^t \quad i \in S \\
& \quad \sum_{i \in S} x_{ij}^t = d_j^{t+1} \quad j \in D \\
& \quad x_{ij}^t \geq 0 \quad i \in S, j \in D
\end{align*}
\]

This problem can be repeated over any time frame for which data is available — hours, days, weeks,
etc. The synthesizer dataset provides 24 hours of data that can be utilized.

In the formulation of a transportation problem, the objective function seeks to minimize cost of the transportation. The cost in this system will be the distance the vehicle must travel to satisfy the demand. The constraints are simply that 100% of the demand has to be met. The shortest cumulative reroute that meets all demand is the solution. Other factors could be used but distance adequately measures the disutility in routing.

Finally, there will likely be an imbalance in supply and demand at a given time period for the successive runs of the model. That is, there will be either more supply than demand or more demand than supply at a particular time period. Therefore, a dummy source node and dummy sink node will be introduced in order to provide the balancing supply or demand. The cost to go to this dummy node is 0.

5.4 Extensions to Basic Routing Implementation

Section 5.3 presents a fundamental framework to approaching the transportation problem in the simulated aTaxi system setting. However, many improvements can be made to the routing.

For instance, this approach made a fundamental assumption that all vehicles can reach another node within the hour. A more robust analysis will create tighter stipulations to prohibit traveling over set number of miles in a given period of time. This implementation can use binary decision variables in order to determine whether an empty vehicle is routed to a particular pixel. This will allow an additional constraint that only allows certain distances to be traveled: \( c_{ij} \times x_{ij} < \text{maxDistance} \).

Also, in the deterministic setting, excess or dearth in supply was bypassed by simply creating dummy nodes that have no cost. However, in a more advanced setting, multi-period lookaheads can be utilized to see whether a routing can happen at the present time period for all future time periods. For instance, at time \( t \), instead of just looking at routing opportunities at time \( t + 1 \), time periods up to \( t + n \) can be checked where \( n \) is some integer greater than 1. In this way, an empty that has not utilization even in the next time period can be assigned to meet a demand that is several time periods in the future instead of floating to the sink.

Finally, all factors in this framework is deterministic. However, the synthesizer produces trips as a result of drawing from distributions. Hence, the synthesizer output is really an expectation of the travel demands instead of the true value. Optimizing in a stochastic framework is different since analysis must account for the variability in the measurements. Moreover, other factors not present in the deterministic formulation such as the cost of overage or underage, must be assessed. Solutions to stochastic problems can be developed through techniques in dynamic programming [20].
Chapter 6

Conclusion

6.1 Summary

Through the simulation of a system of autonomous vehicles, it was ascertained that rideshare potential does exist which would allow the alleviation of congestion and ultimately better mobility for all. An autonomous taxi system is especially valuable in regions where there is heavy demand such as train stations or times of the day when there are rushes such as during the morning commute to work. Both metrics that were used to analyze the ridesharing consequences indicated that an institution of an autonomous vehicle system could reduce congestion. Reasonable parameter sets for ridesharing produced results that far surpassed the national average vehicle occupancy and the vehicle miles saved as a result of the rideshare is significant.

The case study of Mercer County and the NJ statewide analysis are just a starting point. An analysis should be conducted on each county and rideshare and optimal parameter sets should be determined with this granularity. Moreover, this process should be extended in other regions, both across the United States and the world. Finally, the efficient routing in the system is crucial for its success. This can reduce the fleet size required to service travel demand and reduce congestion even further. The framework for addressing the problem is presented in this thesis. More realistic constraints should be imposed and the process should be refined in order to obtain the optimum solution.

6.2 Improvements and Next Steps

Several changes could be made to improve both the framework of the aTaxi system and the implementation simulation. Currently, the optimum rideshare match does not always occur when rideshare is identified. Rather, the first aTaxi that passes the criteria is simply used as the rideshare vehicle. This causes the suboptimal pairing phenomenon which is reflected in lower MSR values. Instead of starting at the head of the queue and selecting the first rideshare match, an efficient iteration should be made through all the aTaxis to determine optimal ridesharing. This will allow the maximum value to be extracted from the system's perspective. Ideally, a long trip will be matched with a long trip and a short trip will be matched with a short trip.

Moreover, Manhattan distances are used in this simulation as the measure of distance. This is
a better approximation than Euclidean distance since the extra distance built into the Manhattan
can account for the circuity in a real path. Unfortunately, the Manhattan distance tends to
erate distance. This can be seen by observing the trip distribution by distance CDF plot.
The ideal solution is to link the framework to the actual underlying road systems of New Jersey.
This will provide the most accurate distance measures and improve the quality of the simulation.

More realistic features can be added to this framework. Perhaps travelers can be picked up
from intermediate location between two nodes. Perhaps a new traveler can be picked up from a
destination of one of the current passengers. Perhaps trains to locations other than just New York
or Philadelphia can be utilized. These are all adjustments that would make for a more realistic
system and would optimize its performance.

The implementation of the simulation can also be improved. Currently, the entire trip set is
read into memory, sorted by the appropriate columns, and then parsed according to the appropriate
logic. If the sorting can be done outside the module and trips can be analyzed without reading in all
the data first, space and time can be utilized much better. Moreover, creation of a GUI that allows
interaction with data and permits toggling of parameters would improve analysis capabilities.

6.3 Future and Implications

The analysis presented has revealed that there is definite feasibility in congestion reduction capabilities of autonomous taxis. The technology simply needs to be commercially available for the system to be actually implemented. With the prototypes for the Google Car, the Lincoln MKZ auto driving car, and Toyotas self-driving Lexus, a definite investment in the technology is present and milestones are being reached and shattered. Every day, it seems that the case for autonomous vehicles progresses. For instance, California, Nevada and Florisa have even passed acts that allow for the licensing of autonomous vehicles to be tested.

Financially, the autonomous vehicle and an aTaxi system is feasible since the only cost is in the technology. Unlike other modes of transportation which require an investment in infrastructure, autonomous vehicles use the existing guideways for mobility. However, technological, legal, financial and hosts of other barriers do exist. Nevertheless, the convenience, safety, mobility and utility that autonomous vehicles confer makes the adoption of such a transformative technology seem inevitable.

Autonomous vehicles could very well be the future of transportation that emancipates the human from the driving function, and it will be a much welcomed innovation. Within this generation, the vehicle will shift from the “ultimate driving machine” to the “ultimate riding machine.”
Chapter 7

Appendix

7.1 Mercer County Figures

<table>
<thead>
<tr>
<th></th>
<th>CD = 0</th>
<th>CD = 1</th>
<th>CD = 2</th>
<th>CD = 3</th>
<th>CD = 4</th>
<th>CD = 5</th>
<th>CD = 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>DD = 0</td>
<td>1.00</td>
<td>1.02</td>
<td>1.06</td>
<td>1.07</td>
<td>1.07</td>
<td>1.07</td>
<td>1.07</td>
</tr>
<tr>
<td>DD = 1</td>
<td>1.00</td>
<td>1.07</td>
<td>1.73</td>
<td>1.99</td>
<td>2.09</td>
<td>2.13</td>
<td>2.17</td>
</tr>
<tr>
<td>DD = 2</td>
<td>1.00</td>
<td>1.10</td>
<td>1.91</td>
<td>2.33</td>
<td>2.52</td>
<td>2.61</td>
<td>2.70</td>
</tr>
<tr>
<td>DD = 3</td>
<td>1.00</td>
<td>1.12</td>
<td>2.01</td>
<td>2.54</td>
<td>2.82</td>
<td>2.95</td>
<td>3.11</td>
</tr>
<tr>
<td>DD = 4</td>
<td>1.00</td>
<td>1.14</td>
<td>2.09</td>
<td>2.70</td>
<td>3.05</td>
<td>3.23</td>
<td>3.46</td>
</tr>
<tr>
<td>DD = 5</td>
<td>1.00</td>
<td>1.16</td>
<td>2.15</td>
<td>2.83</td>
<td>3.23</td>
<td>3.46</td>
<td>3.76</td>
</tr>
</tbody>
</table>

Figure 7.1: Mercer County Average Vehicle Occupancy Table.

<table>
<thead>
<tr>
<th></th>
<th>CD = 0</th>
<th>CD = 1</th>
<th>CD = 2</th>
<th>CD = 3</th>
<th>CD = 4</th>
<th>CD = 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>DD = 0</td>
<td>1.00</td>
<td>1.00</td>
<td>1.04</td>
<td>1.05</td>
<td>1.05</td>
<td>1.05</td>
</tr>
<tr>
<td>DD = 1</td>
<td>1.00</td>
<td>1.03</td>
<td>1.63</td>
<td>1.86</td>
<td>1.95</td>
<td>1.99</td>
</tr>
<tr>
<td>DD = 2</td>
<td>1.00</td>
<td>1.04</td>
<td>1.78</td>
<td>2.14</td>
<td>2.31</td>
<td>2.39</td>
</tr>
<tr>
<td>DD = 3</td>
<td>1.00</td>
<td>1.05</td>
<td>1.86</td>
<td>2.31</td>
<td>2.55</td>
<td>2.67</td>
</tr>
<tr>
<td>DD = 4</td>
<td>1.00</td>
<td>1.06</td>
<td>1.91</td>
<td>2.43</td>
<td>2.73</td>
<td>2.89</td>
</tr>
<tr>
<td>DD = 5</td>
<td>1.00</td>
<td>1.07</td>
<td>1.96</td>
<td>2.53</td>
<td>2.87</td>
<td>3.07</td>
</tr>
</tbody>
</table>

Figure 7.2: Mercer County Miles Saved Ratio Table.
### 7.2 New Jersey Figures

#### 7.2.1 NJ Figures: All Day

<table>
<thead>
<tr>
<th></th>
<th>CD = 0</th>
<th>CD = 1</th>
<th>CD = 2</th>
<th>CD = 3</th>
<th>CD = 4</th>
<th>CD = 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>DD = 0</td>
<td>1.00</td>
<td>1.02</td>
<td>1.07</td>
<td>1.08</td>
<td>1.08</td>
<td>1.08</td>
</tr>
<tr>
<td>DD = 1</td>
<td>1.00</td>
<td>1.09</td>
<td>1.77</td>
<td>2.06</td>
<td>2.17</td>
<td>2.22</td>
</tr>
<tr>
<td>DD = 2</td>
<td>1.00</td>
<td>1.13</td>
<td>1.96</td>
<td>2.40</td>
<td>2.61</td>
<td>2.71</td>
</tr>
<tr>
<td>DD = 3</td>
<td>1.00</td>
<td>1.16</td>
<td>2.08</td>
<td>2.63</td>
<td>2.92</td>
<td>3.07</td>
</tr>
<tr>
<td>DD = 4</td>
<td>1.00</td>
<td>1.19</td>
<td>2.16</td>
<td>2.80</td>
<td>3.16</td>
<td>3.36</td>
</tr>
<tr>
<td>DD = 5</td>
<td>1.00</td>
<td>1.21</td>
<td>2.23</td>
<td>2.93</td>
<td>3.36</td>
<td>3.60</td>
</tr>
</tbody>
</table>

Figure 7.3: NJ AVO Table.

<table>
<thead>
<tr>
<th></th>
<th>CD = 0</th>
<th>CD = 1</th>
<th>CD = 2</th>
<th>CD = 3</th>
<th>CD = 4</th>
<th>CD = 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>DD = 0</td>
<td>1.00</td>
<td>1.00</td>
<td>1.06</td>
<td>1.06</td>
<td>1.06</td>
<td>1.06</td>
</tr>
<tr>
<td>DD = 1</td>
<td>1.00</td>
<td>1.04</td>
<td>1.60</td>
<td>1.82</td>
<td>1.91</td>
<td>1.95</td>
</tr>
<tr>
<td>DD = 2</td>
<td>1.00</td>
<td>1.06</td>
<td>1.74</td>
<td>2.07</td>
<td>2.23</td>
<td>2.31</td>
</tr>
<tr>
<td>DD = 3</td>
<td>1.00</td>
<td>1.07</td>
<td>1.82</td>
<td>2.24</td>
<td>2.45</td>
<td>2.57</td>
</tr>
<tr>
<td>DD = 4</td>
<td>1.00</td>
<td>1.09</td>
<td>1.88</td>
<td>2.36</td>
<td>2.63</td>
<td>2.78</td>
</tr>
<tr>
<td>DD = 5</td>
<td>1.00</td>
<td>1.10</td>
<td>1.93</td>
<td>2.46</td>
<td>2.77</td>
<td>2.95</td>
</tr>
</tbody>
</table>

Figure 7.4: NJ MSR Table.
7.2.2  NJ Figures: Select Periods

<table>
<thead>
<tr>
<th></th>
<th>CD = 0</th>
<th>CD = 1</th>
<th>CD = 2</th>
<th>CD = 3</th>
<th>CD = 4</th>
<th>CD = 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>DD = 0</td>
<td>1.00</td>
<td>1.02</td>
<td>1.11</td>
<td>1.11</td>
<td>1.12</td>
<td>1.12</td>
</tr>
<tr>
<td>DD = 1</td>
<td>1.00</td>
<td>1.09</td>
<td>1.88</td>
<td>2.31</td>
<td>2.52</td>
<td>2.61</td>
</tr>
<tr>
<td>DD = 2</td>
<td>1.00</td>
<td>1.12</td>
<td>2.05</td>
<td>2.66</td>
<td>3.01</td>
<td>3.20</td>
</tr>
<tr>
<td>DD = 3</td>
<td>1.00</td>
<td>1.16</td>
<td>2.15</td>
<td>2.87</td>
<td>3.33</td>
<td>3.61</td>
</tr>
<tr>
<td>DD = 4</td>
<td>1.00</td>
<td>1.18</td>
<td>2.23</td>
<td>3.03</td>
<td>3.57</td>
<td>3.92</td>
</tr>
<tr>
<td>DD = 5</td>
<td>1.00</td>
<td>1.20</td>
<td>2.29</td>
<td>3.15</td>
<td>3.76</td>
<td>4.17</td>
</tr>
</tbody>
</table>

Figure 7.5: NJ AVO Table - 4PM to 7:30PM.

<table>
<thead>
<tr>
<th></th>
<th>CD = 0</th>
<th>CD = 1</th>
<th>CD = 2</th>
<th>CD = 3</th>
<th>CD = 4</th>
<th>CD = 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>DD = 0</td>
<td>1.00</td>
<td>1.05</td>
<td>1.09</td>
<td>1.09</td>
<td>1.09</td>
<td>1.09</td>
</tr>
<tr>
<td>DD = 1</td>
<td>1.00</td>
<td>1.28</td>
<td>1.97</td>
<td>2.16</td>
<td>2.21</td>
<td>2.22</td>
</tr>
<tr>
<td>DD = 2</td>
<td>1.00</td>
<td>1.37</td>
<td>2.30</td>
<td>2.66</td>
<td>2.78</td>
<td>2.81</td>
</tr>
<tr>
<td>DD = 3</td>
<td>1.00</td>
<td>1.44</td>
<td>2.50</td>
<td>3.01</td>
<td>3.20</td>
<td>3.27</td>
</tr>
<tr>
<td>DD = 4</td>
<td>1.00</td>
<td>1.49</td>
<td>2.66</td>
<td>3.28</td>
<td>3.54</td>
<td>3.64</td>
</tr>
<tr>
<td>DD = 5</td>
<td>1.00</td>
<td>1.53</td>
<td>2.78</td>
<td>3.50</td>
<td>3.83</td>
<td>3.97</td>
</tr>
</tbody>
</table>

Figure 7.6: NJ AVO Table - 7AM to 9AM.
Figure 7.7: NJ AVO vs CD - 7AM to 9AM.

Figure 7.8: NJ AVO vs CD - 7AM to 9AM.
Figure 7.9: NJ AVO vs CD - 4PM to 7:30PM.

Figure 7.10: NJ AVO vs DD - 4PM to 7:30PM.
Bibliography


