SUPPORTING ROUTE CHOICES VIA REAL-TIME VISUAL TRAFFIC INFORMATION AND COUNTERFACTUAL ARRIVAL TIMES

BY DAEHAN KWAK

A dissertation submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Computer Science

Written under the direction of

Badri Nath

and approved by

____

____

____

____

New Brunswick, New Jersey

May, 2017

ABSTRACT OF THE DISSERTATION

Supporting Route Choices via Real-time Visual Traffic Information and Counterfactual Arrival Times

by Daehan Kwak

Dissertation Director: Badri Nath

Mobility plays an integral role in modern lives, yet with the ever-expanding number of cars, traffic congestion poses various negative effects, causing vast economic loss, air pollution, and commuter stress. As live traffic information is becoming ubiquitous, route guidance systems are used to inform drivers of route capacities to avoid traffic congestion. Navigation systems compare several different routes and provide the user with options to choose from, from a list of best possible route recommendations. Drivers' route choice decisions are typically based on the route that minimizes their travel cost (e.g. travel time). However, there are three main limitations for route guidance and information systems. First, as travel time reliability plays an influential role in the driver's route choice decision-making, the difference in the travel time estimations and/or recommended routes may vary across navigation systems, which can contribute to the uncertainty in the route choice. Second, as the estimated travel time is the dominant deciding factor in route choice, the impact of uncertain, inaccurate, and variable travel time estimations can render it useless, negatively influencing the drivers' compliance to

the information system's recommended route. Third, as drivers cannot assess and compare their actual route choice to the non-chosen foregone alternatives, they face frequent dilemmas over their route-choice decisions, especially when route alternatives recommended by navigation systems are not consistent with their own previous driving experiences.

In this dissertation, our focus is to explore these three limitations. First, we present a comparative analysis on the route recommendations given from four popular online map providers: Google Maps, HERE, MapQuest and Bing Maps. We analyze traffic data collected from all four of the different map providers for 71 days for two cities, each with two origin-destination pairs. Statistical analysis show that the estimated travel times on identical routes are significantly different among the map providers. This in itself has the potential to create uncertainty in route choices and travel time variability, in addition to a decrease in the credibility and compliance with the map provider's route choice. Second, to complement the deciding factors (e.g., Estimated Time of Arrival (ETA)) in route decisions, we propose a system called Social Vehicle Navigation. This system incorporates a secondary level of detail into the vehicle navigation system by providing other semantically rich information that drivers can share with one another. This user-shared visual traffic information assists in the decision-making process and also improves the efficacy in route determinations. Third, we introduce a rationale for counterfactual thinking in route choice, where drivers receive feedback information about the actual travel times on forgone alternatives (i.e. non-chosen routes), so that at the end of the day, drivers have the ability to exercise reinforced learning and self-assessments of their route choices. We propose DoppelDriver, a system that offers a direct, actual travel time comparison among chosen and non-chosen routes, which determines the actual travel times from probe participatory vehicles on the non-chosen routes.

The main conclusion of this dissertation is that existing navigation systems have limitations and can potentially introduce uncertainty in route choice. To support and improve the driving experience, we address the use of visual traffic information for pre-trip route choice and the use of counterfactual travel times as post-choice feedback information on the forgone alternatives.

Acknowledgements

My deep gratitude first goes to Professor Liviu Iftode, who is no longer with us, but resting in heaven. He has been a great mentor, guiding me through my graduate education, always pushing me forward and motivating me on my research and studies. I have been very fortunate to have an advisor, who has given me freedom to explore first on my own while at the same time, teaching me to strive for constant improvements. His continuous support for the last seven years at Rutgers has been tremendous and I will miss him dearly. I know he will be proud of me.

I also want to thank Professor Badri Nath, for taking me under his wings and continuing to support me by being available for discussions and providing helpful feedback for me. Our discussions have turned out to be very constructive, and have helped me sort out the details of my projects. His insights and advices have been invaluable. Professor Nath has provided a wealth of benefits to me as I struggled to complete this work.

My gratitude extends to Professor Thu Nguyen and Professor Vinod Ganapathy as well. Professor Thu has always been available for great conversations and guidance for my track for Ph.D. and also job search advices. I have had the pleasure to work with Professor Vinod Ganapathy and he has constantly helped, supported, and shown me encouragements along the way. I would also like to thank Dr. Stephen Smaldone for his support and help as my external committee member.

During my time at Rutgers, I have had the pleasure of working with many excellent people.

My appreciation also goes out to my colleagues, Ruilin Liu, Nader Boushehrinejadmoradi, Daeyoung Kim, and Hai Nguyen, whom I have had the pleasure to learn various things from. I especially would like to thank Ruilin Liu and Daeyoung Kim, who have helped me days and nights on my experiments and research and showing great teamwork. Throughout my long journey all of these people have contributed to make my Ph.D. life special. I am grateful to the staff of the department, especially, Carol DiFrancesco, for assisting me in many different ways.

I also send my thanks to Gregory Hill, my supervisor, my mentor, and my friend. Getting me through my military services, helping me set up at Rutgers, reviewing my work... You have helped and supported me in endless ways. I greatly appreciate all of what you have done for me.

Lastly, I would like to thank my parents for their unconditional love, patience, and support. They add onto my purpose and inspiration. Last but not the least, I am thankful to Amber Kim, for her patience and love while I was working on my Ph.D. I dedicate and owe my dissertation to them.

Thank you all!!!

Dedication

To whom I owe everything, my mother and father.

To Amber Kim, for your love, trust, and support.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Ever since the invention of the first automobile, the power of mobility has improved the frequency at which people meet to maintain social relations over even greater distances. However, with the ever-expanding number of automobiles throughout the world, traffic congestion is a pervasive problem that exacts severe economic cost and negatively impacts on an individual's well-being. Countless countermeasure solutions exist to alleviate traffic congestion, for example, the infrastructure-based Intelligent Transportation System (ITS) such as, on-ramp flow meters, cameras that detect traffic flows, optimized traffic signals, number plate recognition systems, road pricing (e.g. road tolls), electronic informational displays (e.g. Variable Message Sign (VMS) [100]) along the roadways. Today, the most widely used ITS solutions are infrastructure-less, named so because they use floating car data to determine traffic speed and to identify traffic congestion. As ubiquitous computing and real-time traffic flow data becomes pervasive in cars, they are integrated with navigation systems which has become the prominent solution to avoid traffic congestion. Apart from traditional navigation systems, users have gravitated toward smartphone navigation services, such as Google Maps [6]. A J.D. Power and Associates study reported that 47% of 20,704 car owners polled had downloaded a navigation application to supplement their car-based navigation system in 2012 [77]. Most drivers rely on GPS navigation, be it through their cars or their smartphones where a rapidly growing market far from its ceiling is speaking volumes to the ubiquitous presence of navigation services.

1.1 Travel Information

Traffic information can influence drivers' route choice behavior, and consequently, guide them to less congested routes. Traditionally, traffic reporting has been done primarily by the police, state departments of transportation, drivers reporting by phone, and also by traffic reporting companies. Such information is aggregated and then either resold or redistributed directly to the public, broadcast on-air by radio and TV stations, or used as traffic data for in-car navigation systems. Live traffic status reports are becoming more common and easily accessible, with traffic congestion maps available via online maps, mobile phones, and GPS devices. Nowadays, participation by the public in providing traffic reports is becoming popular, because it is easy for users to report and share traffic information with one another via smartphones.

1.1.1 State-of-the-art on Travel Information

Traffic Information via Social Networking Services

Due to the increased usage of social networking services (SNS) like Facebook [5] or Twitter [16], users post events to inform or share information with one another. Researches into extracting useful information from SNS, especially traffic information, have been conducted [87, 19, 85, 106, 81]. In [106], the authors proposed extracting traffic events from Twitter and, using natural language techniques, and classified them by the location mentioned and the type of traffic event. Similarly, Sakaki *et al.* [85] extracted traffic-related Tweets along with location information based on keywords. Schulz *et al.* [87] suggested detecting small incidents, such as car accidents from microblogs by utilizing machine learning and semantic web page techniques, and localizing the microblogs by location and time to provide real-time detection of incidents. Ribeiro *et al.* [81] investigated the relationship between Foursquare and Instagram check-ins to detect traffic movement and built a more efficient traffic condition predictor. Andrea *et al.* [19] proposed a real-time system to detect from Twitter events relevant to traffic and to classify them by event type, such as traffic congestion, crash, or heavy traffic due to an external event like a football game.

Traffic Information via Participatory Crowdsensing

Crowd-sourced participatory sensing applications are becoming popular because users can easily use the sensors that are built in smartphones (e.g. cameras, GPS, accelerometer and motion sensors) to collect data which is then aggregated and interpreted for a common purpose. Micro-blog [41], an early work in participatory sensing, provides a system that connects sharing users and querying users to generate and share geo-tagged multimedia called micro-blogs. Authors in [40] proposed a navigation service called GreenGPS that uses participatory sensing data from on-board diagnostics (OBD) systems to compute fuel-efficient routes, which may differ from the shortest and fastest routes. Smartphone navigation apps such as Waze [17], differ from traditional navigation systems in that crowdsourcing via users provides traffic reports to a central server, where such information is used to provide real-time routing and traffic reports. Here, push buttons are used to share road reports like the degree of traffic, police speed traps or accidents. More recently, apps such as Inrix Traffic [8] have started to incorporate user-based crowdsourced traffic reports, and after Google's recent acquisition of Waze, it started to add social traffic reports into its mapping business. Trapster [15] is also a community-based crowd-sourcing app that improves the commute when users report and share the location of speed traps, red light cameras and speed radar, etc. Without the complex infrastructure to predict bus arrival time, the authors in [111] proposed a bus arrival time prediction system based on participatory contribution of users with commodity mobile phones. In addition, recent researches [38, 92, 89, 45] also talk about the potential from users sharing information regarding traffic, and we are starting to witness widespread use of these techniques.

1.2 Route Choice

In the transportation field, route choice is concerned with the decision-making process of route selection from origin to destination. The driver route-choice model is one of the most influential and challenging topics, because the decision by a user makes affects traffic. We can already witness the pervasiveness of navigation systems, which mainly use the Estimated Time of Arrival (ETA) . These systems take into account real-time traffic flow and provide estimated travel time route comparisons to generate a set of route choices to support the driver in the decision-making.

A myriad of research has been conducted on route choice behavior and the effect of real-time information on route choice [96, 64, 51, 107, 34, 91, 55, 39, 101, 90]. [91] conducted studies to understand the traveler's route choice behavior under uncertainty, and [39, 101] looked into the impact of real-time information that influences influence drivers' route choice. In [90], the authors analyzed drivers' route-choice behavior with pre-trip information. Main findings were that drivers found to be more reluctant to be influenced by the information provided if they had more experience with the routes, and when exploring alternatives increases, individuals tend to prefer a route characterized by lower average but greater variance of travel time.

1.2.1 Limitations in Route Choice

The pervasiveness, efficiency, and market potential of ITS and map service providers strictly depend on users' compliance with the systems and the information they offer. Despite efforts to support decision-making, travelers tend to disregard the suggestions these types of systems offer [27, 28]. In fact, based on their findings, Zhu and Levinson [112] investigated the basic question: "Do people take the fastest route?" In this empirical study, GPS devices were installed in the vehicles of 190 participants for a 13-week period, where no particular instructions were given; they were free to make their own route decisions. They compared the travel time from routes that participants actually chose against the fastest route, which was calculated from other probe vehicles. Their results show that only about 13.5% of commuting trips match the fastest route. Also, in [79], the authors investigated the relationship between driver's behavior and information provision via real-time traffic information from GPS manufacturer TomTom. With 32 participants, 897 trips were analyzed based on GPS traces. They conclude that the driver's perception of the route choice is biased in favor of the chosen route, where the participants consider their chosen route to be reliable when, in reality, they are among the most unreliable.

There are several explanations for user compliance with real-time information (i.e. the decision to follow the recommended route supplied through real-time information). From the transportation literature, we list a few limitations in current route recommendation systems as a means to alter drivers' route choice behavior to make smarter decisions, which mainly rely on real-time ETA information.

Reliability

The quality of real-time information influences the users' compliance with route choice, and as ETA is used as a decision-making tool for route choice, the impact of uncertain, inaccurate, and variable travel times can render it useless [79]. Similarly, differences in ETA and/or best routes offered by various map providers can also contribute to the uncertainty in route choice, and this uncertainty can make it challenging for users to choose the true optimal route.

Rationality

In situations of uncertainty, most people may simply decide on routes they have experienced previously because they have no time or method for rational assessment [28]. We observe this quite often as commuters routinely take the same route to and from work [64]. To some extent, this may be seen as rational; however, in the long run, the decision may cost the driver a significant amount of travel time, thus resulting in irrational decision-making [63].

Perception

Arguably, people may think they are on the fastest route, even when their decision is wrong and do not follow the system's recommended route [112, 79]. People often experience these events in their ordinary commute, where they think they are better route planners than their navigation system, and persist in using "their" route [79, 63]. Consequently, their false beliefs result in poor decision-making if it turns out they were wrong. One behavioral factor that explains this is habit, which appears to have a strong influence on drivers' behavior. Even when real-time traffic information is provided, and even after experiencing long delays, drivers are still willing to adhere to "their" route [79].

Experience

In many situations, drivers face dilemmas over their route-choice decision, especially when real-time information regarding the recommended alternatives does not coincide with previous driving experience [79, 24, 22]. In addition, repeated cycles of accumulated experience on continuous commutes can be a better decision-making factor than route-choice systems [22] because real-time information cannot foretell whether traffic will get better or worse.

Information

Real-time traffic information is becoming more ubiquitous and is used in many systems and services. It is also used in route choice systems to provide better recommendations so that the user can make better decisions. However, such information cannot foretell you that traffic will get better or worse.

Though route choice is a very important aspect, drivers tend to spend less time on thinking, assessing and learning from their route decisions and they eventually wind up spending more time in traffic. People are not willing to spend time comparing travel times for alternative routes. Most do not believe it would be a rational use of their time to explore multiple alternatives which may turn out to only save a few minutes. However, rationally, it might be worth trying as it pays off in the long run [63].

1.3 Thesis

This dissertation states that:

> Existing route recommendation systems have the potential to introduce uncertainty in route choice. Real-time visual traffic information can support the user's pre-trip route-choice decision making and, with counterfactual travel time feedback on forgone alternative routes, users are capable to make strategic decisions on, and self-assessments of, their route choices.

1.4 Summary of Dissertation Contributions

This dissertation presents three main contributions in exploring and identifying the research challenges to support route choices in route recommendation systems, which were published as follows:

- Social Vehicle Navigation: Integrating Shared Driving Experience into Vehicle Navigation, *in Proceedings of the 14th Workshop on Mobile Computing Systems and Applications*, (HotMobile 2013), pp. 16:1–16:6. [89];
- Tweeting Traffic Image Reports on the Road, *in Proceedings of the 6th International Conference on Mobile Computing, Applications and Services*, (MobiCase 2014), pp. 40–48. [59];
- DoppelDriver: Counterfactual Actual Travel Time for Alternative Routes, *in the Proceedings of the 2015 IEEE International Conference on Pervasive Computing and Communications*, (PerCom 2015), pp. 178–185. [60];
- Seeing Is Believing: Sharing Real-Time Visual Traffic Information via Vehicular Clouds, *in IEEE Access*, vol. 4, pp. 3617-3631, 2016. [61];
- A Comparative Analysis of Real-time Travel Information from Online Map Services, [submitted for peer review].

First, we identify the potential of uncertainty in route choice in existing popular online map recommendation systems: Google Maps, HERE, MapQuest and Bing Maps. We analyze analyzed traffic data collected from all four of the different map providers for 71 days for two cities, each with two origin-destination pairs. An extensive comparison analyses are conducted on several aspects, including the difference in estimated travel times across map providers on identical routes, and the number of recommend routes introduced and their differences across map providers for a given origin and destination. Statistical analysis shows that the estimated travel times on identical routes are significantly different among all the map providers ($p < 0.001$). This in itself has the potential to create uncertainty in route choices and travel time variability, in addition to a decrease in the credibility and compliance with the map provider's route choice.

Second, we present a system called *Social Vehicular Navigation* that integrates user-shared visual traffic reports called NaviTweets into a vehicle navigation system in order to provide a secondary level of detail to identify traffic congestion. The users' shared traffic reports are geo-tagged onto a map, called Social Traffic Map (a map representation). Based on these traffic reports, Traffic Digests (concise snapshot summaries on the route of interest) are delivered to drivers to provide rich and reliable information supporting the route choice. These digests will complement factors such as the estimated travel time and assist the driver on their route choice decision making. Once the Traffic Digest is received, the information is displayed in a user-friendly way to the drivers to assist them with route selection. We explain the functions of the proposed SVN model in abstraction layers and present the architecture along with a prototype implementation. Also, we have conducted a questionnaire survey to evaluate the usage of traffic images in route choice and as well a user study to investigate the real behavior of traffic image information sharing.

Third, we propose a system called *DoppelDriver*, a system that answers these users' route choice uncertainties. Through the provision of instantaneous position relative to participatory users on alternative routes as well as actual travel time (ATA) for any alternative routes to a given destination, DoppelDriver offers a direct comparison on chosen and non-chosen routes. We show the potential use of post-trip feedback information on chosen and non-chosen routes based on insights through a literature review in the transportation domain. We design an algorithm to obtain actual travel times on non-chosen routes that combines segments of the actual travel times on the alternative routes crowdsourced by other participating users. We present results from a feasibility study based on GPS traces from a taxi data set. The hypothesis is that the total travel time along a route, as experienced by a single driver, can be accurately approximated by adding the travel times along successive portions of the route as experienced by multiple drivers traversing those route portions at approximately the same time as the single driver. We design and implement a prototype DoppelDriver on the Android platform. When an origin and destination are inputted, a recommended choice list of alternative routes is shown corresponding with ETAs based on current traffic. Users can select which route to take and which route they are interested in receiving feedback from, based on actual travel time. Also, users can save their trip travel times, along with the non-chosen route's travel times, in the trip log. Finally, based on the trip log, the average of, and deviation in, the trip's actual travel times are shown in the trip diary.

1.5 Contributors to the Dissertation

The following is a list of contributors by chapters who co-authored papers from which material was used in this dissertation. Chapter 2 of this dissertation are the result of a collaboration with my colleagues, Yu Yang, Ruilin Liu, and my advisors Prof. Badri Nath and Prof. James Abello. Yu Yang contributed to cleansing the traffic data, Ruilin Liu and both Prof. Badri Nath and Prof. James Abello provided many essential insights. For Chapter 3, Wenjie Sha and Daeyoung Kim contributed in the design and implementation to the prototype, Ruilin Liu contributed in the evaluation analysis, and both my advisors Prof. Badri Nath and Prof. Liviu Iftode contributed to the motivation and design of Social Vehicular Navigation. For Chapter 4, Daeyoung Kim contributed in the implementation to the prototype, Ruilin Liu contributed to provide the taxi data, and both my advisors Prof. Badri Nath and Prof. Liviu Iftode contributed in many ways to the idea and motivation.

1.6 Organization of the Dissertation

The outline of this dissertation is as follows. Chapter 2 presents the potential of route-choice uncertainty through a comparative analysis on the route recommendations and estimated travel times given from four popular online map providers: Google Maps, HERE, MapQuest and Bing Maps. Chapter 3 and 4 describe the use of two systems to support route choice, Social Vehicular Navigation and DoppelDriver, respectively. Chapter 3 presents Social Vehicular Navigation, a system that crowdsources geo-tagged visual traffic information to support drivers in the selection of routes. Chapter 4 presents DoppelDriver, a system that answers route-choice uncertainty through post-travel feedback information. Finally, Chapter 5 presents the conclusion along with future work to this dissertation.

Chapter 2

Comparative Analysis on Online Maps

One of the most essential services in transportation systems is a route recommendation system, which provides optimal origin/destination (OD) route alternatives based on real-time estimation and prediction of travel times. Online maps, such as Google Maps, Bing Maps, MapQuest, HERE, Apple Maps, and Waze, are the de facto standards for this type of transportation service. Such online map providers use historic traffic information and/or live traffic information to construct the fastest route based on the expected time of arrival (ETA). Given the variety and diversity of travel time estimations and route recommendations, it is important to understand and gain insight into a routing service's quality. In this paper, we study four popular online maps (Google Maps, Bing Maps, Here, and MapQuest) for two cities (Los Angeles and New York) selected from cities with the worst traffic, requesting routes for two OD pairs in each city. Online map data for the recommended routes were collected for 71 days for each online map. A comparison analysis was conducted on these online maps based on the recommended routes and the ETAs.

2.1 Introduction

Navigation systems are mostly used to make better use of the transportation network, and enable users to be better informed about route capacities in order to avoid traffic congestion.

Navigation system manufacturers are integrating traffic-related, specialized functions into on-board vehicle navigation systems. The number of in-vehicle navigation systems is predicted to quadruple in North America by 2019, growing to about 13 million [36]. Apart from the traditional navigation manufacturers, such as TomTom or Garmin, drivers are moving to less expensive smartphone navigation apps, such as Google Maps and Apple Maps [35]. Vehicles equipped with these types of devices are expected to be an integral part of the Internet in the near future [45].

Today, most navigation systems and apps can calculate the best route, taking into account real-time travel data, as well as historic data, to predict traffic flow. Travel time information has been the subject of considerable research because, from the traveler's perspective, it supports better route decision making, better scheduling of departure times, reduces vehicle operating costs, and improves the overall experience. With respect to transportation agencies, this information provides criteria to optimally manage and control traffic to reduce congestion, and thus, provide a more sustainable environment.

In particular, real-time travel information provided by online (or mobile) maps, such as Google Maps, Bing Maps, HERE (Nokia), MapQuest, and Waze, are services that are getting better every day, making it more and more possible to avoid heavy traffic and to know when you will arrive at a certain destination. However, to use and comply with recommendations provided by online maps, it is essential that travel time estimations be reliable, as travel time reliability is a major factor in route choice [97, 48]. Also, it is important to understand the travel time variations (the differences in travel time estimations provided by one recommendation system compared to another), since users do not precisely know when they will arrive at a destination [73]. Thus, both travel time reliability and travel time variability are related to, and can introduce uncertainty associated with, route choice [48, 21].

To understand the characteristics of online map services, the objective of this study is to explore and gain insight into travel time variations and route recommendations by conducting a comparative analysis among state-of-the-art online maps. For example, what is the ETA difference among the map providers for a particular fixed route? Are different routes recommended among the map providers for a given origin and destination? To the best of our knowledge, we are the first to conduct a comparison study of online maps.

This chapter is organized as follows. Background and related work are presented in Section 2.2. Then, we explain the setup of our experiment in Section 2.3, followed by a description of the data that was collected during the experiment in Section 2.4. Section 2.5 presents a descriptive study on our collected data set. In Section 2.6, we shows the results from a comparative study on the online maps. Finally, the study summary is presented in Section 2.7.

2.2 Background and Related Work

In the transportation field, route choice behavior is connected to the decision-making process in route selection, and research has been conducted to understand this complex behavior [79, 99]. Many factors influence the decision process. For example, there are observable attributes, such as travel time, cost, distance, fewest turns, trip purpose, and traffic information availability; and there are unobservable characteristics, such as age, gender, income, attitude, perception, personality, spatial abilities, and road network familiarity [47, 53].

Since travel time estimation is the key indicator in the user's route choice, inaccurate travel times can render recommendation systems useless. Drivers may find it difficult to determine if the route chosen was the most expedient upon arrival, making it important to assess the quality of travel time estimations in general. As a matter of fact, it is becoming common to witness drivers using both in-vehicle navigation systems and smartphone navigations apps at the same time. Wang and Ju [105] investigated this behavior in order to understand the reasons for using two navigation systems at the same time. One of their findings was that, during traffic congestion, participants found that it was necessary to use both navigation systems to decide what the best route is. Also, when both navigation systems recommend different routes, participants tend to rely more on the smartphone navigation app, compared to in-vehicle navigation. Although the objective of the study was to understand the various reasons why drivers use two navigation systems, we can see that this is an example of route uncertainty.

Several studies have been conducted to assess travel time by comparing the ground truth travel time collected from static traffic sensors or probe vehicles, compared to the estimated travel time provided by online maps [97, 56, 71, 98, 103, 29]. However, such studies are constrained to freeways and arterials due to limitations in field data collected by static sensors [49, 82].

In addition, a study of 10,156 emergency ambulance journeys in England determined that discrepancies between actual and estimated travel times from commercially available navigation systems across hospital emergency departments were an under-prediction of 1.6 minutes and a standard deviation of 4.9 minutes [70]. Given the small radius of service that a hospital is responsible for, and given that ambulances can exceed general traffic speed, an under-prediction of 1.6 minutes and a rather large standard deviation of 4.9 minutes raises the question of accuracy in estimated travel times.

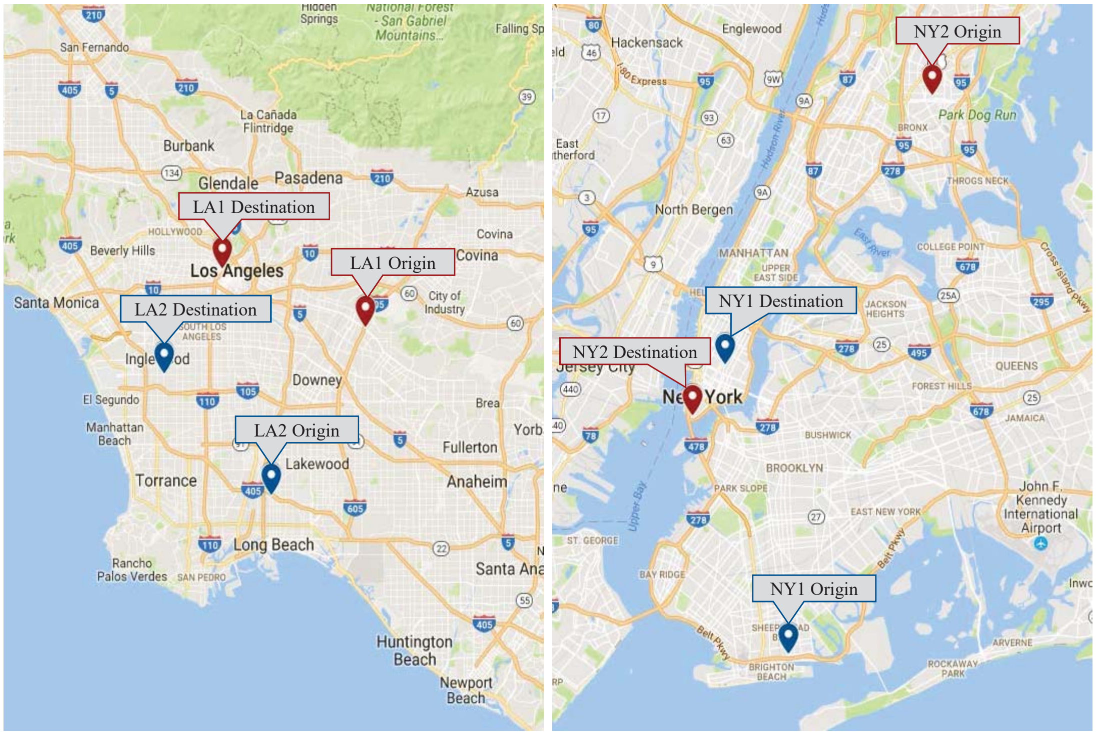
An early work to compare estimations of travel times among different mapping systems was conducted [104]. A comparison was done between Google Maps and the ArcGIS Network Analyst Module [3], where a set of origins and a set of destinations, varying by distance, were input to both mapping systems, and the resulting travel times were compared. The authors identified several advantages with Google Maps in frequent updated road networks, and travel time estimations took congestion into account. However, the limitation of this work was that the authors compared the travel times reported from Google Maps, which accesses updated traffic information, versus ArcGIS, which relies on static data extracted from the software DVD.

Travel time can easily be under- or overestimated due to dynamic changes in traffic volume or by simple and faulty algorithms. Additionally, there is a lack of clarity in the algorithms in estimating travel times across routing services. The fact is that data on current traffic conditions come from a wide variety of sources, ranging from government departments of transportation, to users who anonymously contribute speed and traffic information through crowdsourcing, to private data providers, or even traffic data companies. So, a wide discrepancy in sources of data for each travel time prediction may surely be the case across different online map providers. Each of these sources of variability in travel time across route recommendation systems may make it challenging for the user to choose the best route, potentially introducing uncertainty in route choice.

2.3 Experiment Setup

Selection of Online Maps and Cities

Popular de facto online maps in the U.S. are Google Maps [6], Bing Maps [4], MapQuest [10], OpenStreetMap (OSM) [11], HERE [7], Waze [17], and Apple Maps [2]. Out of these online maps, we selected the ones that are supported by both web browsers and mobile devices and that use real-time traffic information. OSM does not use real-time traffic data, Apple Maps is not supported by web browsers, and Waze map's web browser functions were not stable to properly collect data at the time of the experiment. Thus, we conducted a comparative study based on the other four online maps (Google Maps, Bing Maps, HERE, and MapQuest) for the two cities which were selected from cities with the worst traffic, Los Angeles and New York



LA NY

Figure 2.1: Map location of origins and destinations for LA and NY.

[52, 80].

Selection of Location and Distance for Origin/Destination Pairs

We carefully selected the location of, and the distance between, the origins and destinations in the two cities studied. The selections of the two OD pairs for each city (a total of four OD pairs) were based on the following criteria. a) *Destination*: The destination was selected from among the most popular areas in each city: Downtown LA and Orange County for LA [58], and Midtown Manhattan and Lower Manhattan for NY. b) *Distance*: Since the average one-way commute time is 26 minutes (over an average distance of 16 miles) [62, 76], we selected approximately 16 miles for the distance between origin and destination. c) *Origin*: For the origins in LA, we selected two locations that were east and south of downtown LA, because there are mountains north of LA and the Pacific Ocean is west of LA, as shown in Figure 2.1

Table 2.1: Coordinates and descriptions for origins and destinations.

City	Origin	Description	Destination	Description
LA1	33.992548, -118.067819	East of LA	34.052482, -118.263706	Downtown LA
LA2	33.818773, -118.195800	Long Beach	33.945049, -118.343907	Orange County
NY1	40.586315, -73.948369	Brooklyn	40.727958, -73.990979	Mid Manhattan
NY2	40.859623, -73.851198	Bronx	40.702850, -74.012957	Lower Manhattan

(left). For the origins in NY, as shown in Figure 2.1 (right), we selected two locations that are northeast and southeast of Manhattan, because the Hudson River is west of Manhattan. d) *Route*: From the recommended route options, we checked to see if at least one route is the same for all four maps. Details of the OD pair coordinates and their descriptions are shown in Table 2.1. Figure 2.1 depicts the map location of the OD pair coordinates.

2.4 Data Description

A server was set up to query route recommendations from four online maps: Google Maps, HERE, MapQuest, and Bing Maps. Route requests were queried with the coordinates of the OD pairs as input (Table 2.1). Each online map was queried for 71 days at five-minute intervals, starting from 11/29/2014 (01:20) to 02/07/2015 (15:20). The total data set size is 804,568, where Table 2.2 shows the breakdown according to city and online map provider.

The result of each query is a list of recommended routes, generally one to three routes for Google Maps, HERE, and MapQuest. However, for Bing Maps, only a single route was recommended for every query. Thus, although each map service was equally queried, the data sizes are different among the map providers shown in Table 2.2. This is due to the fact that sometimes the total number of routes that were reported differed among online maps, or there were query failures.

Data cleansing was conducted on the data sets. The MapQuest data had approximately 2.4% abnormal data for a total of 5,958 (LA1: 5454, LA2:304, NY1:66, NY2:134) where ETAs had

Table 2.2: Dataset size (Total: 804,568).

City	Google	HERE	MapQuest	Bing Maps
LA1	60,056	61,028	61,066	20,350
LA2	61,013	61,037	61,109	20,372
NY1	52,850	61,002	60,996	20,332
NY2	61,001	61,014	60,998	20,344

Table 2.3: Sample dataset.

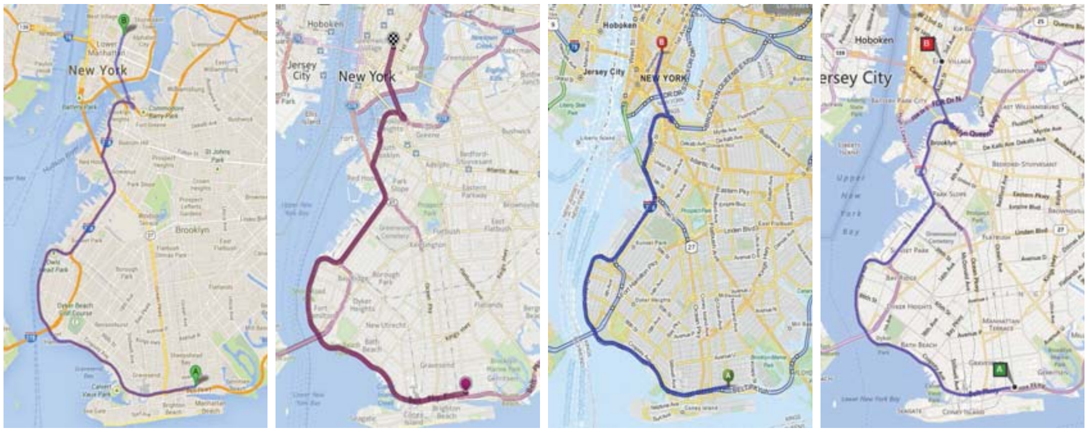
ID	Date/Time	Idx	Dist	Tm_s	Cur_tm_s	Rid	Map	Location
53348	11/29/13 15:30	0	18.3	1260	1560	G3	google	I-5 N
53349	11/29/13 15:30	1	19.9	1260	1320	G4	google	CA-60 W
53350	11/29/13 15:30	2	20.9	1320	1320	G5	google	I-605 N and CA-60 W
53351	11/29/13 15:25	0	18.3	1260	1680	G3	google	I-5 N
53352	11/29/13 15:25	1	19.9	1260	1320	G4	google	CA-60 W
53353	11/29/13 15:25	2	20.9	1320	1380	G5	google	I-605 N and CA-60 W

extremely high values, and thus, were excluded from the analysis. In addition, missing data from query failures were negligible, and thus, ignored.

The response to each query was recommended routes along with their corresponding ETAs. A sample of the dataset for a query response is shown in Table 2.3, which consists of nine attributes. *ID* is the record number; *Date/Time* is the response date and time of the route query; for each Date/Time, online maps may recommend more than one route, so we assigned a different index, *Idx*, to distinguish them; *Dist* is the distance of the recommended route in miles; *Tm_s* is the static travel time in seconds; *Cur_tm_s* is the estimated travel time taking into account live traffic information in seconds; for each distinct route, *Rid* is assigned; Location is the main description of the route.

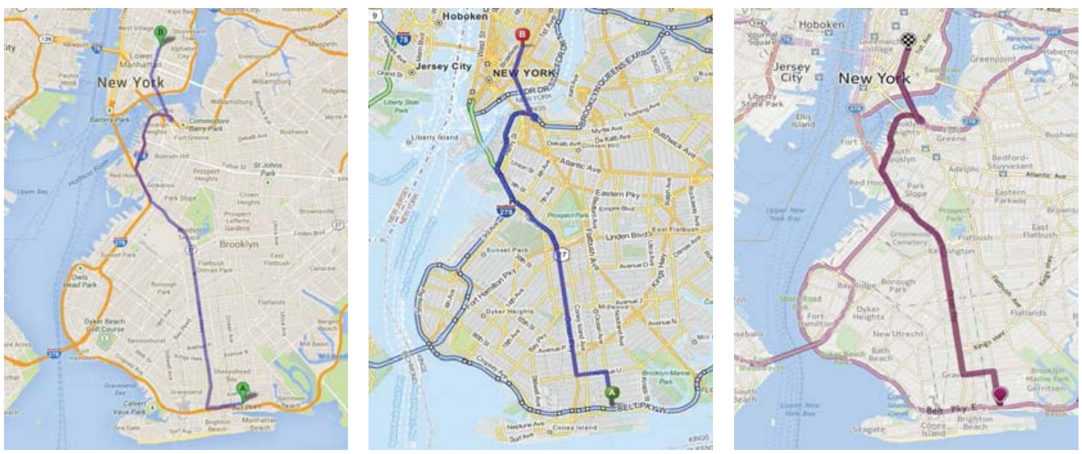
Generally, a total of three route options were recommended per route query for Google Maps, HERE, and MapQuest, whereas Bing Maps provided a single best recommended route per route query. Table 2.4 shows the total number of distinct routes that were recommended throughout the 71 days, along with the corresponding range for *Rids* in parentheses.

Each distinct route may have identical routes that match a route given by other online maps.



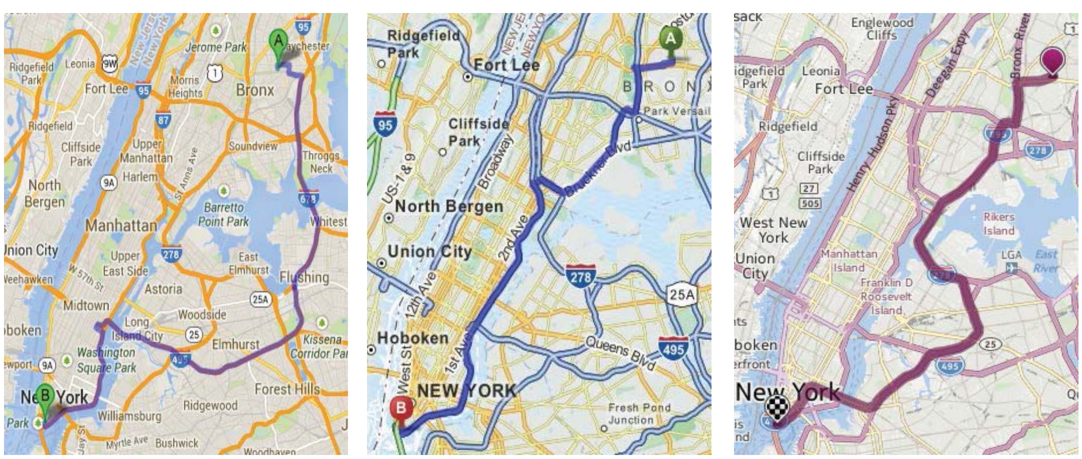
Google Maps HERE MapQuest Bing Maps

Figure 2.2: Example of identical routes in NY1 (G11=H5=M5=B242).



Google Maps HERE MapQuest

Figure 2.3: Example of similar, but not identical, routes in NY1 (G6≈H3=M3).



Google Maps HERE MapQuest

Figure 2.4: Example of a route in NY2 with no identical routes from other services.

Table 2.4: Total number of distinct routes.

City	Google	HERE	MapQuest	Bing Maps
LA1	7 (G1 – G7)	4 (H1 – H4)	3 (M1 – M3)	208 (B1 – B208)
LA2	5 (G1 – G5)	3 (H1 – H3)	3 (M1 – M3)	38 (B1 – B38)
NY1	15 (G1 – G15)	5 (H1 – H5)	6 (M1 - M6)	305 (B1 – B305)
NY2	9 (G1 – G9)	4 (H1 – H4)	3 (M1 – M3)	282 (B1 – B282)

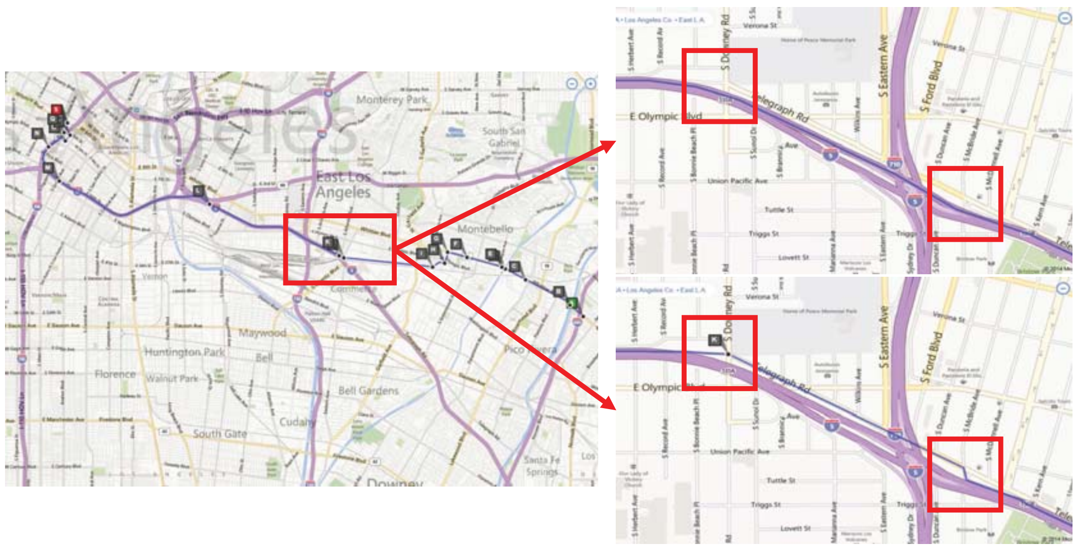


Figure 2.5: Example of micro-changes in routes for Bing Maps in LA1.

Figure 2.2 is a visual example of identical routes from each online map provider for NY1, where G11=H5=M5=B242 indicates that route IDs, G11, H5, M5, and B242 are identical. Figure 2.3 is an example where one route is similar, but not identical, to routes from other online map providers. We can see that the route near the origin from Google Maps is slightly different from the identical routes presented by HERE and MapQuest. In this study, when comparing routes, if one leg of a route is different from another, then that route is considered not identical. Figure 2.4 is an example of a route that has no identical routes from the other online maps.

As mentioned, when a user requests a route, online maps (except for Bing Maps) generally recommend options (more than one route). Bing Maps recommends only a single route per route request. When route options are recommended, we logically inferred that online maps

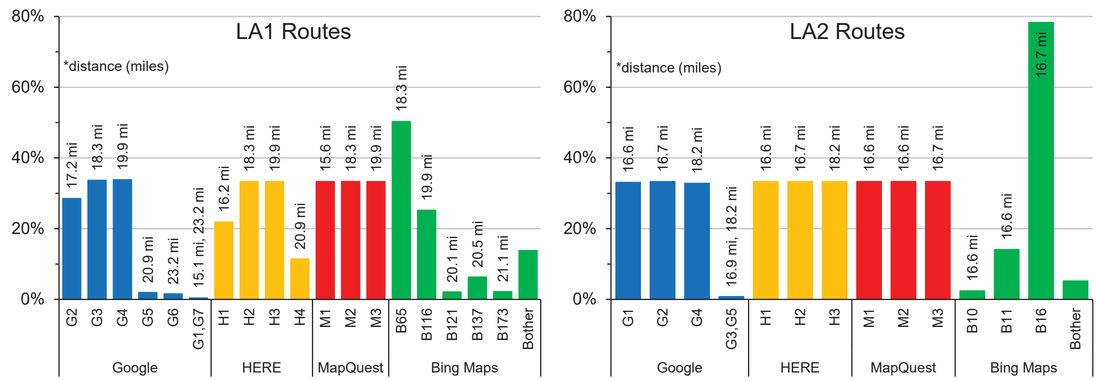


Figure 2.6: Percent of the recommended routes that appear for LA.

other than Bing Maps provide the top three fastest route options from a set of fixed routes, whereas Bing Maps provides the best route that is distinguished by micro-changes. Figure 2.5 gives an example of such micro-changes. The top-right route and the bottom-right route differ in that the top-right route continues on the highway, whereas the bottom-right route exits to local roads and then re-enters the highway. Although Bing Maps, recommends a single best route per query, such micro-changes makes Bing Maps have a large amount of distinct routes.

2.5 Descriptive Study

In this section, descriptive study is conducted to study the ETA average, ETA variability, number of distinct routes, route changes, and number of identical routes. Each aspect is covered in the subsections below.

2.5.1 Recommended Route Options

In this subsection, we first show the total number of routes each online map recommended and the details for each OD pair in LA and NY.

Los Angeles

Throughout the data collection for 71 days, we show the total number of distinct routes that appeared for LA1 and LA2. For LA1, a total of 7, 4, 3, and 208 distinct routes were recommended by Google Maps, HERE, MapQuest, and Bing Maps, respectively. The majority of routes that appeared are G2, G3, and G4 from Google Maps, which account for 96.1%; H1, H2, and H3 for HERE, which account for 88.5%; M1, M2, and M3 for MapQuest, which account for 100%; and B65 and B116 for Bing Maps, which account for 75.5%. Detailed information of the distinct routes that appeared for LA1 is shown in Appendix A Table A.1.

For LA2, a total of 5, 3, 3, and 38 distinct routes were recommended by Google Maps, HERE, MapQuest, and Bing Maps, respectively. The majority of routes that appeared are G1, G2, and G4 from Google Maps, which account for 99.2%; H1, H2, and H3 from HERE, which account for 100%; M1, M2, and M3 from MapQuest, which account for 100%; and B11 and B16 from Bing Maps, which account for 92.4%. Detailed information of the distinct routes that appeared for LA2 is shown in Appendix A Table A.2.

A descriptive summary of the percentage of distinct recommended routes for LA1 and LA2 based on Appendix A Table A.1 and Appendix A Table A.2 is shown in Figure 2.6. We can imply that for LA1, a greater variety of routes appeared, compared to LA2. Routes for MapQuest were the only ones that appeared evenly for LA1. For LA2, three routes for Google, HERE, and MapQuest appeared evenly, whereas one route (B16) was mainly recommended by Bing Maps.

New York

We show the total number of distinct routes that appeared for NY1 and NY2. For NY1, a total of 15, 5, 6, and 305 distinct routes were recommended by Google Maps, HERE, MapQuest,

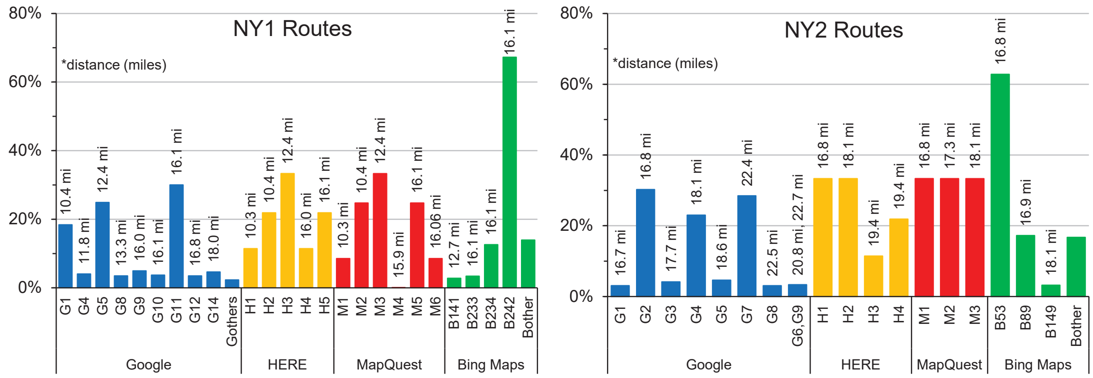


Figure 2.7: Percent of the recommended routes that appear for NY.

and Bing Maps, respectively. The majority of routes that appeared from Google Maps were G1, G5, and G11, which account for 73.3%; H2, H3, and H5 from HERE, which account for 77.1%; M2, M3, and M5 from MapQuest, which account for 82.8%; and B234 and B242 from Bing Maps, which account for 79.8%. Detailed information on the distinct routes that appeared for NY1 is shown in Appendix A Table A.3.

For NY2, a total of 9, 4, 3, and 282 distinct routes were recommended by Google Maps, HERE, MapQuest, and Bing Maps, respectively. The majority of routes that appeared are G2, G4, and G7 from Google Maps, which account for 81.7%; H1, H2, and H4 from HERE, which account for 88.5%; M1, M2, and M3 from MapQuest, which account for 100%; and B59 and B89 from Bing Maps, which account for 80.1%. Detailed information on the distinct routes that appeared for NY2 is shown in Appendix A Table A.4.

Compared to LA, we can see that NY has many more distinct routes. This can be inferred owing to the fact that Manhattan is a grid type of street map, where streets run at right angles to each other, leading to a greater variety of possible routes.

A descriptive summary of the percentage of distinct recommended routes for NY1 and NY2 based on Appendix A Table A.3 and Appendix A Table A.4 is shown in Figure 2.7. We can see for NY1 that a greater variety of routes appeared, compared to NY2. This is because, for NY2,

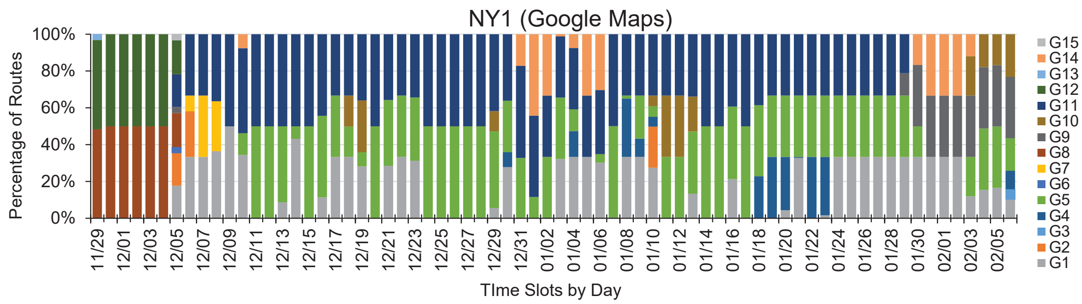


Figure 2.8: Routes viewed by date for NY1.

the OD routes consist mainly of highways. For NY2, routes from MapQuest were the only ones that appeared evenly, whereas one route (B53) was mainly recommended by Bing Maps.

2.5.2 Routes Viewed by Date

Not all the routes appeared every day, with some routes only showing up during a short period of time. Therefore, we show which routes appeared by date. The routes viewed in Google Maps throughout the 71-day period for NY1 is shown in Figure 2.8. All the other online maps and for both LA and NY are shown in Appendix A Figure A.1 to A.4. The x-axis is time slots by date for the 71-day period, and the y-axis is the percentage of routes viewed.

For MapQuest and HERE in all of LA and NY, the routes that appeared were evenly distributed throughout the experiment. For Bing Maps, although there were many distinct routes, only one or two appeared throughout the experiment. Google Maps for NY had the most dynamic route usage, compared to all other maps. However, it is interesting to see that some routes appeared for a short period of time and never appeared again throughout the experiment.

To sum up, it can be inferred that routes for MapQuest and HERE are nearly constant, and are the most static. Although Bing Maps has the most distinct routes, apparently only one or two representative routes are used. Google Maps tended to be the most dynamic, compared to other online maps, but that depends on the road network.

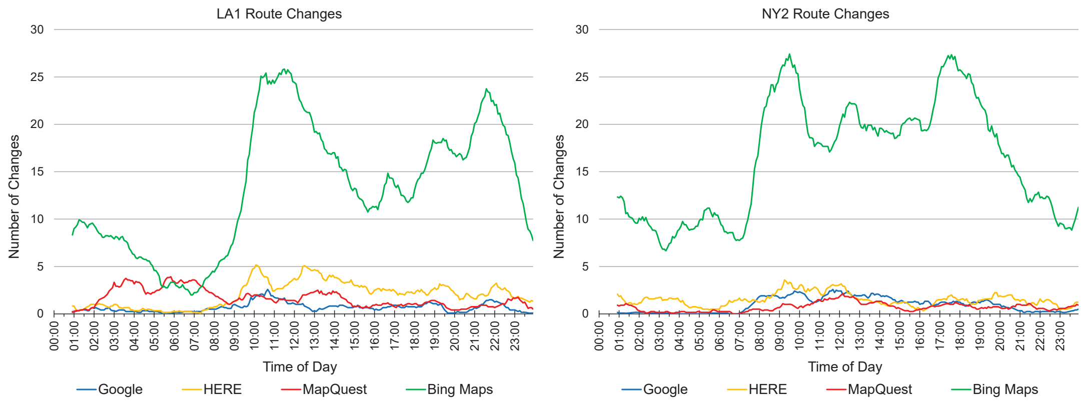


Figure 2.9: Frequency of best-route changes for LA1 and NY2.

2.5.3 Route Changes

In this subsection, we focus on understanding how frequently the best recommended routes change over time. Figures 2.9 shows the frequencies of change for the best routes in LA1 and NY2. The frequency change regarding the other OD pairs are in Appendix A Figure A.5. We used a twelve-period moving average, which is the average in the frequency change every hour, to smooth out the fluctuations. The x-axis is the time of day, and the y-axis is the total number of changes. The findings here also show that the best recommended routes change more frequently during the peak morning and evening hours.

2.6 Comparative Study

2.6.1 Case of Best Routes

In our dataset, Google Maps, HERE, and MapQuest recommended two to three routes whereas Bing Maps recommended a single route for every five-minute interval. We define the best route as the route that has the shortest ETA for the given interval. Thus, for every five-minute interval there would be four best routes from each online map. Note that there may or may not be identical routes that were selected as the four best routes. In this subsection, we present a

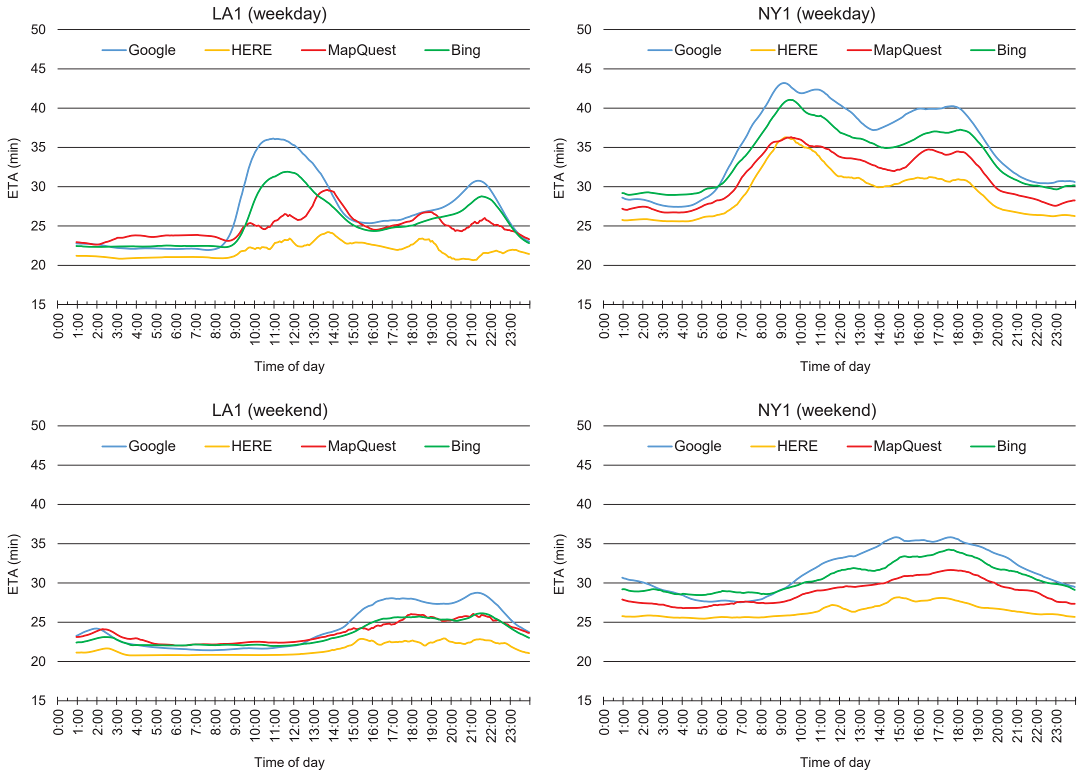


Figure 2.10: Weekday and weekend average ETAs for LA1 and NY1.

descriptive analysis on the best ETA's average and variability along with a statistical analysis for each pair of online map's best ETA.

ETA Average for Best Routes

We plot the average of the best routes' ETAs for each online map. The average of the travel times for each OD pair queried from the online maps for LA1 and NY1 are shown in Figure 2.10. For each query, the shortest ETA was selected to be included in calculating the average, despite the fact that the routes may differ between the online maps. Here, the average ETAs are shown with a five-minute interval, and we separated the data into weekdays and weekends. In each figure, the x-axis is the time of day, and the y-axis is the average of the best ETA for each time each day. Figure 2.10 (top) illustrate the weekday average ETAs for LA1 and NY1. It is noticeable that both figures have the rush hour peak trend (peak in the morning and

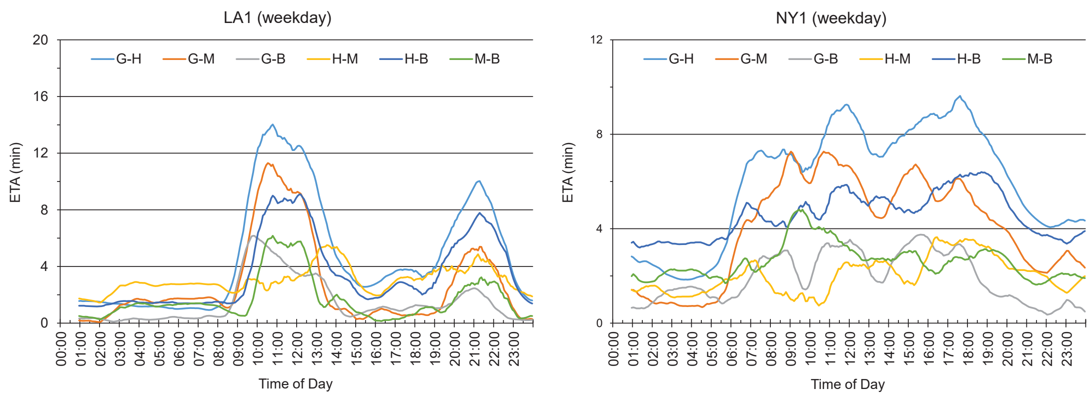


Figure 2.11: Absolute difference for best ETAs on weekdays for LA1 and NY1.

another in the evening), yet the peaks are much higher for NY1. For the weekends, as shown in Figure 2.10 (bottom), we can see that peaks start in the afternoon and are much higher for NY1 compared to LA1. Information regarding the ETA average plots for the reaming OD pairs are shown in Appendix A Figure A.6.

ETA Variability for Best Routes

We show the travel time variability for the recommended routes provided by the online maps. In order to do so, we first show the absolute differences in the best ETAs given for each online map pair. This is the average of the absolute differences for Google versus HERE (G-H), Google versus MapQuest (G-M), Google versus Bing Maps (G-B), HERE versus MapQuest (H-M),

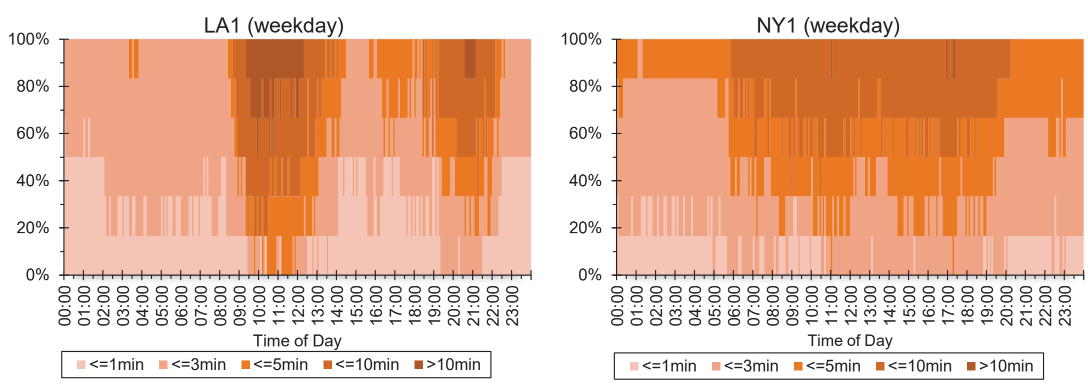


Figure 2.12: Heat map on the difference on best ETAs for all online map pairs (LA1 and NY1).

HERE versus Bing Maps (H-B), and MapQuest versus Bing Maps (M-B). Note that since we are looking at the route that has the minimum ETA from each map provider, routes may or may not be identical. Only LA1 weekday and NY1 weekday are shown in Figure 2.11, and the rest of the other plots for the remaining OD pairs are shown in Appendix A Figure A.7. We used a twelve-period moving average, which is the average in the frequency change every hour, to smooth out the fluctuations We can visually see that G-H has the highest difference, and the differences for each online map pair increase during the peak morning and evening rush hours. Details to test the statistical differences is presented in the next section.

Similarly, to visualize the differences in ETAs provided by online map providers, we plotted the heat map shown in Figure 2.12. Different from Figure 2.11, Figure 2.12 shows the density of differences in terms of minutes. We can see that those in the dark red area, having ETA differences of more than five minutes, are mainly observed during the peak morning and evening hours. The other heat map plots for the remaining OD pairs are shown in Appendix A Figure A.8.

Statistical Analysis

To verify that the best ETA is significantly different among the maps, we first tested whether the travel time distribution follows a normal distribution. We conclude that it does not follow a normal distribution, and our findings are consistent with other research [73, 98] that states travel time distribution is not normally distributed. Therefore, we used the Wilcoxon signed-rank test, a non-parametric statistical hypothesis test, to statistically examine the paired samples for each pair of online map providers. The null hypothesis, H_O, is that two paired data sets are equivalent. All test results showed that p-values were less than 0.05, indicating that we reject H_O, meaning that a significant difference exists. We find that the ETAs for Google Maps were

Table 2.5: Identical routes for LA and NY.

LA1	LA2	NY1	NY2
G3=H2=M2=B65 G4=H3=M3=B116 G5=H4=B164	G2=H2=M2=B16 G1=H1=M1=B11 G4=H3=M3=B27	G11=H5=M5=B242 G5=H3=M3=B84 H1=M1 G1=H2=M2	G2=H1=M1=B53 G4=H2=M3=B149

significantly higher than HERE; those from Google Maps were significantly higher than those from MapQuest; Google Maps ETAs were significantly higher than Bing Maps; HERE ETAs were significantly lower than MapQuest; HERE's were significantly lower than Bing Maps; and MapQuest's were significantly lower than Bing Maps. To sum up, Google Maps, Bing Maps, MapQuest, and HERE is the order from highest to lowest in terms of ETA.

2.6.2 Case of Identical Routes

When different online maps providers recommended routes, there were instances when the recommended route were identical to each other. In this subsection, we identify the identical routes that were recommended by different online map providers. The objective of this is to examine whether online maps have differences in ETA for identical routes. For example, in LA1, route IDs, G3, H2, M2, and B65 are identical. We present all the identical routes observed for LA and NY in Table 2.5.

First, we show when the identical routes were observed throughout the time of day for LA in Figure 2.13 (NY plots are shown in Appendix A Figure A.9). The x-axis is the time of day, and the y-axis is the percentage of identical routes appearing. Note that if the percentage is 100%, it means identical routes were observed for all 71 days. For LA1, we can see that identical routes G3=H2=M2=B65 appear mostly during the night. This means that all four map providers recommended the same route, mostly during the night hours when traffic volumes are low. As the time of day approaches morning rush hour, the appearance of identical routes

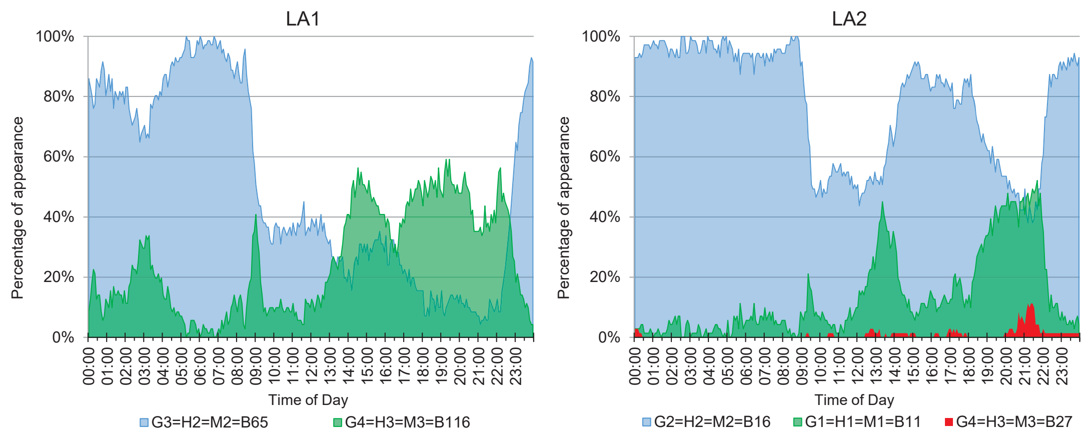


Figure 2.13: Percentage of identical routes that appeared simultaneously for LA.

G3=H2=M2=B65 was reduced. On the other hand, identical routes G4=H3=M3=B116 appear more frequently during the daytime.

Statistical Analysis

In order to test whether different online map providers show similar ETAs for identical routes, we conducted a statistical analysis to compare the ETAs. To get a microscopic view of whether ETAs provided by different online map providers for identical routes were equivalent, we first grouped times of day into five categories, just like Wallace *et al.* [103]. The categories are:

- rush morning (6:00 AM to 10:00AM)
- non rush midday (10:05AM to 2:55PM)
- rush afternoon (3:00PM to 8:00PM)
- non rush night (8:05PM to 5:55AM)
- weekend (all day)

Like the test procedure mentioned in Section 2.6.1, we first conducted a normality distribution test on the travel time distribution for identical routes, and concluded that it does not

Table 2.6: Wilcoxon signed-rank test on identical routes.

City	Route pair	Wkdy rush morning	Wkdy rush afternoon	Wkdy nonrush midday	Wkdy nonrush night	Wknd	Overall
LA1	N	1,782	647	613	3,939	3,182	10,163
	G3-H2	< .001	< .001	< .001	< .001	< .001	G > H
	G3-M2	< .001	< .001	< .001	< .001	< .001	G < M
	G3-B65	< .001	< .001	< .001	< .001	< .001	G > B
	H2-M2	< .001	< .001	< .001	< .001	< .001	H < M
	H2-B65	< .001	.683	< .001	< .001	< .001	H < B
	M2-B65	< .001	< .001	< .001	< .001	< .001	M > B
LA2	N	2,036	2,242	1,371	4,950	5,321	15,920
	G2-H2	< .001	< .001	< .001	< .001	< .001	G > H
	G2-M2	< .001	< .001	< .001	< .001	< .001	G < M
	G2-B16	.040	< .001	.202	< .001	< .001	G < B
	H2-M2	< .001	< .001	< .001	< .001	< .001	H < M
	H2-B16	< .001	< .001	< .001	< .001	< .001	H < B
	M2-B16	< .001	< .001	< .001	< .001	< .001	M > B
NY1	N	615	936	813	1,467	1,826	5,657
	G11-H5	< .001	< .001	< .001	< .001	< .001	G > H
	G11-M5	< .001	< .001	< .001	< .001	< .001	G > M
	G11-B242	< .001	< .001	< .001	< .001	< .001	G > B
	H5-M5	< .001	< .001	< .001	< .001	< .001	H < M
	H5-B242	< .001	< .001	< .001	< .001	< .001	H < B
	M1-B242	< .001	< .001	< .001	< .001	< .001	M < B
NY2	N	1068	1,067	1,155	4,350	3,902	11,542
	G2-H1	< .001	< .001	< .001	< .001	< .001	G > H
	G2-M1	< .001	< .001	< .001	< .001	< .001	G > M
	G2-B53	< .001	< .001	.006	< .001	< .001	G < B
	H1-M1	< .001	< .001	< .001	< .001	< .001	H < M
	H1-B53	< .001	< .001	< .001	< .001	< .001	H < B
	M1-B53	< .001	< .001	< .001	< .001	< .001	M < B

the follow normal distribution. Thus, we used the Wilcoxon signed-rank test to examine if the paired ETAs for each pair of online map providers were equivalent. Again, the null hypothesis, H_O, is that two paired data sets are equivalent. Table 2.6 shows that the p-values are less than 0.05 for all categories, except for two slots, meaning that we reject H_O. We conclude also that the ETAs for identical routes given by different online map providers were significantly different.

The common feature we found in both LA and NY is that the ETA for Google Maps was significantly higher than HERE; HERE was significantly lower than MapQuest; and HERE was significantly lower than Bing Maps. We summarize the significant differences for LA and NY as follows:

- LA1: $H < B < G < M$
- LA2: $H < G < B < M$
- NY1: $H < M < B < G$
- NY2: $H < M < G < B$

The Wilcoxon signed-rank test results conclude that ETAs for identical routes are significantly different and are even in a different order for each OD pair.

2.7 Summary

In this study, we conduct an analysis of the recommended routes provided by four popular online maps. A server was set up to query the online maps for four OD pairs every five minutes for 71 days. First, throughout the experiment period, we observed the number of routes that were recommended, when they appeared, and the percentage of the appearance for each one. Second, we averaged the best routes' ETAs provided by each online map, looked at the ETA variability among them, and concluded that each map provider's ETA was significantly different from the others. Third, we viewed how frequently the best route changed over time. Finally, we looked into the ETAs for identical routes provided by different online map providers and also concluded that even the ETAs for identical routes were significantly different. Thus, for example, when a user has a global view of all the routes recommended from different online maps for a single OD pair, we can see that each online map's ETA and recommended routes significantly differ from the others, making it hard for the user to decide which route to take. To that end, route recommendation systems have room to introduce uncertainty in route choice, and thus, it is important to evaluate the accuracy of the data and to improve travel time reliability.

Chapter 3

Real-Time Visual Traffic Information for Pre-trip Route Choice

From today's conventional cars to tomorrow's self-driving cars, advances in technology will enable vehicles to be equipped with more and more-sophisticated sensing devices, such as cameras. As vehicles gain the ability to act as mobile sensors that carry useful traffic information, people and vehicles are sharing sensing data to enhance the driving experience. This chapter describes a vehicular cloud service for route planning, where users collaborate to share traffic images by using their vehicles' on-board cameras. We present the architecture of a collaborative traffic image-sharing system called *Social Vehicle Navigation* (SVN), which allows drivers in the vehicular cloud to report and share visual traffic information called *NaviTweets*. A set of NaviTweets is then filtered, refined, and condensed into a concise, user-friendly snapshot summary of the route of interest, called a *Traffic Digest*. These digests can provide more pertinent and reliable information about the road situation and can complement predictions like estimated time of arrival, thereby supporting users' route decision making. As proof of concept, this research presents the system design and a prototype implementation running on the Android smartphone platform, along with its evaluation.

3.1 Introduction

Advancements in technology are making vehicles smarter as manufacturers continually equip them with on-board computers, global positioning systems (GPSs), collision avoidance systems, and dashboard cameras. Ongoing attempts to alleviate traffic congestion via smart cars use crowdsourced traffic data collected from GPS-equipped devices to determine traffic speed and identify traffic conditions. When provided to navigation systems, this information is used to generate and present a list of recommended routes for trip planning. Such crowdsourcing (or infrastructure-less) services can be classified into two types: push-based and pull-based. The majority of today's systems anonymously pull GPS location information from mobile phone users or navigation systems to generate a live traffic map. The push-based approach depends on social participation, where drivers purposely report traffic information in a richer context (e.g. the location of speed trap cameras, the degree of traffic congestion, and the types of traffic incidents) onto a dedicated server to be redistributed and shared with other drivers. Waze [17] is an example of a navigation app that functions by anonymously pulling GPS data and, at the same time, providing an interface for drivers to push detailed traffic reports.

Conventional ways of reporting traffic conditions have mainly been through the police, transportation officials, drivers on phones, and traffic reporting companies. Nowadays, real-time traffic updates on traffic congestion are becoming widely available and easily accessible via online maps, mobile phones, and GPS-equipped devices. Drivers' route planning can be heavily influenced by such traffic information, which consequently leads them to less congested routes. Such planning is done by selecting a route from a recommended list of alternative routes calculated based on factors like shortest distance or estimated time of arrival (ETA), taking real-time traffic data into account.

ETA is the main deciding factor in route decisions, and this subsequently does not allow the design of vehicle navigation systems to consider other semantically rich information to guide in the decision making and improve satisfaction in the route decisions. For example, if a driver has information that an accident on a certain road will be cleared soon, the driver may choose to stay on the road. But the driver would certainly take a different route if the traffic jam is due to a long-term lane closure. Having this type of traffic information about the road ahead in a timely fashion will alleviate stress and significantly improve the quality of the driving experience.

This chapter highlights the use of geo-tagged traffic images, called NaviTweets, provided by the vehicular cloud to assist drivers in route planning and route decisions. We introduce a vehicular cloud service called Social Vehicular Navigation (SVN), which exploits the mobility of vehicles to expand coverage beyond the limited scope of static sensors, such as traffic cameras. Drivers who are planning a route can opt into the service and request images showing the traffic conditions on the alternative routes ahead. Other drivers whose vehicles are subscribed to the same service collaborate and share their sensing data by uploading NaviTweets concerning the current traffic conditions or any unexpected events. The Internet (central) cloud computes a Traffic Digest that organizes the traffic reports into a user-friendly format to show to the driver and aid the individual in route choice decision-making.

In the near future, rapid development in autonomous or semi-autonomous cars [14, 33] mounted with wide-angle cameras will make such information sharing more pervasive and desirable at the same time. Moreover, hands-free driving experiences will allow drivers to have the time to leisurely participate in sharing traffic information and to engage in more careful route planning.

This chapter is organized as follows. Section 3.2 provides background by presenting related work. Section 3.3 illustrates an example scenario, along with the functions of the proposed

SVN model. Section 3.4 discusses the SVN system design considerations in vehicular clouds, and the prototype implementation is described in Section 3.5. Evaluations on the Traffic Digest and system performance are presented in Section 3.6. Discussed in Section 3.7 and 3.8 are results based on a questionnaire survey and a user study conducted to investigate the real behavior of traffic information sharing, respectively. Future work is discussed in Section 3.9, followed by summary remarks in Section 3.10.

3.2 Related Work

3.2.1 Collaborative Sharing

Drivers can specify their interest in a service, in which other drivers subscribed to the same service can collaborate by sharing necessary information with regard to the request [81, 108]. Gerla *et al.* [44] described Pics-on-wheels, where images taken from on-board car cameras in the vehicular cloud are delivered to the customer on request. Pics-on-wheels is a surveillance service in which several vehicles are selected to take photo images of an urban landscape within a given deadline, as requested by the customer. For example, an insurance company investigating a car accident can request pictures that were taken at a particular location at the time of the incident. Similarly, Hussain *et al.* [55] introduced a service called Vehicle Witnesses as a Service (VWaaS), which utilizes mounted in-car cameras where vehicles act as witnesses to events and provide pictures to law enforcement agencies for forensic purposes. However, the authors mainly focused on the security and privacy of the data exchange between entities [55].

Unlike traditional navigation systems, Waze [17] is a navigation app that collects traffic data from users to provide traffic reports to a central server, where such information is shared with other drivers to provide real-time traffic and road information, such as the volume of traffic, any road hazards, or accidents affecting traffic. Other navigation apps, such as Inrix Traffic [8], have included user-generated traffic reports, and after Google's acquisition of Waze, Google Maps added similar features in its mapping business. Moreover, recent research [38, 92, 89, 60, 67, 66] discussed the potential for traffic information sharing, and we are starting to see greater integration of these techniques.

3.2.2 Route Planning via Traffic Cameras

In the field of transportation, route choice behavior is associated with the decision-making process of route selection, and much research has been conducted to understand this complex behavior [79, 99]. Studies previously did not consider traffic images as part of the criteria for route selection; thus, there has been limited work on their role in route selection behavior. However, there are patent proposals [18, 50] and studies in the literature [75, 94] that apply or identify the usage of traffic photos in route planning. Hanchett [50] proposed a method to install infrastructure that includes a series of camera sensors spaced along major roads to provide traffic images, which are sent to a central station and then distributed to users. Users have a receiver that displays the images so they can preview the route ahead and make route choices. Adam *et al.* [18] proposed a navigation device that displays a route on a road map along with locations where visual traffic information exists. Visual traffic information comes from fixed traffic cameras, and viewing the video feed or still images allows the driver to assess traffic conditions. The Highways Agency in the United Kingdom is making images from traffic cameras available to licensed organizations to provide traffic information to the public to enable better route planning [75, 54]. Speirs and Whitehead [94] presented a research survey to evaluate the influence of providing public access to traffic camera images. Their results showed that combined with other sources of traffic information (e.g. speed/delay information or radio traffic news), traffic images provided an additional secondary level of detail, and they support drivers in making better decisions on their route choice. Real-time traffic camera images are also available to the public in the United States and can be accessed from the corresponding state's 511 website [1].

3.3 Social Vehicle Navigation

3.3.1 Example Scenario

The proposed SVN architecture is hierarchical, which includes three interacting components: (a) the *vehicular cloud* (VC), where a mobile cloud is formed from a group of participating vehicles that collaborate to share its resources, i.e. sensed data from the local environment; each vehicle can opt into the VC and utilize its services; (b) a *roadside unit* (RSU), where the VC establishes wireless connections to the RSU using a vehicle-to-infrastructure (V2I) communications technology, such as cellular, Wi-Fi, or dedicated short range communication (DSRC); and (c) the *Internet cloud* (IC), where the central cloud consists of a group of aggregated servers on the Internet that is responsible for managing the VC's resources; the VC communicates with the IC via the RSU.

Figure 3.1 depicts an example scenario to illustrate the SVN architecture. John commutes to work and prefers to consider safety first when deciding between Route 66 and Route 22. However, information to assess the safety of the roads is not available in current navigators. Therefore, John registers with a vehicular cloud service that allows him to benefit from social feedback shared by other drivers in the VC ahead. John's social navigator accesses traffic updates that are along Route 66 and Route 22 to his destination. Lucy, driving on Route 66, experiences traffic congestion due to an accident ahead and shares this information by posting an image tweet (T1) to the IC via the vehicular cloud service. Similarly, Sam posts a voice tweet (T2) noting that the bridge on Route 22 is slippery, but luckily, there is little traffic. Other

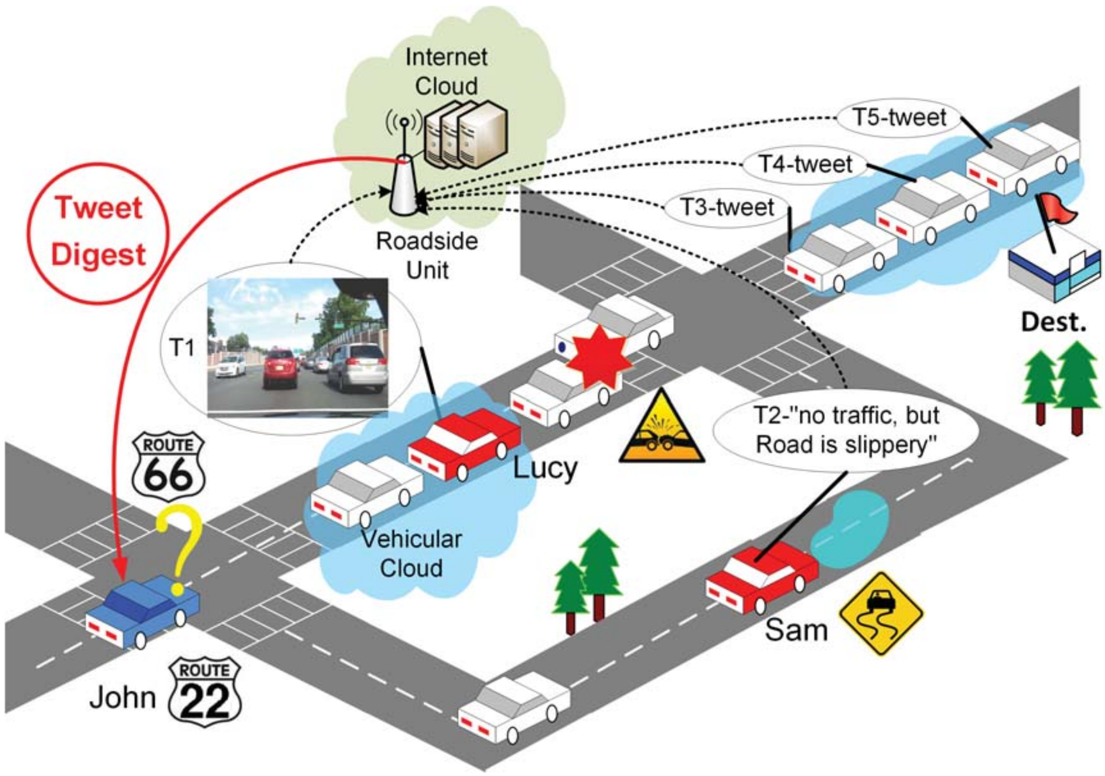


Figure 3.1: Example scenario.

drivers in the vehicular cloud along Route 66 have previously posted tweets (T3-T5) concerning the traffic accident at the same location in front of Lucy. The IC recognizes that T1 and T3 to T5 refer to the same traffic accident, so it discards the older tweets while retaining T1, which is the most up-to-date. The IC then aggregates T1 and T2 into a Traffic Digest and sends it to the querying navigator. John acknowledges both routes' conditions based on the Traffic Digest and plans his route accordingly, where he decides to take Route 66, despite the slow traffic because he prefers a safe, albeit slow journey. Had John's navigation system computed the route based solely on the fastest route, he would have likely taken Route 22.

By using vehicular cloud services, shared real-time sensed data about the environment becomes a possibility. Users can either post or receive other users' real-time sensed data about the traffic in more detail. Then, based on the user's perception of the traffic situation, the navigator can include the driver's preference in the route planning.

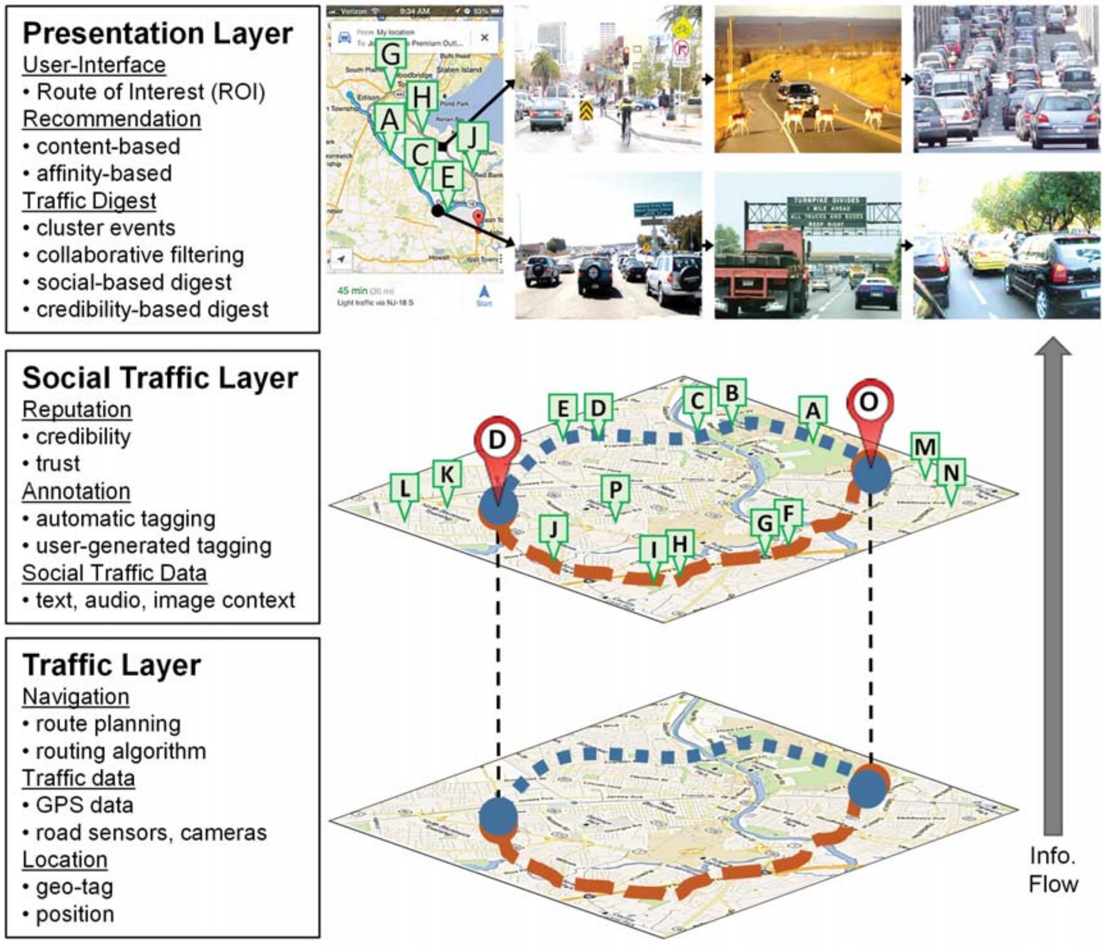


Figure 3.2: Social Vehicular Navigation model.

3.3.2 Layered Architecture Model

Figure 3.2 depicts detailed functions of the proposed SVN model, partitioned into abstraction layers. This is similar to the Open Systems Interconnection (OSI) networking model, where each layer has independent functionalities and passes information to the layer above it.

Traffic Layer

The *Traffic Layer* consists of real-time GPS data from mobile sensors, and traffic data from static road sensors, which are used to update speed or delay information on road segments. Given the origin and destination, the routing algorithm is executed, and ETA is calculated to find a list of alternative routes based on the fastest route or shortest distance. Information regarding route plans and traffic volume for the routes flows to the upper layer.

Social Traffic Layer

The *Social Traffic Layer* consists of social traffic data, i.e. traffic reports in media files, such as text, voice, or images, which are tagged by location and placed on the road network map. Such traffic information is shared by participating drivers in the vehicular cloud who subscribed to it. After the Traffic Layer recommends alternative routes, all corresponding social traffic reports that coincide with the routes are selected and passed to the Presentation Layer. The reason for this is to filter out social traffic reports that are irrelevant to the route the driver is interested in.

Presentation Layer

The role of the *Presentation Layer* is to summarize the information passed from the lower layers. In route selection, users generally seek only information pertinent to their routes of interest. People who receive too much information can easily become overwhelmed and may have difficulty processing the information [79, 88]. It would also be redundant to view similar traffic reports on the route, so the Traffic Digest summarizes the social traffic data set based on geo-tagged locations. For example, in the Social Traffic Layer in Figure 3.2, the blue, square-dot route consists of a set of NaviTweets $\{A, B, C, D, E\}$. The digest summarizes the set, and $\{A, C, E\}$ is displayed, as shown in the Presentation Layer. Details on the mechanism of the Traffic Digest are presented in Section 3.4.3. Also, in the user interface, drivers do not have the luxury of clicking on each geo-tagged image. Thus, to reduce the cognitive load, it is important to provide a user-friendly interface so drivers can sequentially view the information as a virtual tour of the routes, as shown in the Presentation Layer of Figure 3.2.

3.4 System Design

3.4.1 Design Consideration

There are several unique characteristics in a vehicular environment, and we designed the SVN system accordingly, as explained below. The problem domain includes unique features, such as short-time events, increases in traffic communications in congested areas, and data annotation, etc., which provide challenges as well as advantages.

Network Performance

Network performance varies according to the commuting traffic volume, and it is expected that the aggregated network data traffic for uploading and downloading data is associated with peak rush hours. This is under the assumption that higher rates of traffic reporting will occur when roads are congested or during peak rush hours (also shown in the user study results). Thus, communications with the IC should be kept minimal.

Data Annotation

Traffic reports, i.e. NaviTweets can be composed of any type of media file, such as voice messages, photos, or short videos. The size of such data is in the order of kilobytes or, at most, megabytes. Social networking services such as Instagram [9] have shown that uploading and downloading photos via mobile devices can be handled by contemporary network architectures. Photos or videos are taken as raw byte streams and contain little cues to imply the content. Identifying and recognizing the content of the data, e.g. determining whether an image is of an accident or of a moderate to heavy traffic jam, is essential in order to present only pertinent, concise information.

Utilizing recognition techniques, object detection, or user-generated or machine-generated tagging can annotate the data more effectively. Recognition techniques can be used to determine the level of traffic congestion [68, 26] in order to automatically annotate traffic data. Detecting and identifying objects in images or video sequences and annotating events, such as car accidents or the type of road incident, can be accomplished by applying current object detection and identification algorithms [42, 57]. Also, machine-generated annotation can be used to describe the traffic situation in natural language [43]. In our implementation, we provide a platform for users to easily annotate the traffic data they share, i.e. user-generated tagging.

Time-driven Events

A typical traffic event has a time-to-live (TTL) that can last anywhere from several minutes to several days. For example, traffic jams rarely last for more than a few hours, but construction can last for several days or weeks. A TTL feature based on the type of traffic event is incorporated into the design, where data can be safely deleted from the storage/cache system after the TTL expires. Such a feature reduces the requirement to sustain data durability, and means the system needs less storage space. As the average commute time in urban cities is about 30 min [32], we set the TTL to 15 min in our prototype implementation, which is half the average commute time. Further study as to the optimal TTL is left to future work.

Scalability

SVN is designed as a scalable online service for commuters to share traffic information. The architecture should accommodate a large number of users uploading or downloading data simultaneously. A distributed architecture of the ICs (for example, cloudlets [86]) can provide a natural solution to achieve scalability. Moreover, a distributed architecture is favored due to the fact that most commutes occur between home and work within a reasonably stable and small geographic region, where a local IC or cloudlet can handle the majority of requests in its locality.

Consistency vs. Availability

In a distributed system architecture, it is essential to ensure distribution transparency and data consistency. In our SVN system, data consistency can be tolerated for two reasons: drivers generally prefer to obtain up-to-date traffic information, and user-generated shared information can be inaccurate, leading to false negatives or false positives, e.g. false traffic alerts. Hence, emphasizing data consistency among the distributed storage servers is not necessary, and therefore, it is more desirable to provide drivers with timely and inconsistent traffic data rather than outdated and consistent data.

Low Cognitive Load

In order to provide a user-friendly system, it is important to minimize the driver's cognitive load when the driver orders the application to share traffic data, and when the driver needs to comprehend traffic information. The SVN utilizes voice commands or gestures for easy interaction with the application, and the digest of traffic information aids in easier comprehension.

3.4.2 Vehicular Cloud Client

1) Posting NaviTweets

When a user posts a NaviTweet, it is important to gather as much information as possible, while also being able to reduce the cognitive burden on the user. We propose two models for posting: *active mode* and *passive mode*. To minimize the cognitive load, the entire procedure is completed within three commands, where each command is executed by either voice or gesture. Active mode is for users who are actively willing to post NaviTweets. A variable, f, is defined as a threshold where the client device detects potential traffic congestion and takes a picture when f is larger than a predefined value. This value is set by using parameters, such as the current speed, acceleration and deceleration rates, and position. Several car-following models for traffic in stop-and-go conditions [25, 109] can be used to determine a suitable threshold value. Then, the user is prompted to share the image. If agreeing to post the NaviTweet, the user is prompted to annotate it. If agreed to again, a list of recommended tags (e.g. congestion, accident, hazard, construction, and others) is presented for selection via voice command. On the other hand, passive mode is designed for users to post NaviTweets whenever they choose to. The only difference from active mode is that whenever the user wants to share a traffic image, the user can voice-activate the camera. Once the picture is taken, the annotating process is the same as active mode.

The most favorable placements, so the camera can take suitable pictures, is any areas above the vehicle's dashboard. However, the middle area in the front windshield provides the most optimal view.

2) Requesting Traffic Digest

The Traffic Map Layer carries out the route calculation, and based on the routing algorithm, several recommended routes are provided. Users who opt in can ask the VC for the traffic digest on routes of interest. The client device will send to the server the corresponding road segments, and the user will be able to view the digest in a sequential series of events along the route to the destination.

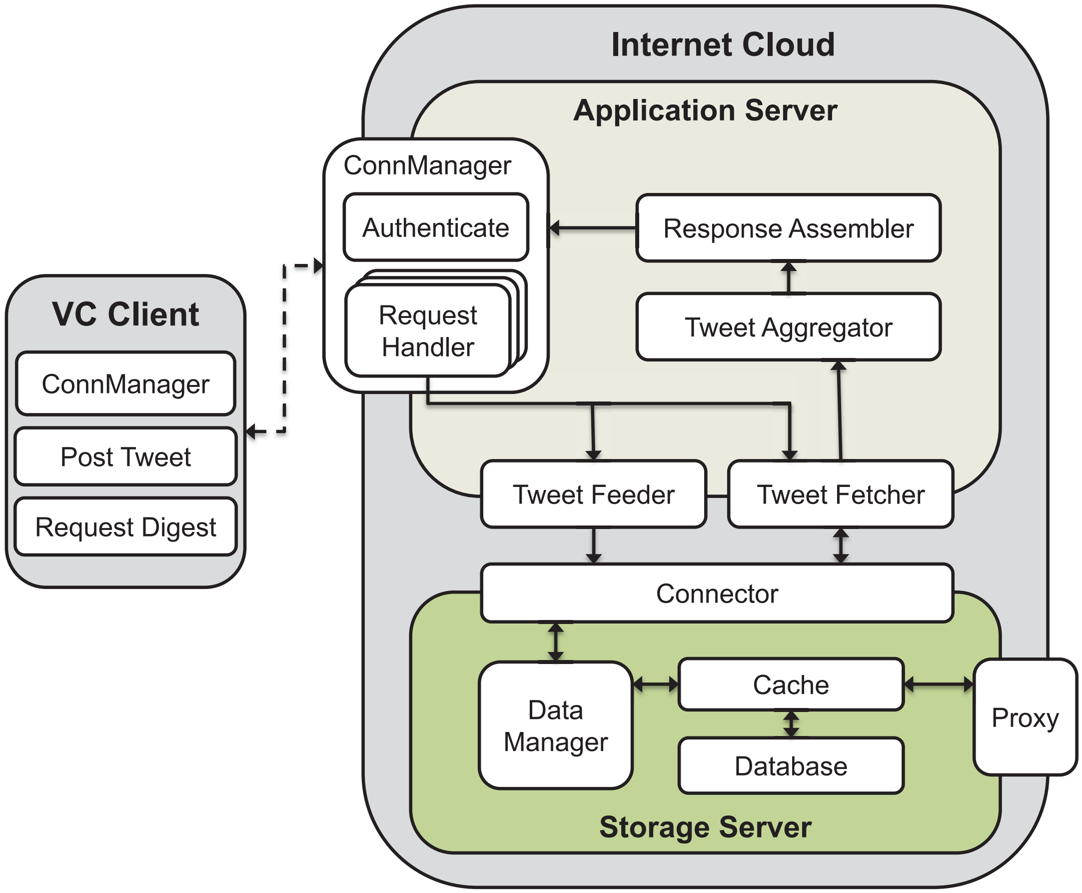


Figure 3.3: System architecture.

3.4.3 Internet Cloud Design

Figure 3.3 presents the design of the system. The application server (AS) sits on top of the storage server (SS) in the Internet cloud, where it receives posts from (or sends digests to) the mobile client subscribed to the VC. The connection manager (ConnManager) authenticates each user and dispatches the job to the handler. Depending on the type of request, the handler updates or retrieves data from the SS. We propose a simple API for communication between the AS and the SS. The SS provides two basic methods to the AS: `put()` and `get()`. The AS is designed to handle two types of request: `post()` and `download()`. Upon receiving post request `T`, the AS simply calls `put(T)`. On the other hand, when a client asks to download a

Traffic Digest, it sends the route of interest to the AS. Upon receiving the digest download request, the AS performs a process of selection, digestion, and composition to satisfy the request. Details are covered in the next following.

1) Cloud Application Server

Receiving NaviTweets

Once the AS receives a tweet, the AS forwards the tweet to the underlying SS via the Tweet Feeder. When a user requests a digest from its associated cloud, the local AS will handle the request by fetching and aggregating the relevant tweets, and finally by responding with a digest containing a series of organized tweets. Since each NaviTweet contains a user-generated or context-inferred tag, the server can index and aggregate the photos based on the tags. We define five tags that can be associated with an image: *traffic*, *accident*, *hazard*, *construction*, and *other*. Moreover, a causal order is defined between images using the tags. A causal order is a happen-before relationship between events x and y. It indicates whether event y is caused or influenced by an earlier event, x. Therefore, it is a semantic causal order relationship. It is safe to assume that drivers prefer to know the reason for a traffic jam. For instance, a lane closed for the remainder of the day leads to a completely different logical predictability of future traffic than from a malfunctioning car. Hence, causal order can also indicate which tweet is more valuable to drivers and should be included in the digest. This is also explained in [93], where knowledge of the causal relationship between a traffic event and slow traffic can be derived. We define the following simple causal relationship between traffic events:

$$x \rightarrow y \text{ (} x \text{ is a cause of } y\text{):}$$

$$Accident \rightarrow Traffic\ ;\ \ Construction \rightarrow Traffic\ ;\ \ Hazard \rightarrow Traffic$$

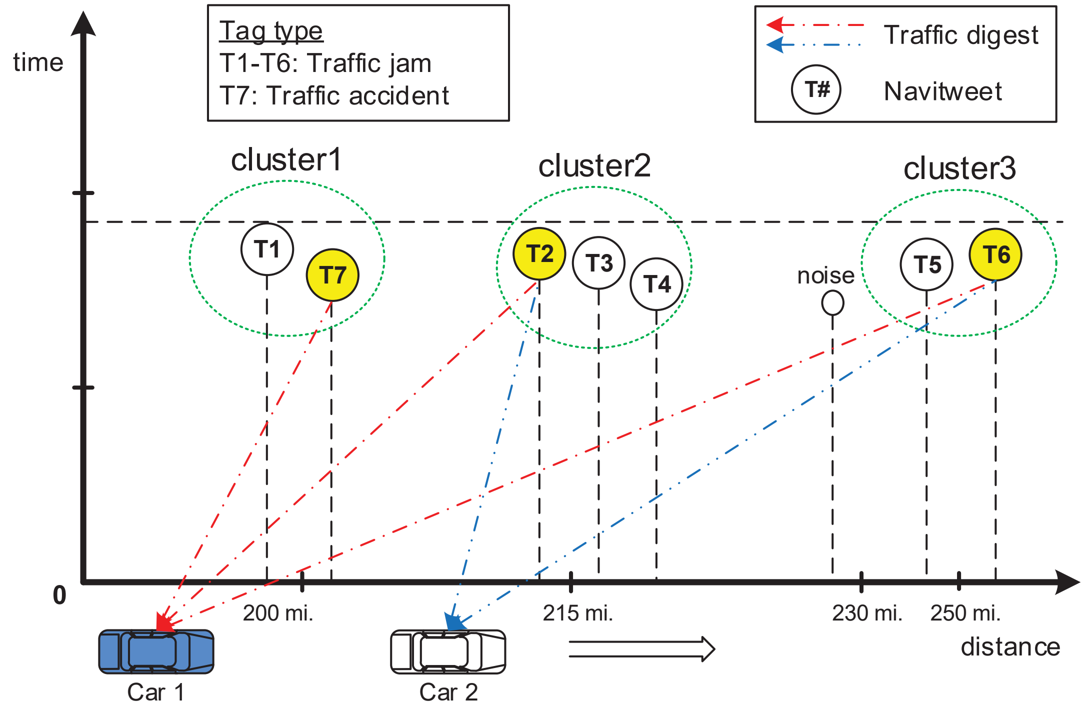


Figure 3.4: Traffic digest. Clustering NaviTweets by location and time. Traffic Digest is based on the casual order of the tag type. T7, T2 and T6 form the Traffic Digest and is sent to Car1.

Traffic Digest

When a user requests a digest, the server will handle the request and respond with a series of NaviTweet results, each of which is a well-formed data structure, containing the location, time, and media content. The Traffic Digest consists of three processes: *selection*, *digestion*, and *composition*. Each process is examined in detail below.

- **Selection.** The Request Handler instructs the Tweet Fetcher to retrieve all the tweets that are posted on the requested road segment within the last TTL. This operation is a simple iterative `get()` call on the SS, with each road segment and the current time as the parameters.

- **Digestion.** The AS then summarizes the potentially large return set from the tweet selections into a readable set of tweets. The resulting set is the Traffic Digest. The motivation for digestion is twofold. First, sending all the relevant tweets to a driver is not feasible, since browsing through them will consume too much time. Secondly, by helping to eliminate redundancy in the tweets, digestion can present drivers with a concise and up-to-date view of the traffic.

The digestion algorithm first assigns each tweet to its corresponding road segment. Assume the length of each road segment has an upper bound. Those road segments that exceed the bound are divided into shorter ones. The algorithm then clusters all the tweets on each road segment using dbscan in a two-dimensional space, i.e. distance and time, to identify possible traffic events that trigger the tweeting. In our prototype, we set the minimum tweets per cluster at two to filter out noise and un-noteworthy events. Another important property of traffic events is that they are transient. As the situation evolves, older tweets reflecting past stages of the event will become less relevant. The latest state of the event is likely to be the most relevant for drivers. Therefore, the digestion first includes tweets with the most recent timestamp in each cluster in the corresponding Tweet Digest. For example, in Figure 3.4, tweets T1, T2, and T6 are the most up-to-date tweets for their respective clusters.

The algorithm then iterates through each road segment to extract and return the selected tweets in each tag category. Moreover, a tweet with the tag "traffic jam" will only be included in the resulting set when there is no tweet tagged accident, construction, or hazard. This rule is used to enforce causal order between the tweets. In other words, if tweet x represents a traffic accident, and tweet y represents a traffic jam on the same road segment, then tweet x is returned while tweet y is not - the reason being that "accident" is more informative than the consequence (traffic jam). Digestion is performed and subsequently produces a concise final set of tweets for the composition stage. Thus, in the first cluster in Figure 3.4, although T1 is more recent than T7, the casual order T7 → T1 results in T7 being included in the Traffic Digest rather than T1.

- **Composition.** The Tweet Digest produced by the Tweet Aggregator is not ordered by either location or time. It is the role of Response Assembler to sort the tweets in increasing distance order and annotate each tweet with location and time. This provides a natural view for drivers as to the reasons for the traffic conditions ahead. The resulting ordered list of NaviTweets is then sent to the client in the VC. Thus, referring to Figure 3.4, Car 1 is shown T7, T2, and T6, whereas Car 2 is shown T2 and T6. In addition, a noise tweet, which is a singular tweet that has no neighbor, will not be selected for a digest.

2) Cloud Storage Server

The SS is the foundation of the Internet cloud, where it is dedicated to provide high-speed, lazy-consistent, and highly available storage services to small media files as well as their metadata. The tweets posted by the clients in the VC will be ultimately stored in the database in the SS. Also, both the metadata used by the Digestion process and the media files used in the Composition process are located in the Database subsystem. We discuss the design of each component in the SS below.

- **Database Separation.** The database in the SS is divided into two components: a metadata database and a file system database. The metadata include the client ID, tag, time, and locations of the media files. The media files corresponding to the metadata are separately located in the file system database. The metadata are used by both Digestion and Composition processes, whereas the media files are only needed for Composition. Thus, the metadata and media files are stored separately. Such separation of databases allows the metadata database to meet the requirements of high-speed search and random access, since the actual media files are not required during the Digestion process.

- **Connections.** Considering the scalability of the SVN system, as well as the locality feature of the request, the SS is designed as a distributed cloud architecture, which means the SS can provide data for the local AS, the remote AS, and the remote SS. The local/remote ASs can access the data in the SS via the Connector shown in Figure 3.3. For each digest request, the AS provides the road segment set and fetches the metadata of the tweet that corresponds to the road segment set. Upon a Composition request, the AS retrieves the media file content that corresponds to the metadata after the digest.

  In cases where the local SS does not have the corresponding media file content of the metadata, but is located in a neighboring cloud SS, we use another connection between these storage servers. The Proxy component is used to enable data sharing between neighboring cloud storage servers. Both metadata and media file content can be sent and received via the Proxy, so that cross-cloud requests can also be handled by the local cloud.

- **Cache Subsystem.** An SS that consists only of databases and connection modules can work alone without a cache subsystem. However, we include the cache subsystem in our design for two reasons. First, the last fetched tweets are usually the most recent ones, and thus, have a higher possibility of being accessed again in the near future. Second, when the required storage space goes beyond a certain threshold, the SS must evolve to a distributed schema, which undoubtedly will increase the response time to fetch data from a neighboring cloud's SS. Therefore, the cache subsystem is deployed between the databases and connection modules, as shown in Figure 3.4. For either metadata access or

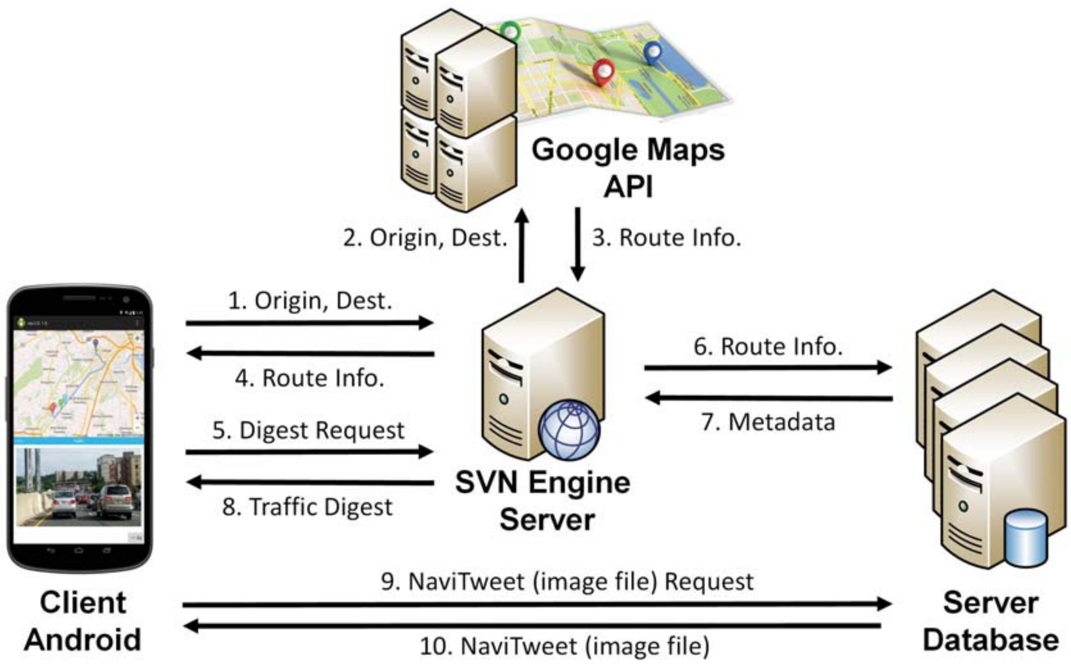


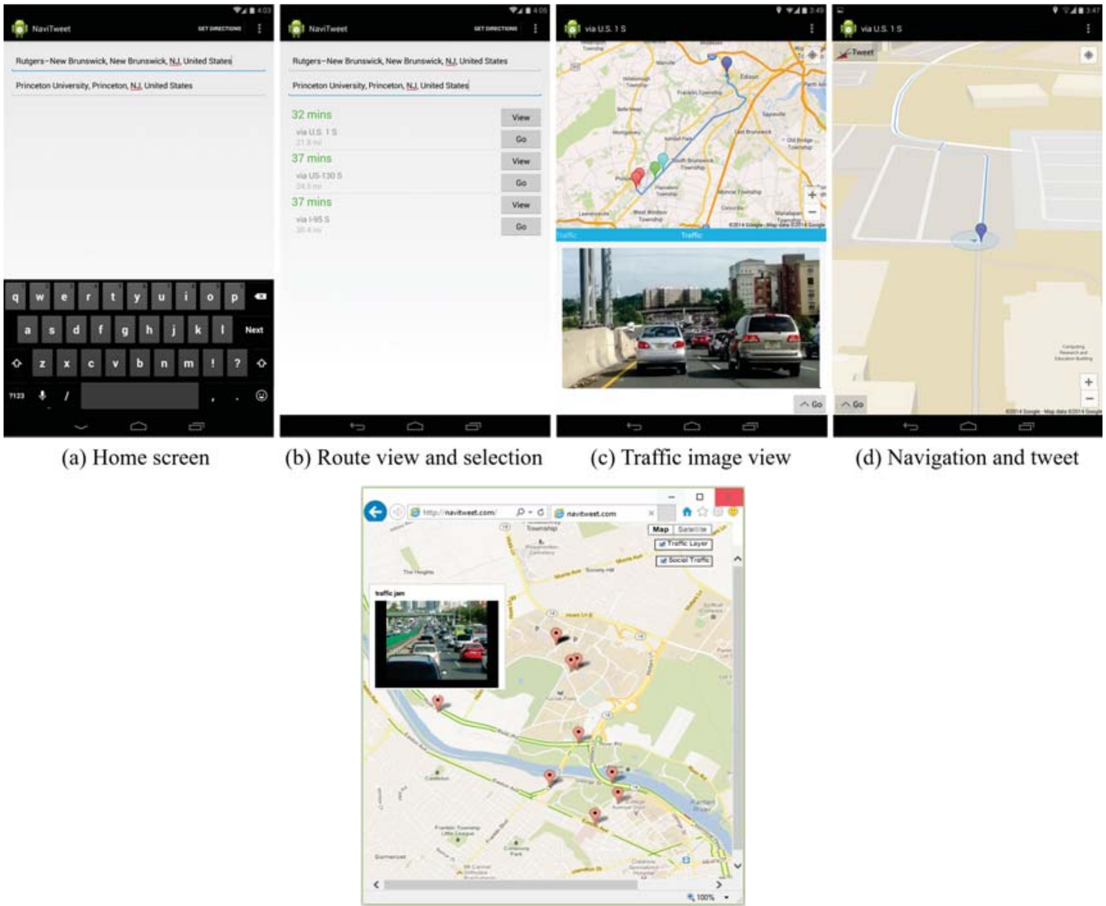
Figure 3.5: SVN prototype system design.

media file access, neither Connector nor Proxy can link to the databases without going through the cache subsystem. Such a design promotes cache hits and reduces the fetch cost from the AS and to the remote SS. In addition, the cache subsystem holds both the metadata and media file contents that are likely to be used in the near future.

3.5 Prototype

3.5.1 Implementation

The prototype SVN system is based on many of the available online services. The SVN client runs on the Andriod 4.1+ mobile platform and uses the Google Places API, the Google Directions API, and the Google Maps API to compute and list the alternative routes, given the origin and destination.



(a) Home screen (b) Route view and selection (c) Traffic image view (d) Navigation and tweet

(e) Web interface

Figure 3.6: Mobile client prototype and SVN web version.

The SVN server engine is currently set up on Ubuntu 12.04 LTS, and the server was configured with Apache 2.2 and PHP 5.3 to process web requests from the client application. In addition, MySQL 5.5 was deployed on the server to store the tweets as well as metadata, such as the timestamp, location, and label categories. The Google Directions API obtains the routes between origin and destination. A set of NaviTweet APIs was developed to provide functionalities for the client. The PostTweet API and the GetDigest API work as mentioned in Section 3.4.3. In addition, the GetTweet API was implemented to support the client in fetching the image itself by using image tags.

3.5.2 System Usage Scenario

In this subsection, a usage scenario for the SVN prototype system is presented to explain the workflow (shown in Figure 3.5) and the user interface (shown in Figure 3.6).

When the user first launches the client application, the origin/destination input screen is shown, seen in Figure 3.6 (a). After inputting the origin and destination, location information for both is transmitted to the SVN engine (Figure 3.5 #1). The SVN engine passes the origin and destination to the Google Maps API and receives the calculated recommended route information (Figure 3.5 #2 and #3). Route information is then returned to the client (Figure 3.5 #4), and thus, the user can see the alternative route list and the corresponding ETAs, seen in Figure 3.6 (b).

Upon selecting a route, users may want to check what traffic events are occurring along the route. In this case, the user presses the "View" button in Figure 3.6 (b), and the request is passed to the SVN engine (Figure 3.5 #5). Based on the route information and the metadata of tweets (Figure 3.5 #6 and #7), the SVN engine computes the digest and returns it to the client (Figure 3.5 #8), which triggers the application to jump to the screen shown in Figure 3.6 (c). The traffic events are geo-tagged on the map, and the image of the first event is shown in the bottom half of the screen. The user can slide the screen to the left or right to navigate through images one by one, and such actions will trigger the client to fetch the next image from the server database (Figure 3.5 #9 and #10). The color of the geo-tag changes from green to cyan as the corresponding image is viewed. (A purple tag is the current location, a red tag is the destination, green tags are the traffic reports, and a cyan tag is the traffic report being viewed.) Finally, after deciding which route to take, the user presses the "Go" button to start the navigation, as shown in Figure 3.6 (d). Also, as shown in Figure 3.6 (e), a desktop web version of the SVN was developed to view the Social Traffic media data shared and geo-tagged

Table 3.1: Input values.

Parameter	Value	Description
λ	28.3	influx rate (1,700 cars/lane/hour)
μ	6	exit rate
p	0.01	prob. of posting a tweet on traffic jam
q	0.1	prob. of posting a tweet on traffic accident
TTL (mins)	15	TTL for all tweets
L (# of cars)	100	length of road segment
N (# of cars)	5,000	total number of cars simulated
V (# of cars)	5	max distance to acknowledge and post a tweet

by users in the VC. As shown in the upper right corner check box, both the traffic congestion data and social media data are viewable.

3.6 Evaluation

3.6.1 Digest Performance

We evaluated the reduction rate in terms of tweet numbers using digestion. Reduction rate is defined as the ratio of tweet numbers with digestion to the number without digestion. In the simulation, we assumed cars move in one direction and on a one-lane highway. In addition, no detour path is assumed to be available, such that the buffer length in this queue is infinite, starting from the location of the traffic accident (i.e. the driver needs to wait until the car can pass through the position where the accident occurred). As a result, the traffic can be modeled as standard M/M/1 queuing system. We further assumed that each car that entered the traffic jam has either a possibility of p (posting a tweet on the traffic jam if the car is at least V cars away from the accident) or a possibility of q (posting a tweet on the traffic accident if the car is at most V cars away from the accident). Table 3.1 lists the values of the parameters in the evaluation. We set an average volume per lane of 1,700 cars/lane/hour [27] as the influx rate.

Figure 3.7 illustrates the change in the reduction rate as the traffic jam progresses, when the

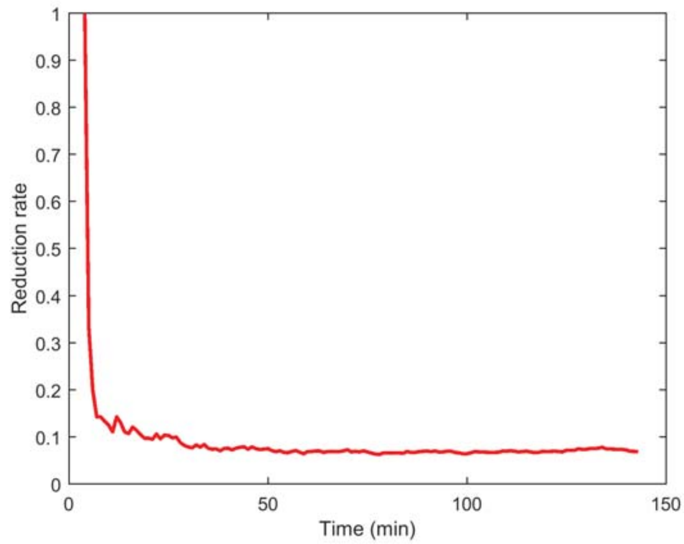


Figure 3.7: Reduction rate varies with time and converges to approximately 6%.

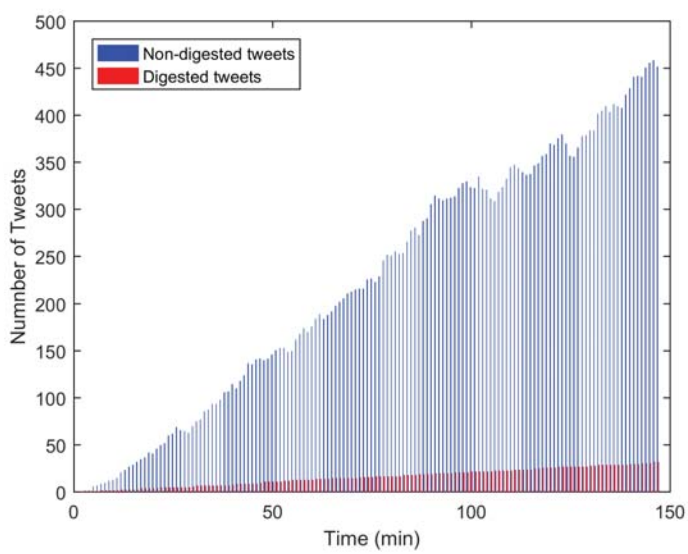


Figure 3.8: Comparison of tweet numbers with and without digestion.

digest algorithm is used. Note that a lower reduction rate actually means better performance, because fewer tweets are contained in the final digest. Figure 3.8 shows that tweet numbers grow to a very large amount if tweets are returned to clients without digestion. For example, after one hour, a driver might need to browse through over 500 images for a big traffic jam. On the other hand, the digest contains only around 40 images, which are equally spread along the jammed road. The resulting savings in reading time and network bandwidth is over a magnitude. There are also two key observations from Figure 3.7. One is that the reduction rate generally decreases over time. The other is that the reduction rate converges at some point. We can show that the converging rate is inversely proportional to road segment length, TTL, and p (under some assumptions) with the following:

$$E\left(\lim_{N\to\infty} \#_{digest}\right) = \frac{N}{L} \times min\Big(L \times TTL \times p, 1\Big) + \varepsilon_1$$

$$E\left(\lim_{N\to\infty} \#_{pre-digest}\right) = N \times TTL \times p + \varepsilon_2$$

and by definition,

$$rr_c = \lim_{N\to\infty} E\left(\frac{\#_{digest}}{\#_{pre-digest}}\right)$$

$$= \frac{E\left(\lim_{N\to\infty} \#_{digest}\right)}{E\left(\lim_{N\to\infty} \#_{pre-digest}\right)} + Cov + \varepsilon_3 \approx \frac{1}{L \times TTL \times p}$$

under the assumption that,

$$L \times TTL \times p \gg 1$$

The assumption above can typically be met by properly setting the road segment length and TTL for a given p. We simulated the impact of two deciding factors on the number of tweets.

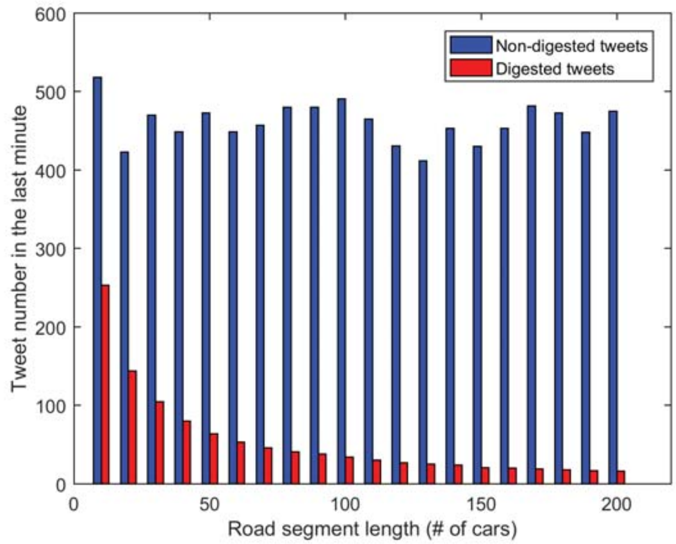


Figure 3.9: Tweet number varies with changing road segment length.

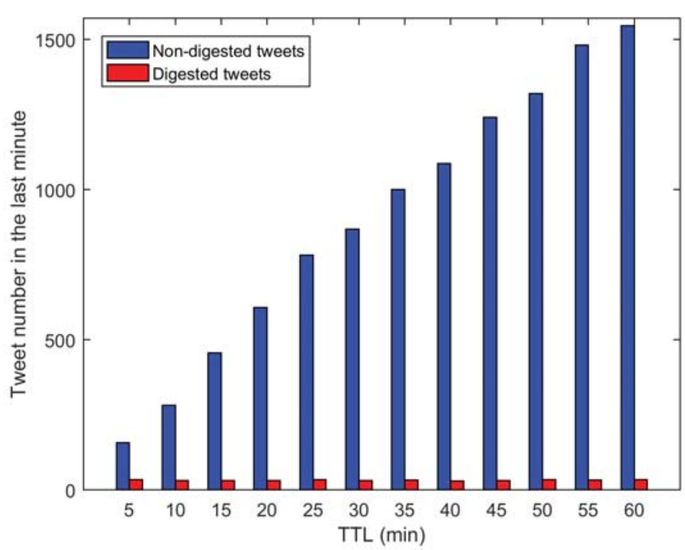


Figure 3.10: Tweet number varies with changing TTL.

Under the default setting, the assumption above is met; therefore, we show that road segment length and TTL are the only deciding factors for the reduction rate, given that p is fixed. Figure 3.9 shows that L only affects the tweet number with digestion. This is intuitively true because with fewer road segments and one tweet for each segment, there is a decreasing number of tweets using digestion. Figure 3.10 illustrates that increasing TTL can significantly increase tweet number without digestion while not affecting tweet number with digestion much at all. Therefore, a good reduction in the tweet number using a digest can be achieved by properly setting road segment length and TTL.

3.6.2 System Performance

In order to explore the feasibility of NaviTweets, we tested our prototype system using simulation experiments. First, the setup of our experimental test bed is presented. With those steps, we tested three main functions (*Tweet Post*, *Tweet Fetch*, and *Tweet Search*) and present the performance evaluations.

Based on the implementation of the prototype SVN system, we used another machine to create multiple test instances emulating smartphone client applications. Each test instance was implemented as a thread that uses the same APIs as those implemented in the SVN application. The client setup was on a Dell machine equipped with an Intel Core i7-3520 CPU @ 2.90 GHz, 8 GB DDR RAM, and a 1TB hard disk. The server side setup was as mentioned in Section 3.5.1. The client/server network connection was over a private 100 Mbps LAN.

By checking the functionality of the SVN system, we can divide the prototype system into three separate functional parts: Tweet Posting, Tweet Searching, and Tweet Fetching. Tweet Posting is the function used when the client application wants to report an event, e.g. post a photo, while Tweet Searching and Tweet Fetching together fulfill the digest task, i.e. when

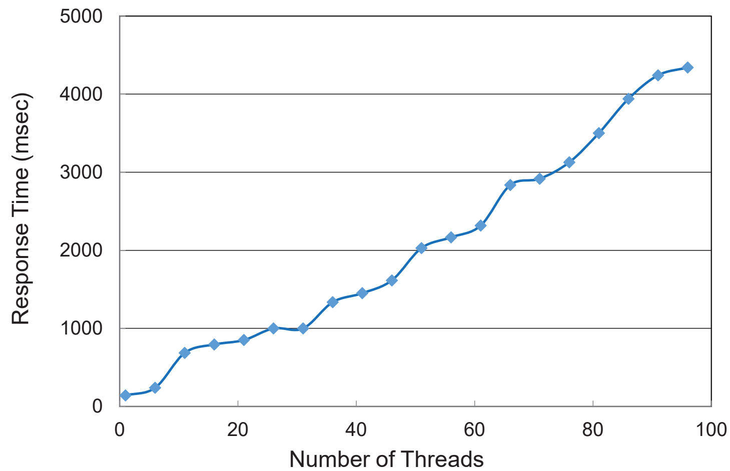


Figure 3.11: Tweet posting response time.

an application user requests a Traffic Digest by providing the origin/destination pair as input. Tweet searching finds the alternative routes and searches the tweets that are relevant to all the road segments contained in the route candidates. If a user wants to check the tweets for details, the client application fetches the corresponding tweets for the user by using the metadata, which is abstracted as a Tweet Fetching function. In the following subsections, performance analysis is conducted individually for these three parts.

1) Experiment Setup

2) Tweet Posting

To test the Tweet Posting function, several experiments were conducted, emulating different workload levels by managing the number of instances of tweet post threads. Every thread had the same image file prepared in memory and posted the image ten times to the server. In addition, each thread inserts a random waiting period of up to 100ms. We measured the response time on the client from posting the request until getting the confirmation from the server. The results under different workloads (i.e. number of concurrent threads) were recorded

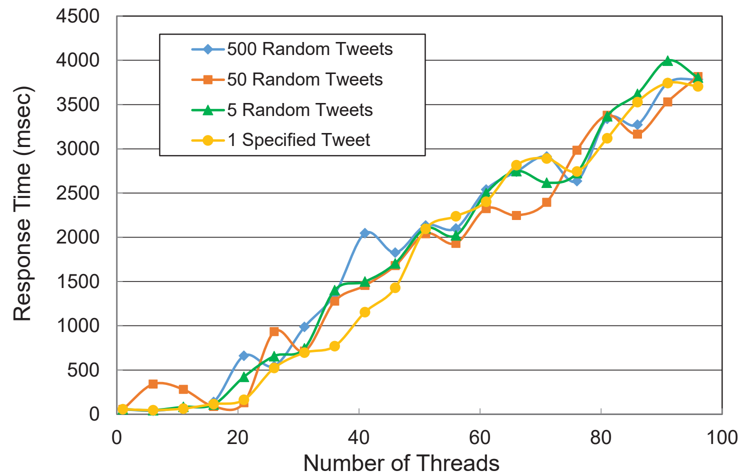


Figure 3.12: Tweet fetching.

and plotted in Figure 3.11.

In Figure 3.11, a linear relationship between the workload and corresponding response time can be observed. In particular, the results show that the prototype system can finish posting ten tweets in one second for each client when the total number of clients is less than 30. Although the response time increases with the number of users, there is no noticeable sign that the system will crash when the number of posting requests reaches some threshold.

3) Tweet Fetching

Similar setups were deployed to evaluate the performance of Tweet Fetching: different workloads were emulated by establishing various numbers of threads in the client desktop. Additionally, each thread was assigned to fetch ten image tweets of the same size from the server, with a random delay of up to 100ms between two consecutive requests.

Compared to Tweet Posting, we face a new question as to whether fetching the same tweet by multiple clients affects response time due to potential access contention or caching in the database. Hence, in this evaluation, we added a new parameter: the range of tweet candidates.

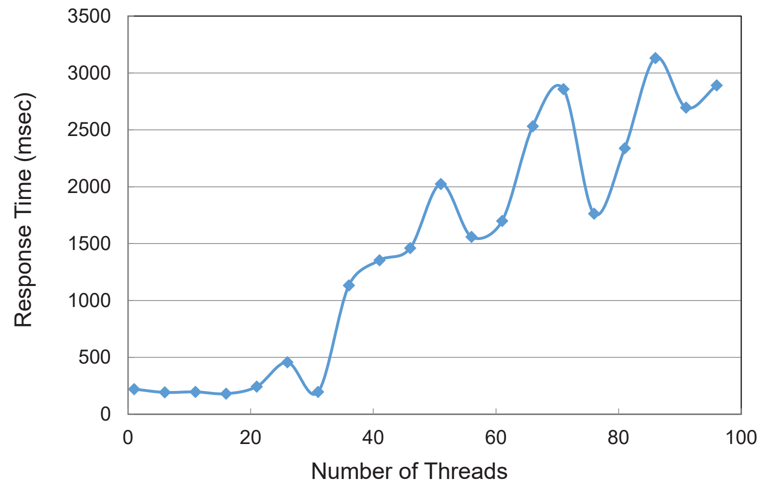


Figure 3.13: Tweet searching response time.

When the number of requests is fixed, random access to a smaller range of tweet candidates means higher probability to access the same tweet.

The results of the experiments are shown in Figure 3.12. The relation between the number of requests and the response time is also linear in general. However, fluctuations in this relationship are more significant than with Tweet Posting. In addition, a different range of settings does not incur meaningful variation. This implies that concurrently reading the same content in the database does not noticeably influence the cost of Tweet Fetching.

4) Tweet Searching

The previous two evaluations tested the two functional modules that mainly deal with the storage systems. However, Tweet Searching is different. Upon receiving the origin/destination pair, the application server fetches the routing information by calling the Google Maps API. Then, for each road segment, the server searches the corresponding tweets in the database and returns the metadata to the client. The cost of this process not only depends on the cost of calling the Google Maps API, but also on the set of road segments as well as the corresponding tweets.

Table 3.2: Tweet searching time corresponding to different origin/destination pairs.

Destination	Distance (miles)	Travel time (mins)	Response time (ms)
Piscataway, NJ	2.9	7	212
Edison, NJ	4.2	11	241
New Brunswick, NJ	4.2	12	180
New York, NY	38.8	48	181
New Haven, CT	119	126	190
Boston, MA	253	251	194

In the first part of this test, we evaluated variation in the response time when given a different number of road segments. We fixed the origin location as the Busch Campus, Rutgers University, and varied the destinations. As a result, both the route segments as well as the corresponding tweet set were changed. For each origin/destination pair, we established a thread that calls the Tweet Searching API ten times and then recorded the average response time of the ten requests.

The destinations' setup and corresponding results are shown in Table 3.2, from which we see that the response time is kept relatively stable, even though the variation in destinations greatly changed the total distance of the road segment and travel time. A possible explanation for this observation is that the tweets are locally distributed in our prototype system: even though a longer route may incur extra search time in the extended area, the vacuity of the tweet set in this area made the cost of additional searches negligible.

According to the results in the first part, there were no noticeable variations in the response times for different origin/destination pairs. Also, the previous subsection shows that potential access contentions have added negligible cost. Thus, when we studied system performance under different workload levels, we chose one of the origin/destination pairs (Busch Campus, Rutgers University to New Brunswick, NJ) as representative. Other origin/destination pairs would show similar results. The simulation setups are similar to those in the evaluations of the first two functions. Similar to previous setups, various numbers of threads were created to emulate the different workload levels; each thread executed ten tweet searches with randomized intervals occurring between two consecutive requests.

In Figure 3.13, the relationship between the response time and the number of threads is shown. Although there is also a linear trend in relations, the fluctuations in response times are even higher than those of the Tweet Fetching function. The reason is that for Tweet Searching, the number of factors that influence the cost is more than for either Tweet Posting or Tweet Fetching, i.e., variations in the costs of calling the Google Map API and the costs for database access resulted in a larger fluctuation. Another finding is that the increasing rate of the response time is smaller compared to Tweet Posting and Tweet Fetching, which implies that when the workload goes beyond a certain threshold, the cost of Tweet Searching will not be the dominant factor for the total cost.

3.7 Questionnaire Study

The motivation for implementing SVN was to develop a crowdsourced application and provide additional detailed information using images to support the driver as to route choice. A user study was designed to answer the following main objectives:

- factors that influence drivers when making initial decisions on route choice.
- whether the use of traffic images influences drivers in route choice.
- occasions that prompt willingness to use traffic images in route choice.

3.7.1 Design and Outline

An online web-based questionnaire using Qualtrics software [12] was developed with guidelines from previous survey questionnaires. Questions were either selected and modified to fit

Table 3.3: Survey questionnaire outline.

Section	Questions
Participant profile	Q1. Gender Q2. Age range Q3. Years of driving
Traffic information source	Q4. Traffic information source preference Q5. Route choice preference Q6. Navigation system preference Q7. Criteria considered in route choice Q8. Occasions for the use of navigation Q9. Accuracy of ETA Q10. Other effective criteria in route choice
Route choice behavior (w/o image)	Q11. Route selection Q12. Rate criteria based on decision in #11
Route choice behavior (w/ image)	Q13. Route selection Q14. Rate criteria based on decision in #13
Influence of traffic images	Q15. Rate how image would support route choice Q16. Uses of traffic images Q17. Influence of traffic images

the study or inferred from results in previous literature based on 1) a study of the decision-making process of drivers considering alternative routes [53], 2) a study on the impact of traffic image camera usage on route choice [94], and 3) a study defining the criteria used in route selection [47].

The survey was divided into four sections: participant profile, preference as to source of traffic information, route choice behavior (with and without traffic images), and finally, the influence of traffic images. Table 3.3 illustrates the outline of the survey sections and the questions. The complete survey questionnaire is shown in Appendix B.

The survey questionnaire consisted of 17 questions, where five had additional sub-question. Questions 1 through 3 collected basic profile information about the participant: gender, age range and years of driving. Questions 4 through 10 asked the participants about their preferences as to the various types of traffic information sources or media, what they rely on when choosing a route (e.g. experience or navigation system), type of navigation system they use, what criteria they currently use when selecting a route (e.g. distance, ETA, name of road or highway, turn-by-turn directions), in what instance they consider using a navigation system, how accurate they thought ETA was, and what other secondary criteria they thought was useful to support them in route selection.

For questions 11 and 12, the participants were shown a map with two routes. The source of the map image was an actual screen shot from the mobile version of Google Maps on a weekday at the 5 pm rush hour, where the origin was the university campus and the destination was a grocery store. Both routes had the same ETA (25 mins), where route 1 was on a highway with a distance of 7.2 miles and route 2 was on a local road at 7.8 miles. Participants were then asked to choose the route they would travel, assuming that the area was unfamiliar. Then, they were asked to rate the criteria they used to make their decision. The criteria Google Maps offered was road or highway name, distance, estimated travel time (ETA), and color-coding roads (green, yellow, red) to indicate the severity of traffic congestion. Questions 13 and 14 asked the same questions as questions 11 and 12, respectively, but images were added to the route. The images were real images taken by the app, where both routes were driven simultaneously by two vehicles at the time when the screen shots were taken, as mentioned above. Participants were then asked to select a route with the additional traffic image information and rate the criteria they used to make their decision. Finally, for questions 15 through 17, the participants were asked about the usefulness of the traffic images and, if the participant was provided with such an app, on what occasions they would use the traffic images, and what influence traffic images would have on their trip.

To measure the participant's preferences, multiple choice questions with single/multiple answers and a Likert-type scale of five points (Strongly Don't Consider; Don't Consider; Neutral; Consider; Strongly Consider) were used for the questions. No information on the NaviTweet

Table 3.4: Participant profile (N = 73)

Question	Response	Count (%)
Gender	Male	42 (58%)
	Female	31 (42%)
Age range	18 - 30 years	56 (77%)
	31 - 40 years	13 (18%)
	41 - 50 years	3 (4%)
	51 - 60 years	1 (1%)
Years of driving	Less than 1 year	10 (14%)
	1 - 3 years	23 (32%)
	4 - 5 years	15 (21%)
	5 - 10 years	17 (23%)
	More than 10 years	8 (11%)

app (usage of traffic images) was mentioned to participants until Question 13. The time to complete the survey was estimated at about 20 mins.

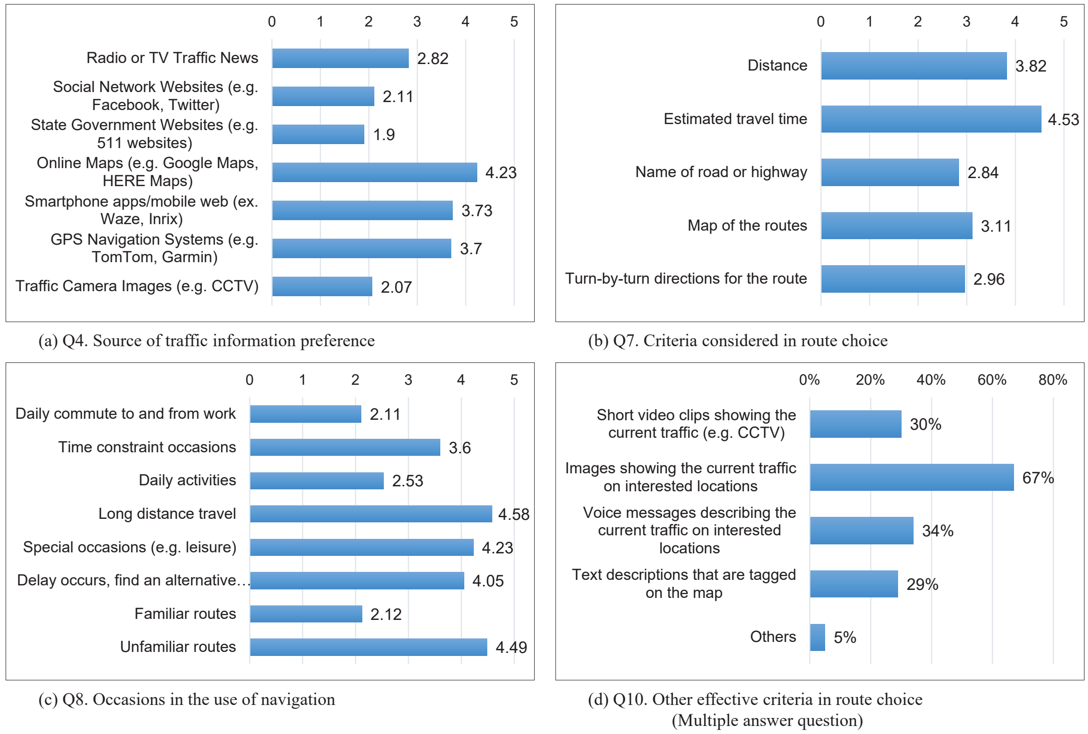
3.7.2 Results

Participant Profile

The web-based questionnaire was distributed to university students and colleagues, and was advertised on social networks. The user study involved a total of 98 respondents, where 73 completed the survey and 25 did not. Survey data from respondents who did not finish the survey were excluded. All respondents resided in the United States and had a driver's license. Table 3.4 summarizes the results for participant profiles.

Traffic Information Source

Participants were asked to indicate the information sources used when they searched for travel information (Q4). Figure 3.14 (a) compares the average rankings of all seven types of information source. Overall, the traffic sources ranked as most considerable were online maps, smartphone apps, and GPS navigation. Also, the participants were asked what they mostly



(a) Q4. Source of traffic information preference

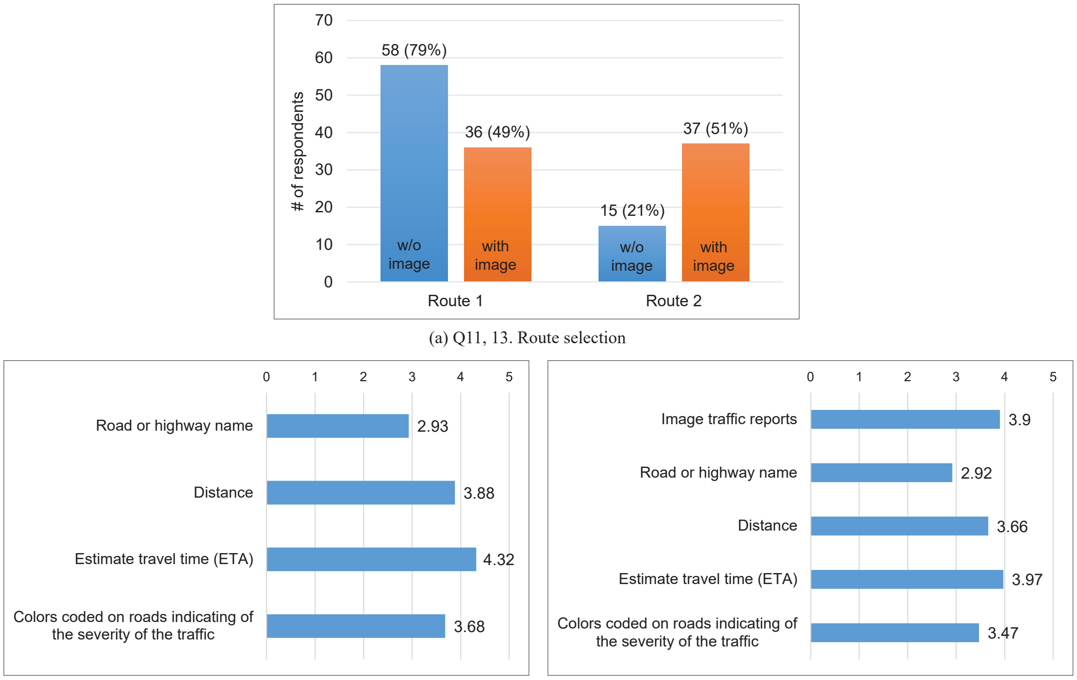
(b) Q7. Criteria considered in route choice

(c) Q8. Occasions in the use of navigation

(d) Q10. Other effective criteria in route choice (Multiple answer question)

Figure 3.14: Survey questionnaire results (Q4, Q7, Q8, Q10).

relied on when choosing a route in terms of percentages (Q5). Approximately 47% were based on past experience, and 52% on recommendations from a navigation system. Only 1% of respondents based reliance on the "others" category (where text entry was allowed), and notable written replies were "directions given by parents/friends/others" or "suggestions from others." The responses for Q6 on the type of navigation the participant used were 55% in-car navigation systems, such as TomTom or Garmin, and it was notable to see that 90% used smartphone navigation apps, such as Google Maps, Map Quest, Waze, etc. In Q7, the participants were asked to rate the criteria when making a route choice. Estimated travel time (ETA) was ranked the highest (4.53) followed by distance (3.82), as shown in Figure 3.14 (b). Q9 asked how likely the participants considered using a navigation system for eight possible instances, and long-distance trips and unfamiliar routes were the top two rankings, as shown in Figure 3.14 (c). This study looked at the impact of traffic images and how they can be used to confirm



(a) Q11, 13. Route selection

(b) Q12. Criteria rate (w/o traffic image)

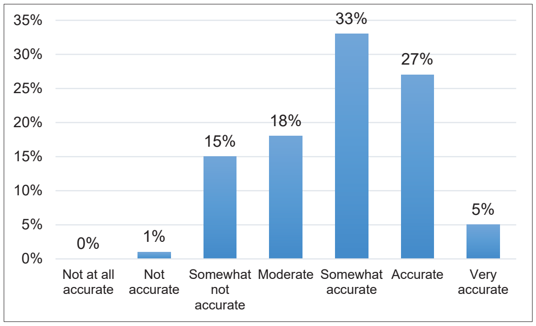
(c) Q14. Criteria rate (with traffic image)

Figure 3.15: Survey questionnaire results (Q11 – Q14).

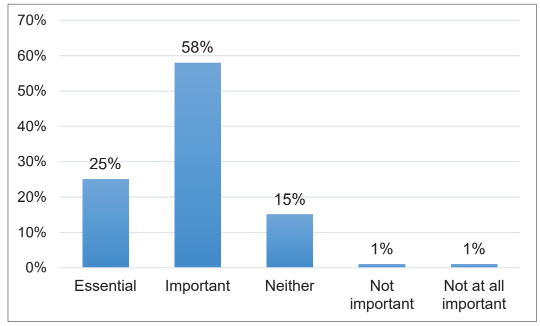
the traffic conditions when there is doubt as to the reliability of ETA, and it ranks what participants thought about the accuracy of ETA. A Likert-type scale of seven points was used for this question, and the results shown are 5%, very accurate; 27%, accurate; and 33%, somewhat accurate. Details are shown in Figure 3.16 (a). Before being provided with any information on the application, the participants were asked what criteria other than the traditional (e.g. ETA, distance) they thought was useful in decisions about route choice (Q10). About 67% answered that traffic images showing the current traffic helped when selecting a route, as seen in Figure 3.14 (d).

Route Choice Behavior

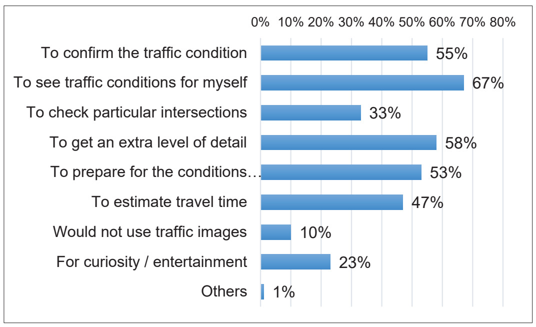
The purpose of Q11 to Q14 was to observe the route choice behavior in a real scenario when users selected a route by using current, traditional route selection criteria versus when users were provided with real traffic images. When users were asked to select either route 1 or route



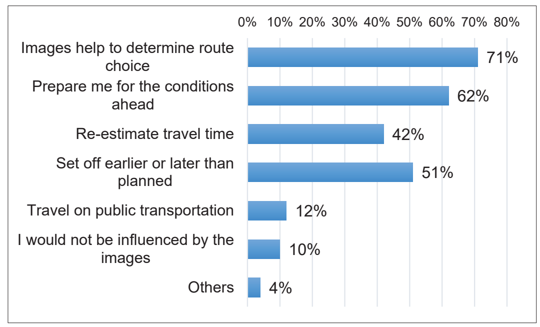
(a) Q9. Accuracy of ETA



(b) Q15. Rate how image would support route choice



(c) Q16. Uses of traffic images (Multiple answer question)



(d) Q17. Influence in traffic images (Multiple answer question)

Figure 3.16: Survey questionnaire results (Q9, Q15, Q16, Q17).

2 by using traditional methods (without traffic images), 79% selected route 1, and 21% selected route 2. Then, participants rated the criteria their decision was based on, and ETA was ranked at 4.32, followed by distance at 3.88. Afterwards, participants were shown the same map and routes along with real traffic images and were asked to select a route. This time, 49% selected route 1 and 51% selected route 2. There was a preference change, with a 30% decrease in selection of route 1 and a 30% increase in selection of route 2 after viewing images. After viewing traffic images and selecting a route, participants were again asked to rate what their decision was based on. The rank for using traffic images was 3.9, and ETA was 3.97 (a 0.35 decrease compared to not using traffic images). Details are shown in Figure 3.15.

Influence of Traffic Images

For Q15, the participants rated how supportive and important traffic images were, with 25% rating them essential and 58% rating them important, as shown in Figure 3.16 (b). Also, to recognize why users would look at real-time traffic images, participants were asked to rate nine items (Q16). The top three items were "To see traffic conditions for myself"; "To get an extra level of detail"; and "To confirm the traffic condition when there is doubt about the reliability of other criteria such as ETA," as shown in Figure 3.16 (c). Finally, participants were asked what influence real-time traffic images would have; 71% responded that images helped to determine route choice, and 62% responded that images prepared them for traffic conditions ahead, as seen in Figure 3.16 (d).

3.8 User Study

The motivation for implementing SVN was to develop a platform where users can easily share traffic data that provides detailed information using images to support drivers for route planning. The best user study approach is to deploy our application to study the real behavior of tweet posting and the efficiency of the Tweet Digest. However, there are several limitations to such a user study. There is the lack of a decent number of traffic images that can be crowdsourced. Also, even if there were to be enough images, there needs to be drivers who take the corresponding route to make use of the images taken.

Therefore, we designed an alternative approach for the field experiment to study users' behavior in tweet posting and using a Tweet Digest. A video was recorded for each of two different routes (the same origin and destination) with a smartphone camera mounted on the windshield of a car when both routes are simultaneously driven by two vehicles beginning at the same start time, as shown in Figure 3.17. Participants were then recruited and instructed to

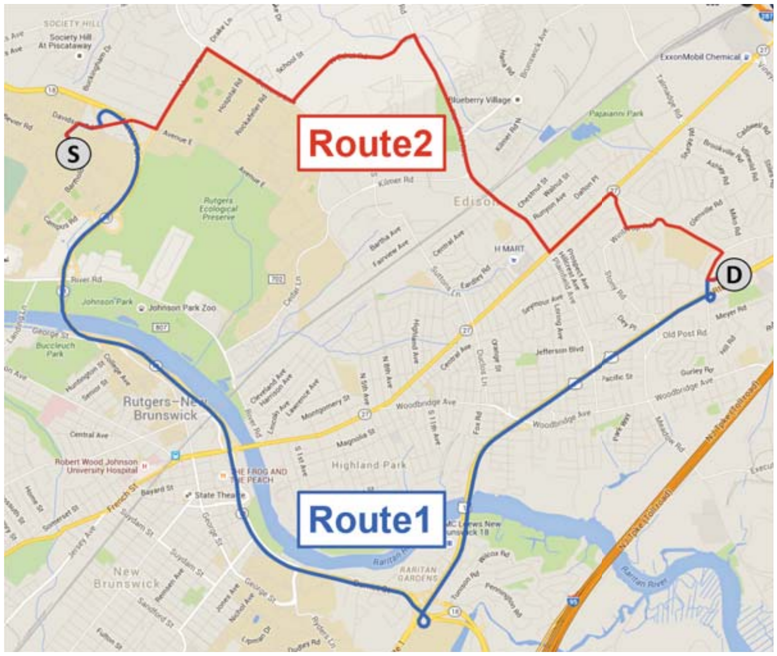


Figure 3.17: Two routes for the user study.

view the two videos (route 1 and route 2) as if they are driving. They recorded the times they felt they would take a photo of the traffic environment to share with other drivers.

3.8.1 Setup and Design

A summary of the details of the user study setup is shown in Table 3.5. The starting location was the university, and the destination was a local grocery store, where route 1 was via the highway and route 2 was via local roads. Both cars set out at 5 p.m., when high-volume traffic was expected due to rush hour. The ETA given with live traffic information was 16 min and 17 min for route 1 and route 2, respectively. However, the actual duration of the travel time was 33 min for route 1 and 18 min for route 2. Route 2's ETA was quite accurate, yet route 1's actual travel time was nearly twice the ETA due to congestion. Also, there were no construction sites or accidents along either route at the time of the experiment (or video recording).

A total of 14 participants (university students with a driver's license) were recruited, and

Table 3.5: Summary details of the user study.

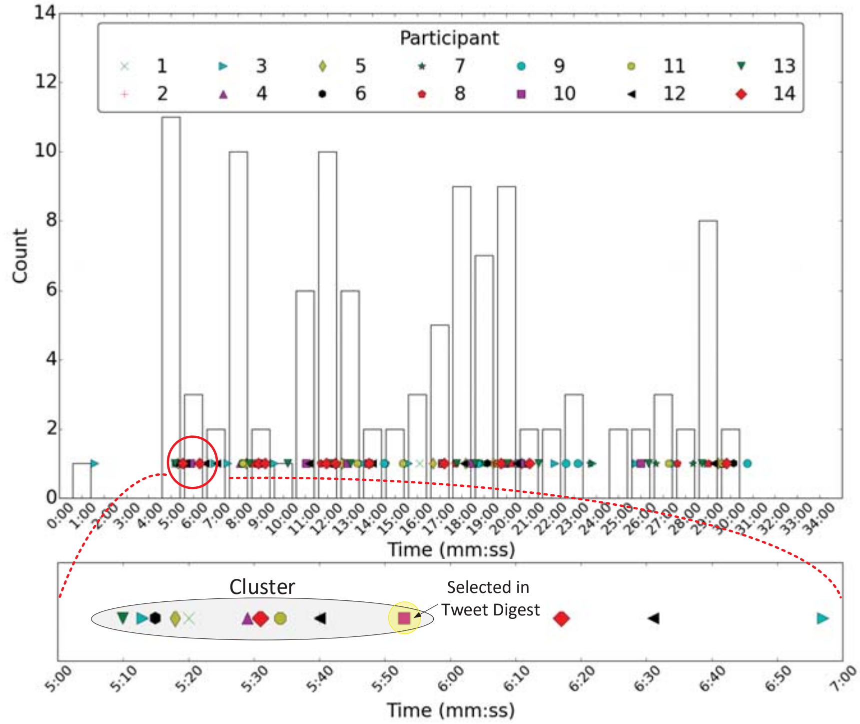
Category	Route 1	Route 2
Distance	8.2 mi	5.9 mi
Estimated travel time (ETA)	16 mins	17 mins
Actual travel time	34 mins	19 mins
Road type	Highway	Local
# of road segments	11	15
Total # of tweets	115	38
Ave # of tweets per participant	8.2	2.7
Min # of tweets per participant	3	0
Max # of tweets per participant	16	6
# of clusters	10	4
# of tweets in the Traffic Digest	9	3

each was instructed in how to use the SVN application. Each participant viewed both videos of the routes and recorded the times they decided to take a photo of the road environment to share with others. Route 2 had a total of 38 tweets, but route 1 had a total of 115 tweets due to traffic congestion. On average, approximately eight tweets and three tweets per person occurred for route 1 and route 2, respectively.

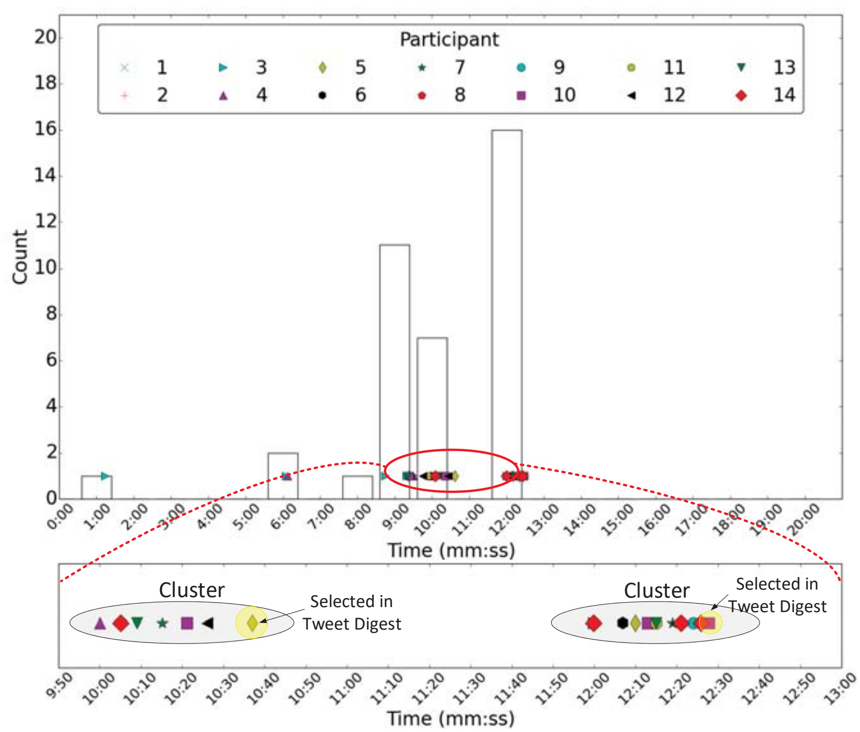
3.8.2 Results

Figure 3.18 (a) shows the results of route 1's tweet count, and the corresponding tweet moments shown (by time) for all of the 14 participants are represented with different markers. After compiling the Traffic Digest, 10 clusters were produced, and the final digest included nine tweets (see Table 3.5). The first cluster containing one tweet was regarded as noise and was omitted from the final digest. Thus, a total of 115 tweets on route 1 was reduced and summarized into nine tweets, a reduction to about one-twelfth of the original number.

An example from the time span between 5 min and 7 min is enlarged in the lower graph of Figure 3.18 (a). This illustrates how the digestion process clusters the tweets, and how tweets are selected from that cluster to be included in the final traffic digest. In this setup, since both



(a) Route 1



(b) Route 2

Figure 3.18: Tweet count, tweet cluster and tweet digest.

routes did not have any unexpected incidents, such as an accident or construction, the tweet with the most updated timestamp was selected in the Tweet Digest because the tag category for all tweets was the same (i.e. traffic jam).

Due to the fact that route 2 did not have much traffic congestion, the total tweets (38) was much less, compared to route 1 (115 tweets). However, as shown in Figure 3.18 (b), route 2's tweets were highly concentrated in two clusters, whereas route 1's tweets were more distributed throughout the route. Route 2's digestion produced a total of four clusters, and the final traffic digest included three tweets (reduced from 38), for a reduction ratio similar to that of route 1 (one-twelfth).

In summary, the user study implies that when drivers are willing to share traffic information (tweet posting), the majority would be sharing road information on the same or a similar event. In our user study setup, for an ordinary travel time of about 20-30 min where about a dozen people share traffic information (from a moderately congested road and a highly congested road), the digest algorithm showed a reduction of about one-twelfth in both cases. Thus, reducing and summarizing this information into a digest will effectively reduce both the shared traffic information and the communication load between the vehicular cloud and the Internet cloud.

3.9 Discussion

Like any other crowdsourcing system, a suitable number of users is necessary for the system to work. Incentives for tweeters, such as "likes" or points, are used in existing apps like Waze [17] so that many users will contribute data, and the system will work. Such mechanisms can also be integrated in our implementation to properly incentivize drivers. For instance, acquired points (or a reputation) for drivers can be used to give priority to their social tweets in the selection for Tweet Digests.

Although digestion can effectively eliminate redundancy, there is no guarantee on the quality of images the digestion algorithm produces. It is quite normal for photos taken from different angles and locations to cause a difference in perceived quality. Therefore, computer vision techniques can be explored to provide a mechanism to select photos with the best quality to represent a traffic event. Also, computer vision algorithms can be tuned to automatically annotate an image as a traffic jam, an accident, etc. Finally, security, privacy, malicious users, and last but not least, passenger safety [65] must also be considered.

3.10 Summary

Advancements in technology have contributed to making cars smarter by equipping them with sophisticated devices, such as cameras and communications devices. Also, significant advances in the production of autonomous cars mounted with wide-angle cameras give most vehicles the potential to gather a significant amount of useful traffic information.

This chapter described a vehicular cloud service for route planning, where cars obtain local traffic information from nearby cars in real time in contexts like text, images, and short videos. However, too much information can make route planning even more difficult when processing it all. As current navigation systems mainly rely on estimated time of arrival, there are limitations to taking other semantically rich information into account to support decision making and improve satisfaction in route selection.

This chapter introduced the use of traffic images provided through the vehicular cloud to assist drivers in route planning and route decisions. We proposed a *Social Vehicular Navigation* system where driver-generated geo-tagged traffic reports can assist other drivers in route planning. The traffic reports are called *NaviTweets*, and summaries are called *Traffic Digests*, which are composed and sent to drivers upon request. The chapter presents the system design and the SVN prototype, along with performance and user-study evaluations.

Chapter 4

Counterfactual Travel Time Information for Post-trip Feedback

"What would have happened, had I taken the alternative route?" This is a question that many drivers often ponder, hoping that their chosen route is the best. This chapter proposes *DoppelDriver*, a system that attempts to answer such questions on non-chosen alternative routes to a given destination by determining actual times of arrival (ATAs) from participatory users on the non-chosen routes. DoppelDriver offers a direct, actual travel time comparison among route choices. With DoppelDriver's collection of actual travel time comparisons, users will be able to make strategic decisions on, and self-assessments of, their route choices. Also, we describe the potential usage and benefits of post feedback (i.e. travel time on non-chosen routes) and how snapshots of travel time comparisons can be used to support strategic decision making on the choice of a route. Using real taxi GPS data, we investigate whether aggregating ATAs for road segments from other users mimics the ATA for the intended origin-to-destination route. Finally, we present our system design for a prototype implemented on the Android platform.

4.1 Introduction

One of the most empowering aspects of advanced navigation systems is their ability to predict the *estimated time of arrival* (ETA) [74, 37] taking real-time traffic flow into account and to provide estimated travel time route comparisons to generate a set of route choices to support the driver in the decision-making. However, when compared to the *actual time of arrival* (ATA), the ETA can easily be under- or overestimated due to, for example, dynamic traffic conditions or the use of simple and/or faulty algorithms. For instance, at the end of or during a trip a curious driver can discover that the ETA given by the online navigator, which he based his initial route selection on, can be far off from the ATA and wonder what would have happened on the other recommend routes that were not taken [46].

This provision of ATA is not unlike what occurs daily in the grocery store. Stuck between choosing two lines in a grocery store, the average customer may choose the line with the shorter number of individuals. Later, as he looks over to the other line, he sees that an individual who stepped in line after him has his goods bagged and walks away. Such a scenario can be replicated across numerous situations, from comparing the time spent in a drive-thru relative to walking inside of a fast-food restaurant, or in choosing the stairs over a slow, crowded elevator to save time.

To the best of our knowledge, up till now, research into predicting accurate ETA based on historic GPS traces or real-time speed and location information from individual drivers have focused only on pre-trip and en route information, while overlooking post information about foregone route alternatives (i.e. actual travel times concerning non-chosen routes). The difference between pre-trip/en route and post information is that the former uses estimated travel times while the latter utilizes actual travel times. Enlightened by the studies in the transportation route-choice behavioral literature [28, 31, 95, 30, 20, 69] that highlight the potential use of post information, this neglect is surprising.

The plausible solution to overcome the limitations of route-choice systems mentioned in Section 1.2.1, and to acquire a higher rate of user compliance, is to *address the misperception*, *motivate drivers to think rationally*, and *combine "experience" with "real-time information"*. In particular, ETA information alone as given by the online navigators may not be sufficient for drivers to make the optimal decision. In addition, this chapter proposes the use of post feedback on the actual travel times for the chosen (i.e. from their own experience) and non-chosen alternative routes (i.e. foregone experience). In fact, the findings in [30] emphasize the importance of post-trip feedback in attaining high compliance rates. They concluded that providing feedback built up the credibility of the information system and achieved high levels of compliance with the system.

Up till now, there has been no systematic way to compare and assess post route choices and to learn from them in order to make more efficient and strategic decisions in the future. Hence, this chapter proposes DoppelDriver, a system that provides comparisons of ATAs for alternative non-chosen routes. The system makes use of other participating users, who act as doppelgängers to provide ATAs for alternative routes of interest to the user. With such feedback, users will have the ability to assess their route choices and log them in a trip diary that can be used in the future for strategic decision-making when selecting a route. To the best of our knowledge, this is the first work that 1) exploits the use of actual travel times from other vehicles on non-chosen routes to be compared to that of the chosen route and 2) provides the ability to assess the route choice and use it as reference for future decision-making.

The outline of this chapter is as follows: Section 4.2 provides the background to post feedback and its potential usage. Section 4.3 presents an intuitive example scenario on how to calculate actual travel time from origin to destination on non-chosen routes. An overview of the DoppelDriver concept, along with its algorithm, is explained in Section 4.4. A feasibility study is conducted in Section 4.5, and the system design and implementation details are presented in Sections 4.6 and 4.7, respectively. Section 4.8 shows the results of a field experiment. Section 4.9 presents a discussion and suggests future work, with the conclusion to this chapter presented Section 4.10.

4.2 Preliminaries and Motivation

As the proposed solution is novel, we particularly emphasize on the potential for post feedback to assess non-chosen routes, which motivates our system's practical usage.

4.2.1 Why Current Information Systems Neglect Post Feedback

Various research has been done on providing quality information to influence drivers' behavior, to guide them to less congested routes, and consequently, to alleviate traffic congestion. However, studies concerning travel time have so far focused only on pre-trip and en-route information, while overlooking post information on forgone payoffs (i.e. travel times concerning non-chosen routes). We argue two key points for the reasons such post feedback is not used in current systems or by service providers such as Google or Waze.

Estimation errors

The estimated error is the measure of information accuracy, defined as the difference between the actual travel time and the estimated travel time. Chen *et al.* [30] studied compliance behavior under real-time traffic information, and their results show that both over- and underestimation errors significantly reduced the likelihood of compliance, although underestimation errors had a greater negative impact than overestimation errors. Therefore, there is no incentive for service providers to incorporate post feedback. However, other studies [112, 28, 30, 69] revealed that post feedback information allows users to compare their actual decisions to the optimal route choices and to assess the quality of their decisions, and emphasizing that evaluation feedback attains high compliance rates.

Regret

Because the recommended route from any information system can turn out not to be the optimal route, post feedback can trigger regret over not choosing the foregone alternative [31, 23]. To avoid regret, people often avoid information about the outcomes of non-chosen alternatives which can be a reason why post feedback is disregarded in current systems. However, very recently, route-choice behavioral studies have started to incorporate regret theory into the travel behavior models. Chorus [31] shows that in repeated choice situations (such as most travel situations), the desire to avoid regret over forgone alternatives means passing up an opportunity to learn. Learning, in turn, implies the opportunity to make better decisions, and hence, lower the levels of regret for subsequent choices. In line with those findings in [31], Ben-Elia *et al.* [23] also indicate the importance of feedback on the perception of regret.

4.2.2 Potential of Post Feedback for Chosen/Non-chosen Routes

Almost all studies in the information technology domain have focused on pre-trip and en route trip information, and because the proposed approach incorporates post feedback on forgone alternatives, we stretched out to find related work in the transportation route-choice behavioral literature. Below, we summarize the findings that emphasize the potential use of post feedback.

In light of the previous studies that highlight the potential use of post feedback, Chen *et al.* [30] developed a model based on empirical studies to capture the effects of commuters' compliance with real-time information by varying the information quality and providing post-trip feedback on actual foregone alternatives. In their experiment, they provided three levels of post-trip feedback: the chosen route (the driver's own experience), the recommended route provided by the system, and the actual best route. They emphasized the importance of post-trip evaluation, where users can compare actual choices to the optimal route choice and assess the quality of their decisions. Their findings show that feedback builds credibility and higher compliance rates with the information system, and thus, they suggest that ITS systems provide such information to users. One of the interesting findings in [95] regarding the effects of feedback was that since users were able to compare their chosen routes with the true optimal choice they were encouraged to select more efficient routes.

Based on the joint relationships among real-time information, post-trip information, learning, and habit, Bogers *et al.* [28] experimented on three scenarios. For all three scenarios, drivers were given real-time traffic information. In scenario 1, post information was based on only the chosen route; scenario 2 provided post information for both chosen and best routes for the latest period; and scenario 3 provided post information for both chosen and best routes for all past periods. Because travelers who receive information on foregone alternatives can also learn about the characteristics of the non-chosen routes, Bogers *et al.* drew the following conclusions. Drivers react primarily to the route queue length information, yet post information helped them learn and improve performance. Also, under the scenario that provided all past travel times for both the chosen and best routes, travelers had significantly better scores versus the other scenarios.

In a recent behavioral model study, Chorus [31] explains that the importance of the acquisition of post information, and the behavioral decision to acquire it, is fundamentally different to the traditional acquisition of pre-trip or en route travel information. Based on psychological theories, when a traveler decides whether to acquire post information concerning repeated route choice, the traveler has to accept a tradeoff between wanting to ignore the potential regret associated with the already executed choice, and wanting to learn in order to minimize potential regret from future decisions. When travelers value the post information it implies that the learning benefits that will reduce levels of regret for the next trip are likely to outweigh any regret associated with the current trip.

4.2.3 Usage of Post Feedback for Chosen/Non-chosen routes

Because ETA serves to determine the travel time based on current traffic conditions, below is an explanation of the usage of post information and how the proposed system provides it.

Counterfactual Thinking

After a decision has been made, the tendency people have is to think about how things would have turned out differently. The "What if?" thoughts that poke at a user given multiple choices is known in psychology parlance as *Counterfactual Thinking* [84, 83, 46]. The concept's premise is that any individual naturally dynamically considers alternatives based on how events, such as the time taken for a certain route to a certain destination, transpire. An example of counterfactual thinking applied to route choice is, "What would have happened on the alternative non-chosen route?" which is the motivation for this work. Counterfactual thinking is anticipated with either *regret* or *rejoice* (i.e. the opposite of regret). For instance, drivers may be stuck in traffic and wonder where they would be, had they been travelling another route. In fact, prior psychology research has asserted that counterfactual thinking is especially activated in the case of negative outcomes. Post information can inform drivers that the other route was even more congested, confirming that the chosen route was actually better, causing rejoice. But regret theory suggests that when provided with post information, drivers may want to ignore the information to avoid regret. In addition to Chorus [31] and Ben-Elia [23], concerning the tradeoffs of balancing regret versus learning, Van Djik and Zeelenberg [102] investigated whether curiosity about an unchosen alternative could overcome potential regret from learning about the unchosen outcome. Their findings show that when information was readily available, curiosity found to overcome regret aversion.

Curiosity and Assessment

In this sense, there is active curiosity and a need for individual assessment on travel-time comparisons with secondary and tertiary routes, which presently remain unfulfilled and unavailable within any web-based or smartphone-based navigation applications or services. Providing real-time travel comparisons with actual travel times for both chosen and non-chosen routes via mobile participatory sensing addresses the inherent user concern activated by counterfactual thinking. Psychology research has shown that humans are predisposed to viewing situations and results from a conditional perspective—that is, to wonder what might have been, if a different choice had been made. Offering all users the chance to answer a nagging curiosity that is innately present, this research has relevance to almost any navigation system user. DoppelDriver allows the user to definitively conclude which route was best, firmly answering the natural human tendency to constantly consider the merits of alternative routes that one could have chosen. Offering all users the chance to answer a nagging curiosity that is innately present, this research has relevance to almost any navigation system user.

Learning

Selecting a route can be considered as a Route Choice Game [96, 51, 107] especially on a day-to-day scenario. In [96, 51], the authors conduct laboratory experiments on humans playing a route choice game. They find that at the beginning individuals tend towards user equilibrium. However, after iterations of repeated games, users often established a coherent behavior.

Iterations of a repeated route choice can be viewed as a learning process. Due to cognitive limitations, people require numerous iterations to detect a pattern, and cannot remember all their choices and what the optimal choice was [28]. Thus, DoppelDriver provides a Trip Diary that logs user selected routes and the true best route, based on the alternative routes' actual travel times. Also, since the trip diary logs all previous commutes that were of interest, this can expedite the learning rate, as shown in an experiment in [69].

DoppelDriver will prove especially useful to commuters or individuals who travel to the same destination more than once. Armed with the ability to see the actual travel times for both the chosen and non-chosen routes, the user is now empowered—he or she can legitimately identify the optimal route, especially when there are similar times for the ETA in the route choice set. For the next commute or trip, the user can make a smarter decision, and confirm that the selected route is truly the fastest each time he or she travels [64].

Personalized Route Choice

In [90], Shiftan *et al.* analyzed drivers' route-choice behavior with real-time pre-trip and post information. Main findings were that risk-seeking individuals tend to prefer a route characterized by a lower average, but greater variance, in travel time. On the other hand, risk-averse individuals preferred lower variance, but better averages. Depending on the route-choice behavior, drivers' preferences can be roughly classified into risk-averse travelers, indecisive travelers, and expected-time minimizers [20]. ETA alone cannot provide users with a personalized choice. Thus, DoppelDriver provides a trip diary that gradually collects personal history of post information with actual travel times on alternative routes from an origin to a destination of particular interest. The trip diary shows the average of, and variance in, the actual travel times, depending on departure time. Drivers can select routes based on their preferences through the actual average travel time and travel time variability for alternative routes.

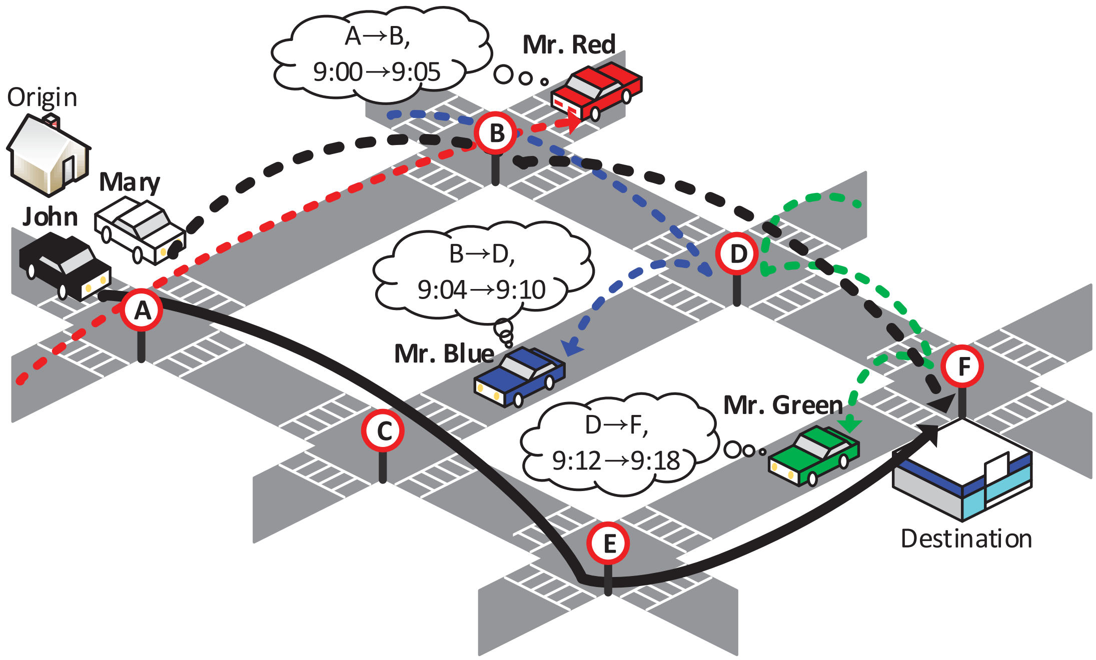


Figure 4.1: Example scenario 1.

4.3 Determining Counterfactual Travel Time

Our algorithm is based on how ordinary people obtain actual travel times on a non-chosen route. If a driver travels from an origin to a destination ($O \rightarrow D$) and would like to compare one's travel time to that of the non-chosen route in order to assess the route choice, the ordinary way would be to ask someone who started at the same time and origin how long it took to travel to their mutual destination. For example, as shown in Figure 4.1, John normally commutes from O to D at 9 am and has two alternative routes, r_1 and r_2. Routes r_1 and r_2 are represented as $A \rightarrow C \rightarrow E \rightarrow F$ and $A \rightarrow B \rightarrow D \rightarrow F$, respectively. John cannot simultaneously travel both routes at the same time, so he would not be able to compare and assess his choice. However, if Mary, John's neighbor, traveled on route r_2 while John traveled on route r_1 to the same destination, they would be able to collaborate to figure out what the best route turned out to be by alternatively driving each route in their daily commute.

Realistically, people would not be willing to go through such a hassle, and it is nearly

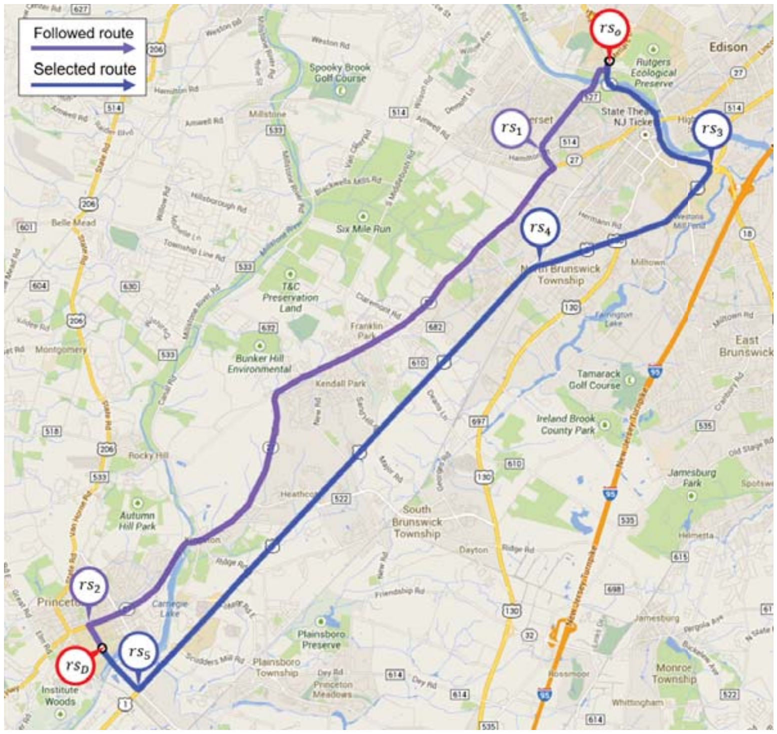


Figure 4.2: Example scenario 2.

impossible to find someone who has the same start time along with the same origin-destination pair. This research provides a solution to this problem by combining portions of three trips: Mr. Red, who passes through road segments $A \rightarrow B$ from 9 to 9:05 am; Mr. Blue who passes through road segments $B \rightarrow D$ at 9:04 to 9:10 am, and Mr. Green who passes through road segments $D \rightarrow F$ from 9:12 to 9:18 am. The aggregation of the actual travel time of Mr. Red, Blue and Green can represent Mary's route r_2. Thus, it is possible to use the aggregation times of road segments traveled by different users to calculate the actual trip time for the alternative route. John can eventually compare the route he took to the alternative route he is interested in. However, selecting the best-fitted participants to provide the actual travel times and whether combining those travel times to represent the actual end-to-end trip is not obvious and thus, this research explores further into the feasibility of the proposed solution.

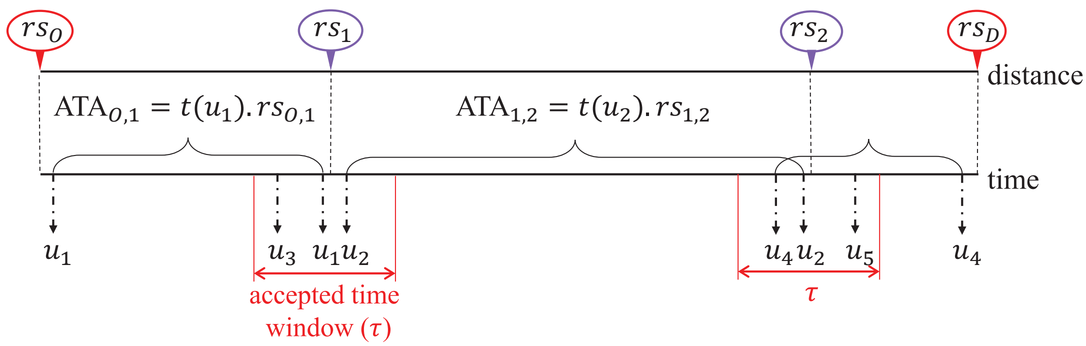


Figure 4.3: Algorithm.

4.4 DoppelDriver Overview

4.4.1 Preliminary

We call our system DoppelDriver, because other participating users act as doppelgängers who are simultaneously "driving" on alternative routes that you are not traveling on. Also, we denote the alternative routes that are of interest (but not taken) as counterfactual route(s). Using Figure 4.2 as an example, below are some preliminaries introducing terms that are used throughout the chapter.

Definition 1: A *waypoint* is a latitude and longitude coordinate set denoted as $wp = (lat, long)$. A GPS point is denoted as $p = (u, wp, ts, o)$ where u is user $u \in U$, wp is the waypoint, ts is the timestamp and $o \in \{1, 0\}$ is the state of passenger occupancy.

Definition 2: A *road segment* is the edge of a road network that does not share a junction. Terminal waypoints, defined as either the start or end of each road segment, are denoted as rs_i. Traversing two terminal waypoints of a road segment is denoted $rs_{i,j} : rs_i \rightarrow rs_j$. Dot notation is used to refer to an element. So, the travel time for $rs_{i,j}$ is $t.rs_{i,j}$, and the start time and end time when traversing from waypoint rs_i to rs_j are $st.rs_{i,j}$ and $et.rs_{i,j}$, respectively. Also, note that $et.rs_{i,j} = st.rs_{j,k}$.

Definition 3: A *route* r is a set of consecutive road segments, $r : rs_1 \rightarrow rs_2 \rightarrow \cdots \rightarrow rs_n$.

The origin and destination point of the route can be represented as r_O and r_D, respectively.

Definition 4: A *trip* *tr* is the trajectory of a sequence of GPS points, i.e. $tr : p_1 \rightarrow p_2 \rightarrow \cdots \rightarrow p_n$.

4.4.2 Concept

We call our system DoppelDriver, because other participating users act doppelgängers who are simultaneously "driving" to a destination on alternative routes that you are not traveling on (because individuals cannot traverse more than one route at a time). The definition of counterfactual is "expressing what has not happened but could, would, or might under differing conditions," so we denote the alternative routes that are of interest (but not taken) as counterfactual route(s).

Once an origin and destination are put into a navigation system, a list of routes is provided to the user, usually ordered incrementally by ETA. Figure 4.3 illustrates two alternative routes from the origin (rs_O), to the destination (rs_D). The user will designate a selected route that he intends to take, and the followed route(s), are the alternative routes to the same destination for comparison of actual travel times. For example, followed route fr consists of four waypoints $rs_O,\ rs_1,\ rs_2,\ rs_D$ and the three road segments make up route $rs_{O,1},\ rs_{1,2},\ rs_{2,D}$.

4.4.3 Algorithm

This subsection explains the DoppleDriver algorithm in detail, along with an example to illustrate it. Figure 4.3 is a simplified form of the counterfactual route $cr : rs_O \rightarrow rs_1 \rightarrow rs_2 \rightarrow rs_D$, to explain the algorithm. First is an explanation of how to elect the best-fitting sharing user for each road segment, and how the end-to-end route travel time is calculated.

Selection Algorithm. Referring to Figure 4.3, the algorithm first checks the initial start time of the querying user at origin rs_O. A sharing user is in the *candidate user* $CU \subseteq U$ if $\{u_m \in U \mid | st(u_{m'}).rs_{O,i} - st(u_m).rs_{O,i}| < \tau\}$. This is interpreted as finding all sharing users who are in the corresponding road segment and nominating those to candidate users if the difference in the initial start time of the querying user and the sharing user falls within a given window time frame, τ. Then, among the candidate users, one is selected as the *elected user* whose difference between the initial start time of the querying user and the start time of candidate user is at a minimum, which can be expressed as $\{cu_m \in U \mid min(|st.rs_{O,i} - st(cu_m).rs_{O,i}|)\}$. To summarize, the selection algorithm can be explained such that it is seeking to elect one sharing user out of many to represent the travel time on the corresponding road segment. The selection algorithm is reiterated until all road segments are covered.

An example is illustrated in Figure 4.3, where if we assume the initial elected user at the origin is u_1, we now find the next elected user at location rs_1. Users u_2 and u_3 fall within the window time frame, and thus, are nominated as candidate users. Among the candidate users, the difference between u_2's start time and u_1's end time is minimal, and thus, u_2 becomes the elected user whose travel time on road segment, $t.rs_{1,2}$ is used. The end-to-end travel time for the entire counterfactual route can be computed as $\sum_{eu \in U} et(eu).rs_{i,i+1} - st(eu).rs_{i,i+1}$ which can be explained as the sum of each elected user's road segment traverse times.

4.5 Feasibility Study

This section presents a feasibility study on real taxi GPS data to study whether the aggregated travel times of road segments can be represented as the total travel time of a complete trip. The methodology, test data settings, and results are described in the subsections below.

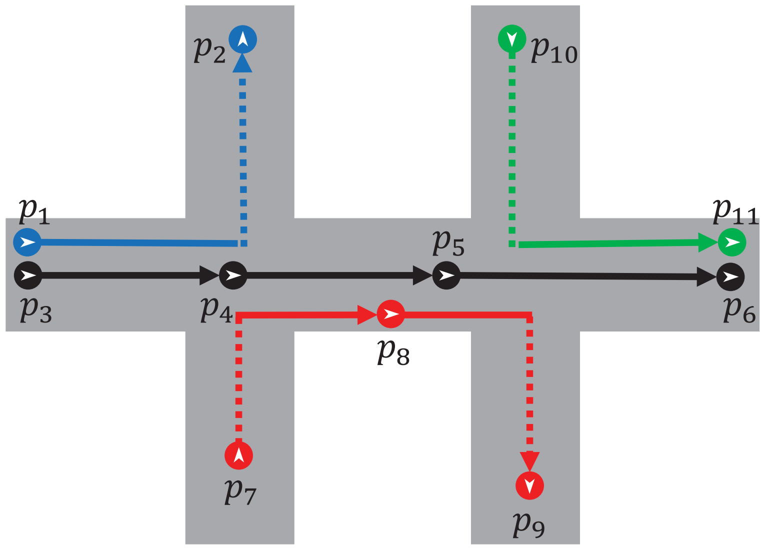


Figure 4.4: Comparison of travel times (Black vs. $\sum$ (Blue, Red, Green)).

4.5.1 Methodology

A simple illustration of the methodology for the feasibility study is shown in Figure 4.4, which compares the travel time of a complete taxi trip as shown with the black solid line, to the sum of the travel times of the other taxi trips shown in solid blue, red, and green lines in each corresponding road segment. The travel time of a taxi corresponding to a road segment is estimated as the portion of the road segment that the taxi traversed. Also, note that GPS trace intervals that exceed 5 mins were filtered out to reduce noise.

The *actual trip* (AT) travel time is defined as the travel time for a complete trip, shown as the black solid line in Figure 4.4. The AT time from its origin and destination for taxi user $u \in U$ is calculated as $t.atr\,(u) = \sum_{i=O}^{D-1} (ts.p_{i+1} - ts.p_i)$. The *counterpart trip* (CT) travel time is defined as the travel times for all road segments that correspond to the actual trip (i.e. the solid portion of the blue, red, and green lines in Figure 4.4), which is calculated as $t.ctr\,(u) = \sum_{i=O}^{D-1} \left(t.rs^{i}_{i+1}.(\neg u)\right)$, $\neg u \in U$, i.e. the traverse time for all the road segments from taxi users who are not the corresponding taxi user from the AT. An example of the counterpart trip *ctr* is

Table 4.1: Taxi data set.

Dataset	Non-rush	Rush
# of taxis (u)	7,288	8,093
# of trips	2,049	2,797
# of road segments traversed	60,551	71,209
# of GPS points	128,826	147,008
mean road segments traversed	29.55	25.46
standard deviation of road segments traversed	15.30	15.47

illustrated as the aggregated travel time for road segments of the solid blue, red, and green lines shown in Figure 4.4.

4.5.2 Test Data

An empirical feasibility study was carried out using large-scale taxi GPS point sets from Beijing on January 6, 2009, from 1 pm to 2 pm and 5 pm to 6 pm, which are denoted as non-rush hour and rush hour, respectively. We employ a set of methods to preprocess the GPS data set, such as data cleansing, map matching, identifying trajectory and travel time estimations [78, 72, 110]. Preprocessing the taxi data requires a great extent of work before using it for the analysis, however, this topic is beyond the scope of this research, thus the details of the preprocessing techniques are omitted. A summary of the taxi data set and the descriptive statistics for both rush hour and non-rush hour periods are shown in Table 4.1.

4.5.3 Analysis

Comparison of the aggregated road segment travel times was analyzed with respect to the actual trip time, i.e. counterpart trip travel time versus the actual trip travel time. We first determined the distribution, if any, of the travel times and then based on the distribution, we decided what statistical inference to apply.

For each AT time, we obtained the corresponding CT time. If there were at least a single

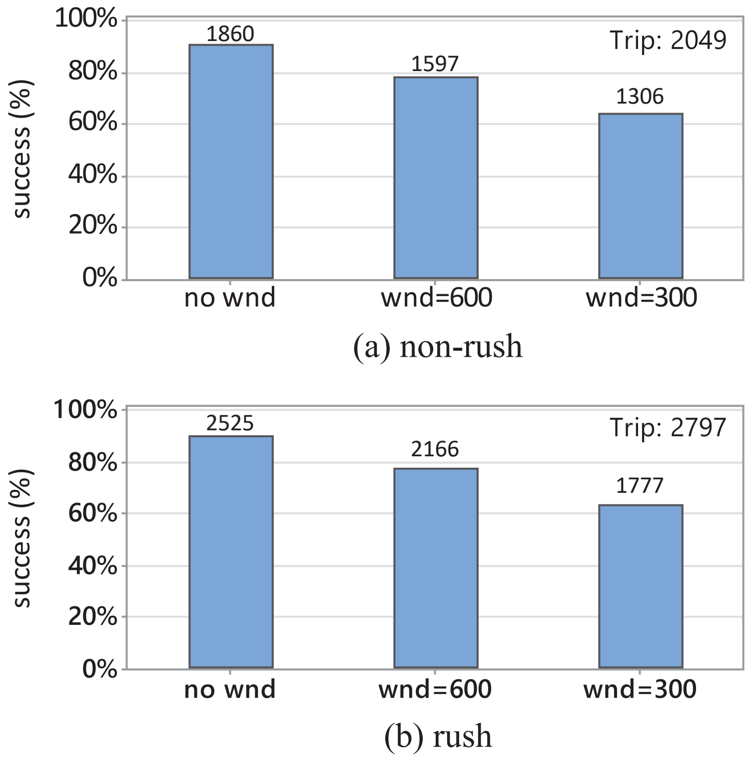
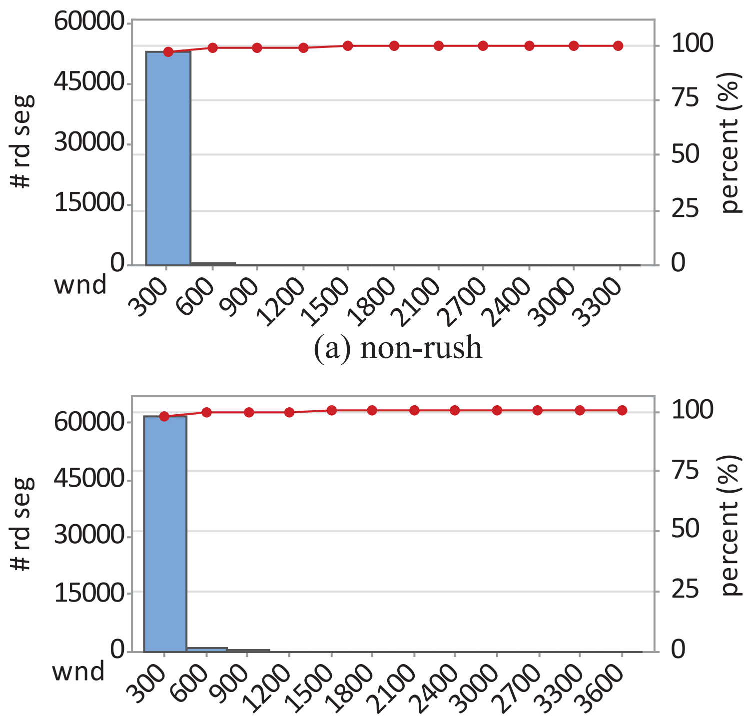


Figure 4.5: Acquiring CT success rate.

road segment that could not be acquired for the complete CT, it was acknowledged as failure to get a CT time. When there is no defined window size (*wnd*), the success rate was 90.8% and 90.3% for non-rush and rush hour traffic, respectively. For a window size set at 600 sec, the success rate was 77.9% and 77.4%, and for a window size set at 300 sec, the success rate was 63.7% and 63.5% for non-rush and rush hour traffic, respectively. As seen in Figure 4.5, as the window size decreases, the success rate also decreases. The taxi data set accounts for approximately 7% of the total traffic. Even with such a low volume of taxis, the results are somewhat satisfying, and the success rate will probably increase dramatically as more users participate.

With no defined window size, we first determined the distribution of window sizes. Figure 4.6 shows that the majority, i.e. approximately 97%, fall within a window size of 300 sec for both non-rush and rush periods. This plot shows the distribution of window size for non-rush



(a) non-rush

(b) rush

Figure 4.6: Distribution of window size.

and rush hour periods for periods where acquiring the travel times for the CT road segments was successful with respect to the AT. For the non-rush periods, the cumulative percentage for *wnd* $\leq$ 300 was 97.9%, and for *wnd* $\leq$ 600, it was 99.2%. For rush periods, the cumulative percentage for *wnd* $\leq$ 300 was 97.4% and for *wnd* $\leq$ 600, it was 99%. This shows that for both non-rush and rush periods, the majority of the window sizes fall under 300 sec.

For each trip where acquiring a CT failed, the number of road segments that were affected by the failure was determined. The distribution is shown in Figure 4.7. For non-rush periods with *wnd* = 300, 46% of the total trips that failed to acquire a CT were due to a failure to obtain just one road segment, and a cumulative 90% were due to a failure to acquire up to five road segments. For *wnd* = 600 in non-rush periods, 50% were due to failure to acquire one road segment's travel time. For rush periods with *wnd* = 300, 42.16% of CT failures were due to lack of one road segment, and similar with non-rush; a cumulative 90% were from failing to

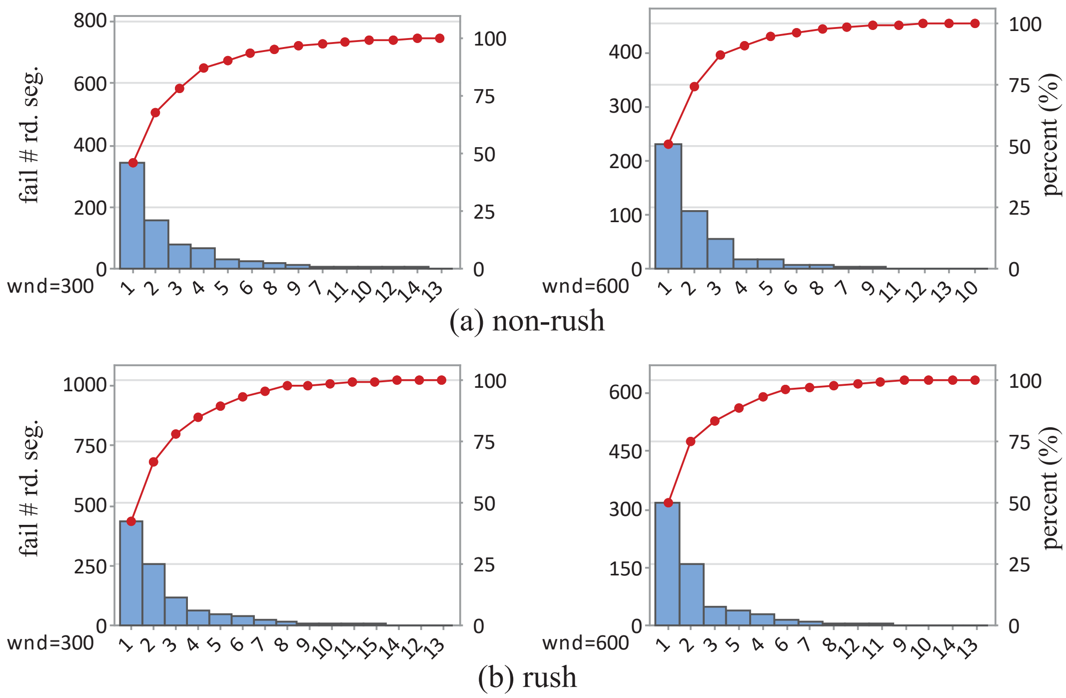
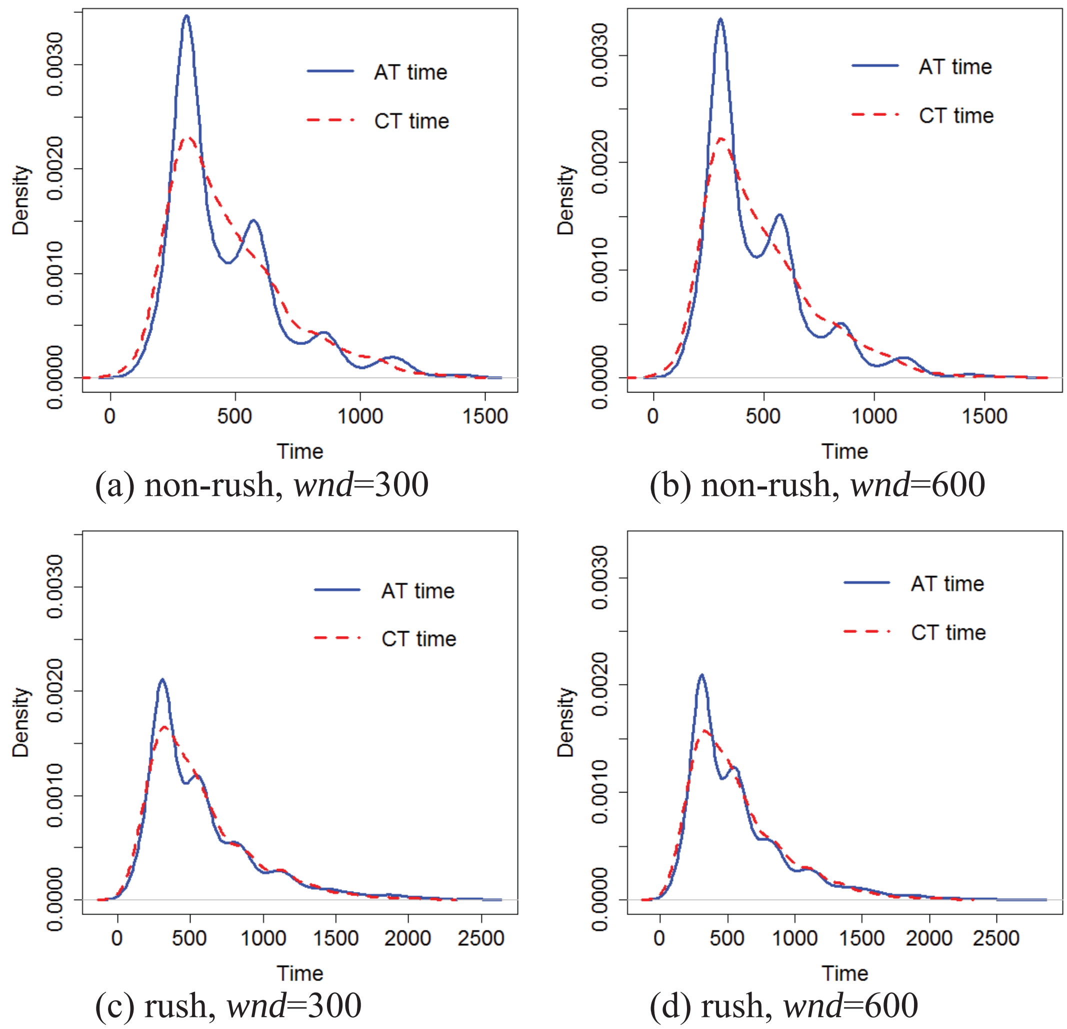


Figure 4.7: Distributions of the road segments affected by CT failures.

acquire up to five road segments. For rush periods with $wnd = 600$, 50.24% of CT failures were due to lack of one road segment, and 93.5% were due to a failure to acquire up to five road segments. Recall from Table 4.1 that the mean number of road segments traversed was 29.55 and 25.46 for non-rush and rush periods, respectively. This can be interpreted as for an average of 30 road segments traversed, 50% of the CT failures were due to missing one road segment and 90% of CT failures were due to missing up to five road segments.

The data set was retrieved for AT times and CT times for those trips that successfully acquired a CT time. Then, kernel smoothing was applied to the distributions of both AT and CT times to visually inspect the distributions, which are shown in Figure 4.8. To determine if the AT and CT data sets were normally distributed, a few normality tests were conducted, such as the Kolmogorov-Smirnov test using R statistical programming [13]. The p-value for all tests on the data sets was < 0.02, and thus, the result was that the distribution is not normal. Also



(a) non-rush, *wnd*=300 (b) non-rush, *wnd*=600

(c) rush, *wnd*=300 (d) rush, *wnd*=600

Figure 4.8: Distribution of AT time and CT time.

conducted were a log-normal test and a chi-squared test, and the results showed that the data set distributions did not fit (p-value<0.001). Such results are not unusual. Equivalent results are also shown in other previous work [98].

A paired equivalence test was applied, which was used to test whether two paired samples are nearly equivalent as to some outcome, such that any difference is insignificant. The paired difference of AT and CT time was computed, and distributions are shown in Figure 4.9. All figures, i.e. non-rush and rush periods, show that the paired differences of ATs and CTs are concentrated near 0, meaning that the differences are trivial. Thus, also conducted were a

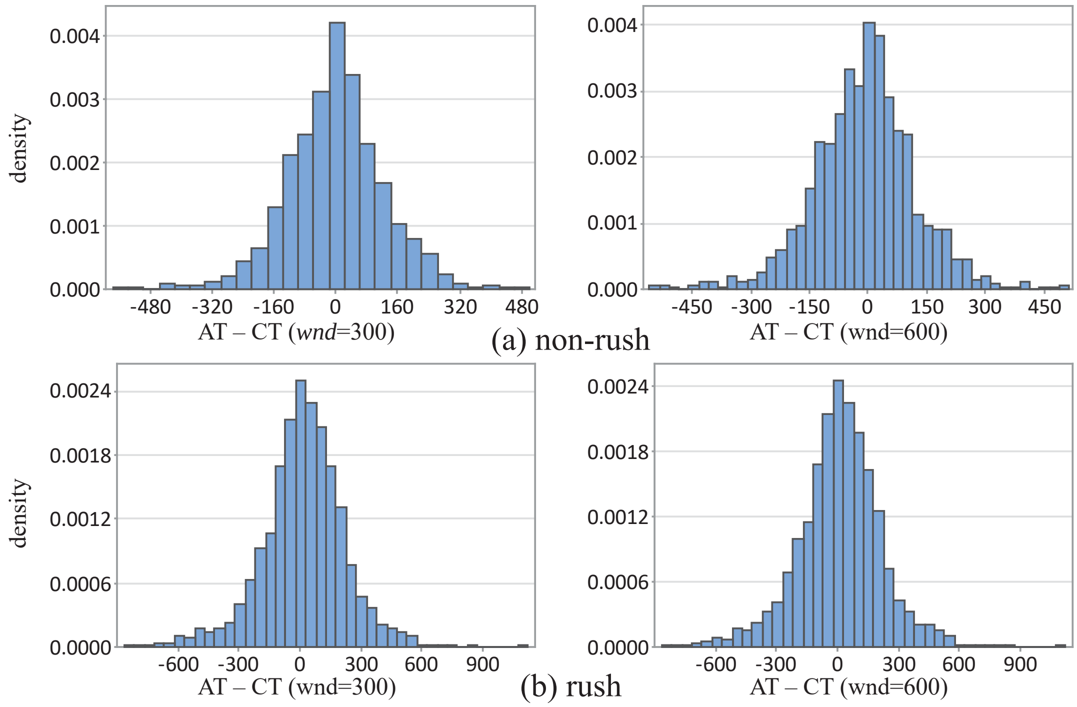


Figure 4.9: Distributions of the differences between AT time and CT time.

non-parametric statistical hypothesis test and a Wilcoxon signed rank test, which assumed as the null hypothesis that the difference between AT and CT is within δ sec, and the alternative hypothesis is where the difference is greater than δ sec, as shown below.

$$H_0 : |\mu_{AT} - \mu_{CP}| \leq \delta, \quad H_A : |\mu_{AT} - \mu_{CP}| > \delta$$

Confidence interval (CI), α was set to 95%, and δ was set to 30 sec, such that 30 sec represents a reasonable figure when saying AT and CT times are equivalent. The results are shown in Table 4.2. It is clear that for non-rush periods, the median difference between AT and CT for the same trip is -4 sec, where differences range between -10.5 and 2.5 sec. For rush periods, the median difference was 16.5 sec for intervals between 8.0 and 25.0 sec. For the equivalence test, all accepted the null hypothesis, meaning that the differences between AT and CT are within 30 sec.

The 95% CI for the Wilcoxon signed rank test was tested for both window sizes on non-rush

Table 4.2: Summary results.

	Non-rush		Rush	
wnd size	*wnd*=300	*wnd*=600	*wnd*=300	*wnd*=600
N	1306	1597	1777	2166
est. median	-4.0	-10.0	16.5	9.5
CI	95.0	95.0	95.0	95.0
CI lower	-10.5	-16.0	8.0	1.5
CI upper	2.5	-4.0	25.0	17.0
equivalence	Yes		Yes	
p-value	≈ 1		≈ 1	

and rush periods. The CIs all fall within the equivalence interval of (-30, 30) accepting the null hypothesis with $p \approx 1$, and thus, it can be concluded that the actual trip times and counterpart trip times are equivalent.

Based on the taxi data, the feasibility study showed that depending on the window size, the success rates to acquire a CT were approximately 80% and 65% for window sizes of 600 sec and 300 sec, respectively, whether they were non-rush or rush periods. Also, for trips that failed to acquire a CT time, approximately 40 to 50% were due to a failure to obtain just one road segment. The distributions of the differences between AT and CT were concentrated near 0 and, based on hypothesis testing, the difference was trivial. Therefore, we conclude that it is possible to aggregate travel time for a route by using segment travel times, and the aggregated travel time of road segments can be represented as the total travel time of a complete trip.

4.6 DoppelDriver System

Figure 4.10illustrates the client/server architecture of the DoppelDriver system. The client-side interface interacts with the Maps API for route recommendations and information, which are passed to the server via the client, along with its location, as sensor data. Location and timestamp data are then used to calculate the travel time for each road segment on the counterfactual

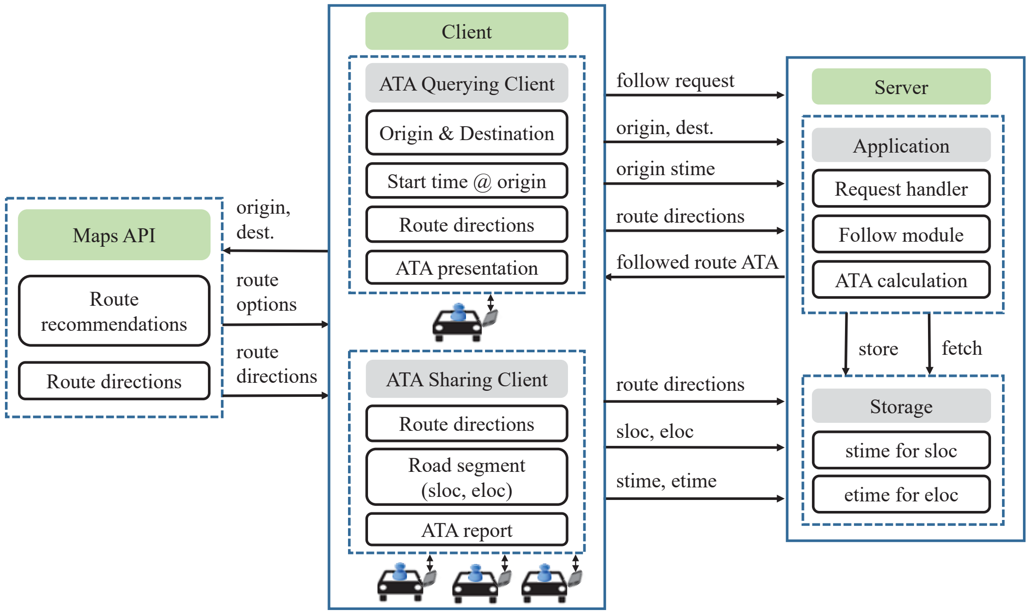


Figure 4.10: DoppelDriver system.

route. The following subsections discuss in detail the primary role of each entity.

4.6.1 Client

As shown in Figure 4.10, a client user can be a sharing client, a querying client, or both. The client first inputs the desired destination, and the mobile device will send the origin and destination to the Maps API. The Maps API will then compute the route and recommend a list of route options to the client. Once the user chooses the selected and counterfactual route, route directions (i.e. waypoint and road segment information) for the routes are passed to the server.

Querying Client

A querying client requests from the server the ATA information from sharing clients who are travelling on the followed route. The followed route's origin and destination, route directions and start time at the origin are sent to the server. After arrival at the destination, the system displays to the querying client the requested counterfactual route's total ATA.

Sharing Client

The sharing client acts as a participatory sensor, and once a sharing client selects a route to travel on, the corresponding route directions are sent to the server. The route directions are used so the server can distinguish which sharing users are traveling which road segments. As the sharing client's location is updated at regular intervals, for each road segment that is travelled, the corresponding start and end locations along with the start and end times are sent to the server.

4.6.2 Maps API

The Maps API consists of three main functions. The first function is to provide map images and add overlays, such as markers or polylines to the map. The second is to provide geocoding to find the corresponding geographic coordinates from an inputted textual geographic location. Finally, the last function is to calculate and recommend a list of alternative routes for a given origin and destination, along with additional information such as ETA or distance. Driving directions for each road segment of the route are also provided.

4.6.3 Server

Three processes are involved in the application server component. One is to handle follow requests by searching and logging the sharing clients' locations and travel times. Second, the follow module's role is to elect users from among shared users to follow for each road segment along the selected route. Finally, the ATA calculation module's role is to record the ATA when a sharing user traverses waypoints, to search for road segment ATAs, and to calculate the corresponding end-to-end selected route's ATA. Details of each process are explained below.

Request Handler

The request handler's role is to store the road segments of the selected routes of sharing users that were calculated from the Maps API and the ATA in the form of start and end times for each traveled road segment. Then, the follow module is called in order to elect users to follow.

Follow Module

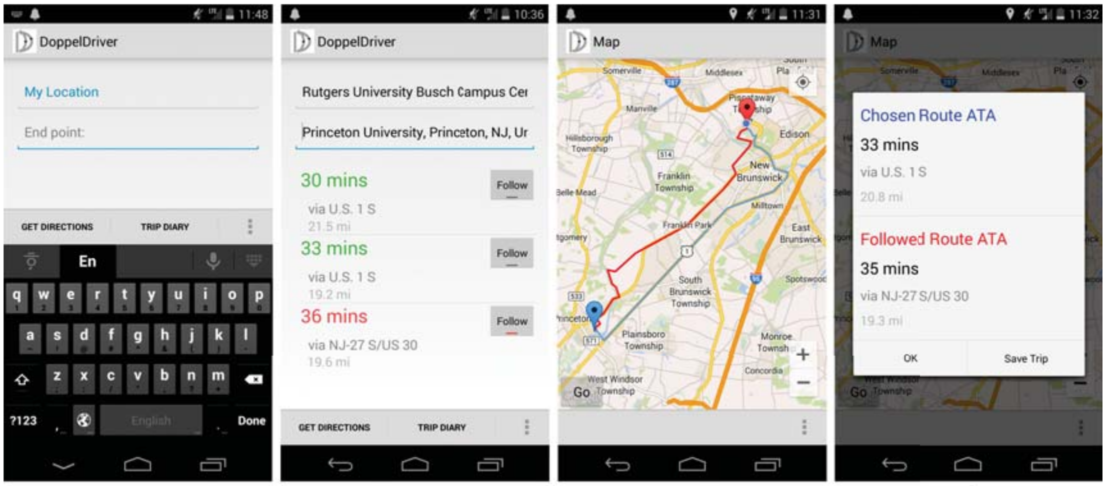
This process basically selects the best set of sharing users' travel times for the followed route. Per request of the ATA of the followed route, the follow module runs its algorithm to sort candidate users based on those who fit in the window time frame. Based on the candidate users, the algorithm will select the sharing user's ATA that best fits the querying user's time frame. The selection of sharing users is repeated for each road segment until all the road segments within the route are covered.

ATA Calculation

Upon arrival at the destination, the arrival times for each road segment along the followed route are calculated based on the profile from the follow module and from the storage server component. These values are added up to provide the final ATA for the entire journey.

4.7 Implementation

The DoppelDriver prototype system was implemented where the client runs on an Android 4.1+ mobile platform. The Google Maps API, the Places API, and the Directions API were utilized for the Maps API. The Google Maps API is used for the base map, the Places API is used for specifying locations as latitude/longitude coordinates, and the Directions API is used to calculate and recommend a list of alternative routes, along with step information. The server



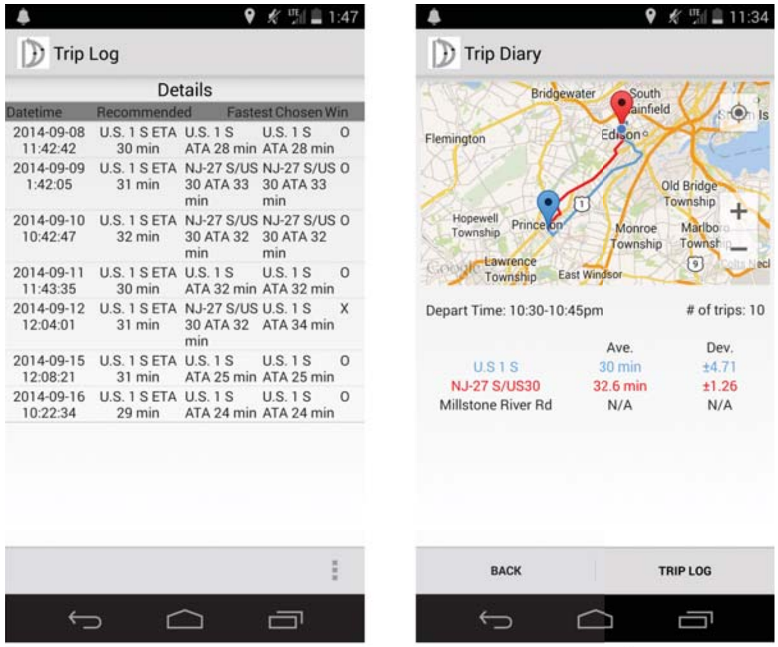
(a) Home screen (b) Route list (c) Map of routes (d) Result screen

Figure 4.11: DoppelDriver client prototype.

side is a Dell Optiplex 780 machine equipped with a 3.16 GHz Intel Core2 Duo CPU, 4GB DDR3 1066 MHz RAM, and a WD 320GB 7200 RPM SATA II disk. Moreover, we have set up server-side scripts developed by PHP5 running on Apache 2 on Ubuntu 12.04.4 LTS, and the data on ATA is stored in a MySQL database. The field experiments were carried out on the aforementioned platform.

Figure 4.11 illustrates the implementation of the client prototype. When an origin and destination are introduced, a recommended list of alternatives is shown and sorted by shortest ETA. The user will then select which route to take and which counterfactual route to follow, as shown in Figure 4.11 (b). Then, the overall view of the selected route and counterfactual route is shown, as in Figure 4.11 (c). Upon arrival, when the selected route arrival time is more than the counterfactual arrival time, then both routes' ATAs are shown, as seen in Figure 4.11 (d). On the other hand, if the selected route arrival time is shorter, the counterfactual route will show "Has not arrived" and a popup alert message will be shown upon arrival, notifying the ATA.

The *Trip Log* is shown in Figure 4.12 (a). Every trip can be saved so that users can utilize



(a) Trip log (b) Trip Diary

Figure 4.12: DoppelDriver Trip Log and Trip Diary.

their previous experiences. The first column states the date and time of the trip. The second column shows the navigation system's recommended best route (the recommended route that has the minimum ETA). The third column shows the route that turned out to have the minimum ATA (when counterfactual comparisons are available). The fourth column shows the route the user chose to drive. Finally, the last column represents whether the selected route was the true fastest route. The design of the trip log is modeled after the travel behavioral experiment with real-time traffic and post information in [28]. The *Trip Diary* is shown in Figure 4.12 (b). Depending on one's departure time, the Trip Diary uses the Trip Log data and past number of trips to calculate the averages and standard deviations of the actual travel times for that 15 min. time slot the departure time falls in (personalized route choice).

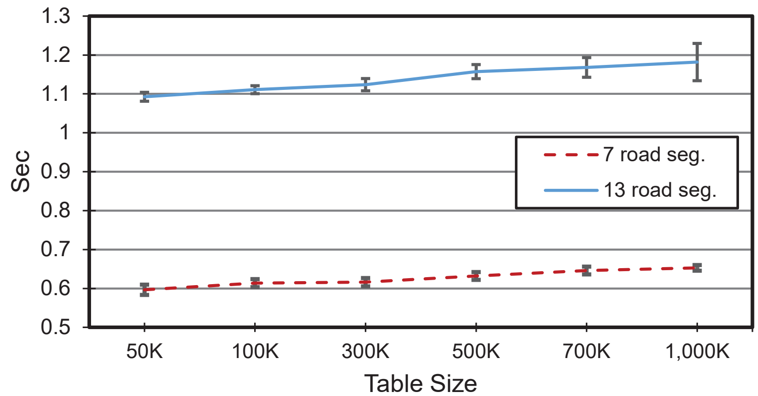


Figure 4.13: Counterfactual route ATA search time.

4.8 Field Experiment

A field study was carried out to validate the prototype by testing whether the ATAs on the counterfactual road segments were correctly shared and retrieved from the database, and to evaluate the search cost on different table sizes. Two drivers, one sharing and one querying client, simultaneously drove on two routes for the same origin and destination. The two routes were recommended from Google Maps where one route was via the highway with an average ETA of 15 mins (9.3 mi) consisting of 7 road segments while the other route was via local roads with an average ETA of 16 mins (7.3 mi) consisting of 13 road segments. A total of ten trips were carried out. Five trips were conducted, where at the origin, the querying client took the local route and requested the ATA on the highway route, i.e. the counterfactual route, where the sharing client was traveling on. The remaining five trips were conducted in reverse where the querying client took the highway route and the sharing client on the local road route. The database of our table that records the ATA was injected with random start and end travel times on random geo-points for start and end road segments. This process mimics how other drivers share their ATA. At the end of the trip, the retrieved ATA on the counterfactual route was compared to the sharing client's travel time to validate that the sharing client's travel time was correctly sent and retrieved from the database table. Also, the search time for the ATA on the counterfactual route was recorded on table sizes of 50K, 100K, 300K, 500K, 700K and 1,000K. Note that the taxi data set mentioned in the previous section was approximately 7% of the total traffic and covered the Beijing area where the table sizes were 309K for non-rush hour and 407K for rush hour.

For all the ten trips, the retrieved counterfactual route's ATAs coincided with the travel times of the sharing client. This shows that the sharing client's travel times for all the road segments within the counterfactual route were correctly updated for each road segment that was traveled and the ATA values were properly calculated to provide the final ATA for the entire journey. Also, Figure 4.13 shows the average search time and the standard deviation in seconds per table size when the local road (7 road seg.) and the highway road (13 road seg.) were the counterfactual route. We can observe that the table size does not affect the search time whereas the number of road segments searched mainly affected the search time.

4.9 Discussion

A problem may arise if the system cannot find a driver on a certain road segment of the non-chosen route within a certain time interval in order to compute the actual travel time. Our current solution is to query the start point of such road segments and use the ETA (provided by online map providers) instead to contribute to the calculation of the entire counterfactual route travel time. Such information can be used where there is no actual arrival-time data so as to provide the entire counterfactual route travel time. The same approach can be taken during the initial deployment of DoppelDriver where there may not be enough users to collect actual travel times. The system could be bootstrapped by taking advantage of most recent ETAs for immediate next road segments.

Furthermore, perhaps this system's counterfactual thinking approach to navigation routing could be extended to public transportation. By additionally considering public transportation options and their respective ATAs, alongside alternative routing, the user can be better informed of all that is available and what might be optimal. If public transportation is found to be timelier, demand for and interest in public transportation could increase. This idea is also discussed by Chorus [31], where such a strategy might be effective in light of the well-known misperceptions among travelers concerning the performance of, for example, public transport options with which they have less experience.

Also, when choosing participating users along alternative routes, more factors can be considered. DoppelDriver could additionally be personalized to the characteristics of the user and his or her driving habits and preferences. One aspect of this personality coefficient could include a user-optional survey based on the Likert scale (1-5), in which the user could indicate driving preferences (such as tendency to drive over the speed limit) or (routing preferences such as avoiding tollbooths or local roads). A second facet of the coefficient could include more objective details, such as vehicle attributes (weight or size, car make, model or miles per gallon) where similar attributes of cars are chosen for comparison. This similarity coefficient could be used to find better fits for participating users along the alternative routes when ATAs are provided, in order to maximize ATA accuracy for each user. Such coefficients could be considered when selecting a user for comparison.

4.10 Summary

This chapter presents a rationale and a system for counterfactual thinking in route selection. We presented a method by which ATAs for alternative routes can be calculated by combining other travelers' ATAs for segments that constitute the complete alternative route. A study based on taxi data shows that by selecting a suitable time window, it is possible to aggregate travel times for a route by using segment travel times.

This work presents the first crowd-participatory sharing service that collects data on actual arrival times. *DoppelDriver* empowers users with the ability to exercise strategic decision making and self-assessments of their route choices. Using data collected from participating users, DoppelDriver offers real-time comparison of alternative routes via real-time instantaneous position comparisons, as well as provision of ATAs in order to inform users in a novel way on how to better make routing decisions. We presented a design for, and prototype of, DoppelDriver where other user's locations and arrival times were used to compare against the time of the selected route.

Chapter 5

Conclusion

Current recommendations systems provide the fastest alternative routes to the destination based on various sources of traffic data. The fastest route which solely depends on the Estimated Time of Arrival (ETA) to a given destination is used as a decision-making tool where users can select from alternative options of routes.

In this dissertation, we have explored the limitations in route recommendation systems as we compared the recommended route options from popular online maps. As we have looked into the recommend routes given by four online map providers, we have identified that travel time variability exists among the maps. Such variability in travel times introduces uncertainty in route choice.

In many situations, users may not have the choice of the route and may prefer to access more information to complement route-deciding factors such as ETA. Such information can be the road condition, or the actual reason and status causing traffic jams. In many cases, such information cannot be determined automatically, while such information can be provided by other drivers driving ahead. We make use of such information by presenting a system called Social Vehicle Navigation, which utilizes visual traffic information provided by other users. Just seeing what is happening on the road ahead in a timely fashion can often alleviate stress and significantly improve the driving experience.

At the end of a trip, it is easy for drivers to receive feedback on their chosen route. However, receiving feedback on the travel times on non-chosen routes has not been available. We have introduced a system called DoppelDriver, which actively utilizes the actual travel time from participant users to calculate and provide the actual travel times on the non-chosen routes. Such experiential feedback on the chosen and non-chosen route feedback effects the experience and learning which play an important role in the decision making process.

The main conclusion of this dissertation is that current recommendation systems have limitations and can potentially introduce uncertainty in route choice. To support the decision-making in route choice, we address the use of visual traffic information for pre-trip route choice and the use of counterfactual travel times as post-choice feedback information on the forgone alternatives.

5.1 Future Work

Comparative Analysis on Online Maps

We have collected traffic data from four different map providers, Google Maps, Map Quest, HERE, and Bing Maps, to conduct a comparative analysis on their recommended routes. We plan our future work to expand our comparative study on online maps. Further data collection will be done on more origin destination pairs all varying by distance. One example might have the origin set to the median of each county and the destination set to a popular landmark.

Also, future work is to study the behavior of real users' route choice. In other words, we plan to compare personal route choices to the recommended routes from online maps to study the driver's compliance and route choice behavior. Comparison of actual travel times and estimated travel times will be compared to understand more on the variability and the accuracy of travel time.

Social Vehicle Navigation

Several issues require further research. Building an optimized user interface to enable drivers to interact with the navigation system while driving is essential. Issues such as passenger safety and reducing cognitive load must be further examined through an analytical user study.

Selecting the most relevant tweets to be included in the Tweet Digest and to effectively capture the semantic meaning of events on the route is a non-trivial task. Other approaches can be explored to improve tweet selection by using other criteria, such as user reputation, or by crowdsourcing this task to people willing to help in real time. Driver feedback on tweets can also help to eliminate improper or malicious tweets.

DoppelDriver

The choice of the threshold for the window size is of great importance as it reflects the computation of the actual travel time. As this research worked on the hypothesis of window size of 300 and 600 sec, future work is left to study with varying window sizes on different GPS data sets in order to maximize ATA accuracy for each user.

The main focus of this research in DoppelDriver is to present the potential usage of post-trip feedback and to analyze the feasibility of the proposed algorithm on obtaining ATAs on the non-chosen routes. Although there has been literature on transportation behavior that discuss in-lab experiments on route choice behavior when provided with post feedback, further work remains to study its impact of the proposed systems in real-world scenarios.

Appendix A

Appendix for Chapter 2

This appendix contains supplement Figures and Tables from Chapter 2.

Table A.1: Recommended routes throughout the data collection period for LA1.

Map (LA1)	Route ID	Distance	Count	Percentage
Google (Count: 60,056)	G1	15.1 mi	132	0.22 %
	G2	17.2 mi	17,139	28.54 %
	G3	18.3 mi	20,215	33.66 %
	G4	19.9 mi	20,336	33.86 %
	G5	20.9 mi	1,175	1.96 %
	G6	23.2 mi	939	1.56 %
	G7	23.2 mi	120	0.20 %
HERE (Count: 61,028)	H1	16.2 mi	13,355	21.88 %
	H2	18.3 mi	20,342	33.33 %
	H3	19.9 mi	20,343	33.33 %
	H4	20.9 mi	6,988	11.45 %
MapQuest (Count: 61,066)	M1	15.6 mi	20,357	33.34 %
	M2	18.3 mi	20,359	33.34 %
	M3	19.9 mi	20,350	33.32 %
Bing Maps (Count: 20,350)	B65	18.3 mi	10,237	50.30 %
	B116	19.9 mi	5,132	25.22 %
	B121	20.0 mi	432	2.12 %
	B137	20.5 mi	1,288	6.33 %
	B173	21.1 mi	449	2.21 %
	Bothers		2,812	13.82 %

Table A.2: Recommended routes throughout the data collection period for LA2.

Map (LA2)	Route ID	Distance	Count	Percentage
Google (Count: 61,013)	G1	16.6 mi	20,179	33.07 %
	G2	16.7 mi	20,337	33.33 %
	G3	16.9 mi	324	0.53 %
	G4	18.2 mi	20,014	32.80 %
	G5	18.2 mi	159	0.26 %
HERE (Count: 61,037)	H1	16.6 mi	20,346	33.33 %
	H2	16.7 mi	20,345	33.33 %
	H3	18.2 mi	20,346	33.33 %
MapQuest (Count: 61,109)	M1	16.6 mi	20,371	33.34 %
	M2	16.7 mi	20,374	33.34 %
	M3	18.2 mi	20,364	33.32 %
Bing Maps (Count: 20,372)	B10	16.6 mi	492	2.42 %
	B11	16.6 mi	2,868	14.08 %
	B16	16.7 mi	15,951	78.30 %
	Bother		1,061	5.21 %

Table A.3: Recommended routes throughout the data collection period for NY1.

Map (NY1)	Route ID	Distance	Count	Percentage
Google (Count: 52,850)	G1	10.4 mi	9,723	18.40 %
	G2	11.2 mi	536	1.01 %
	G3	11.4 mi	48	0.09 %
	G4	11.8 mi	2,140	4.05 %
	G5	12.4 mi	13,164	24.91 %
	G6	12.6 mi	24	0.05 %
	G7	13.0 mi	576	1.09 %
	G8	13.3 mi	1,848	3.50 %
	G9	16.0 mi	2,621	4.96 %
	G10	16.1 mi	1,966	3.72 %
	G11	16.1 mi	15,866	30.02 %
	G12	16.8 mi	1,848	3.50 %
	G13	16.9 mi	18	0.03 %
	G14	18.0 mi	2,448	4.63 %
	G15	18.1 mi	24	0.05 %
HERE (Count: 61,002)	H1	10.3 mi	6,979	11.44 %
	H2	10.4 mi	13,355	21.89 %
	H3	12.4 mi	20,334	33.33 %
	H4	16.0 mi	6,979	11.44 %
	H5	16.1 mi	13,355	21.89 %
MapQuest (Count: 60,996)	M1	10.3 mi	5,227	8.57 %
	M2	10.4 mi	15,105	24.76 %
	M3	12.4 mi	20,327	33.33 %
	M4	15.9 mi	3	0.00 %
	M5	16.1 mi	15,107	24.77 %
	M6	16.06 mi	5,227	8.57 %
Bing Maps (Count: 20,332)	B141	12.7 mi	569	2.80 %
	B233	16.1 mi	693	3.41 %
	B234	16.1 mi	2,561	12.60 %
	B242	16.1 mi	13,675	67.26 %
	B250	16.3 mi	337	1.66 %
	Bother		2,497	12.28 %

Table A.4: Recommended routes throughout the data collection period for NY2.

Map (NY2)	Route ID	Distance	Count	Percentage
Google (Count: 61,001)	G1	16.7 mi	1,871	3.07 %
	G2	16.8 mi	18,462	30.27 %
	G3	17.7 mi	2,534	4.15 %
	G4	18.1 mi	14,049	23.03 %
	G5	18.6 mi	2,809	4.60 %
	G6	20.8 mi	1,338	2.19 %
	G7	22.4 mi	17,358	28.46 %
	G8	22.5 mi	1,872	3.07 %
	G9	22.7 mi	708	1.16 %
HERE (Count: 61,014)	H1	16.8 mi	20,338	33.33 %
	H2	18.1 mi	20,338	33.33 %
	H3	19.4 mi	6,982	11.44 %
	H4	19.4 mi	13,356	21.89 %
MapQuest (Count: 60,998)	M1	16.8 mi	20,339	33.34 %
	M2	17.3 mi	20,332	33.33 %
	M3	18.1 mi	20,327	33.32 %
Bing Maps (Count: 20,344)	B53	16.8 mi	12,780	62.82 %
	B89	16.9 mi	3,508	17.24 %
	B149	18.1 mi	652	3.20 %
	Bother		3,404	16.73 %

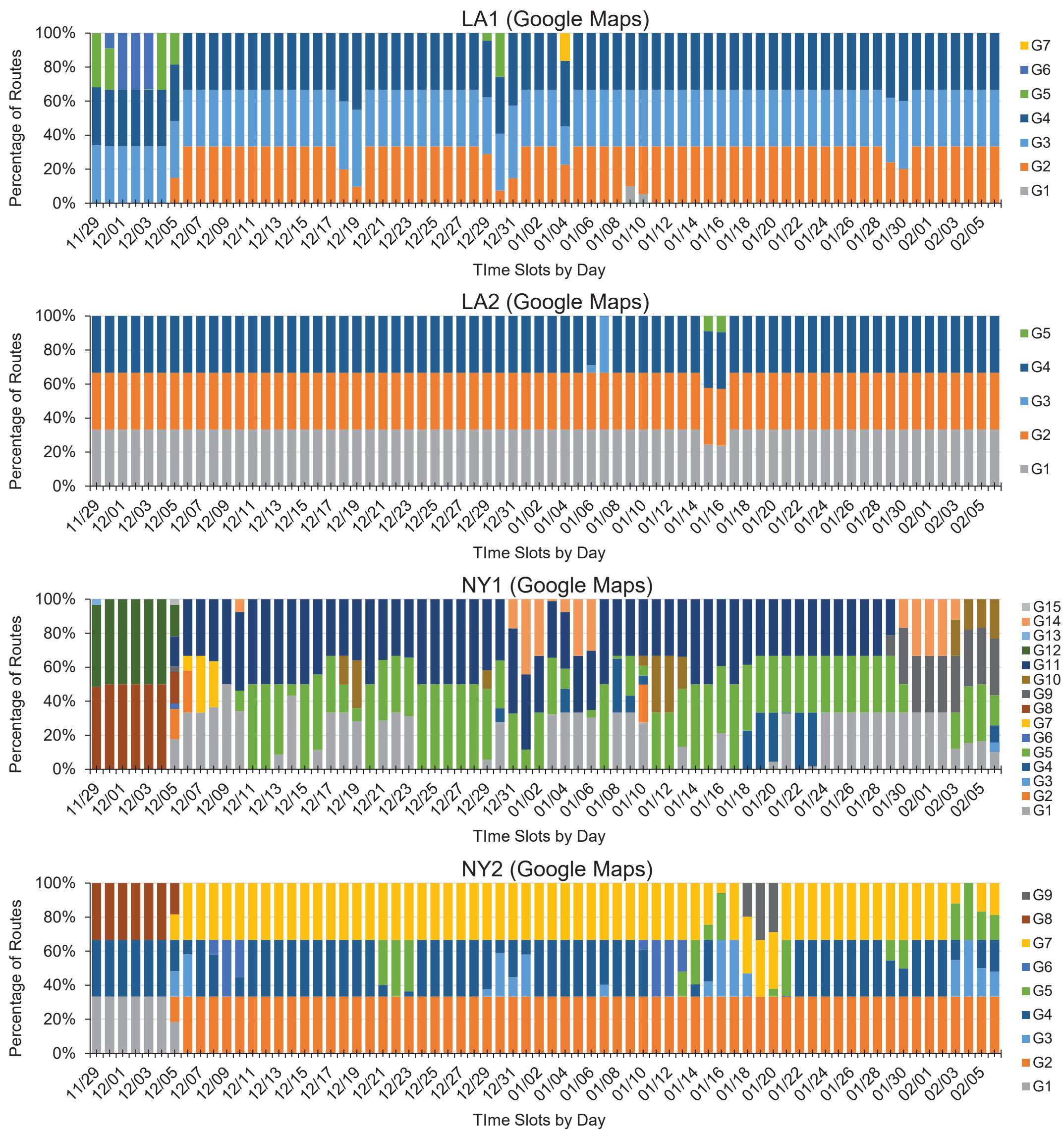


Figure A.1: Routes viewed by date for Google Maps.

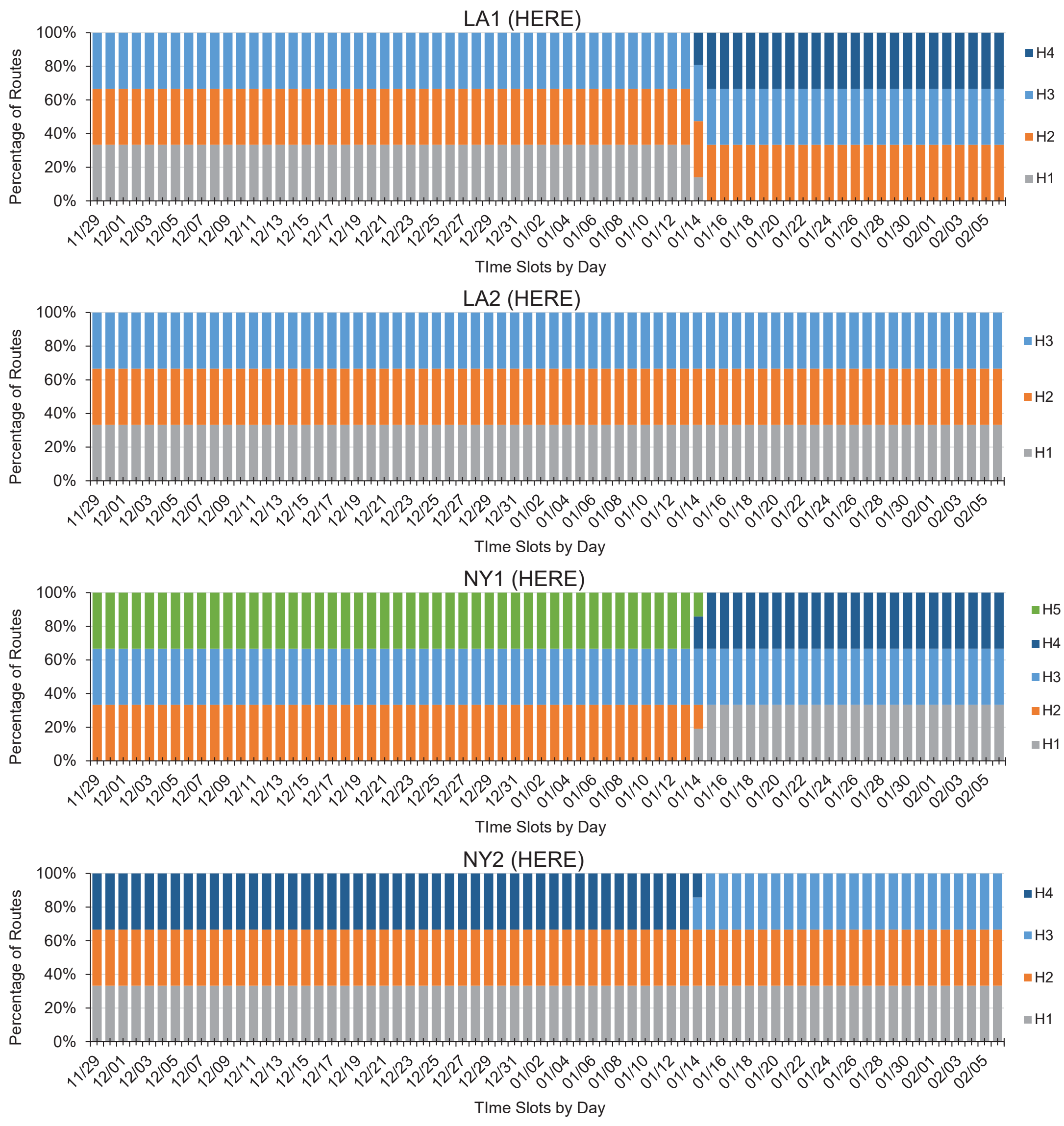


Figure A.2: Routes viewed by date for HERE.

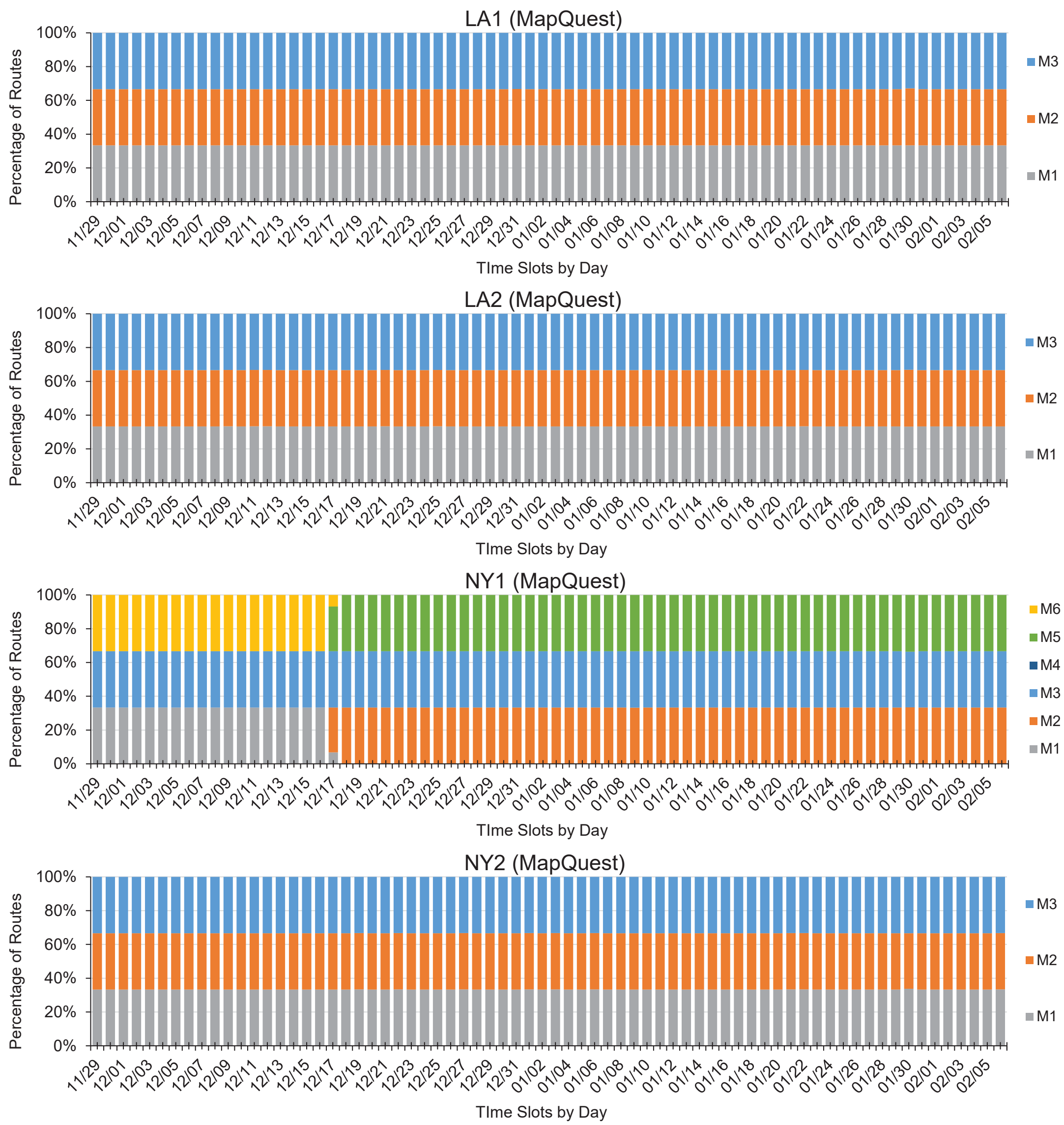


Figure A.3: Routes viewed by date for MapQuest.

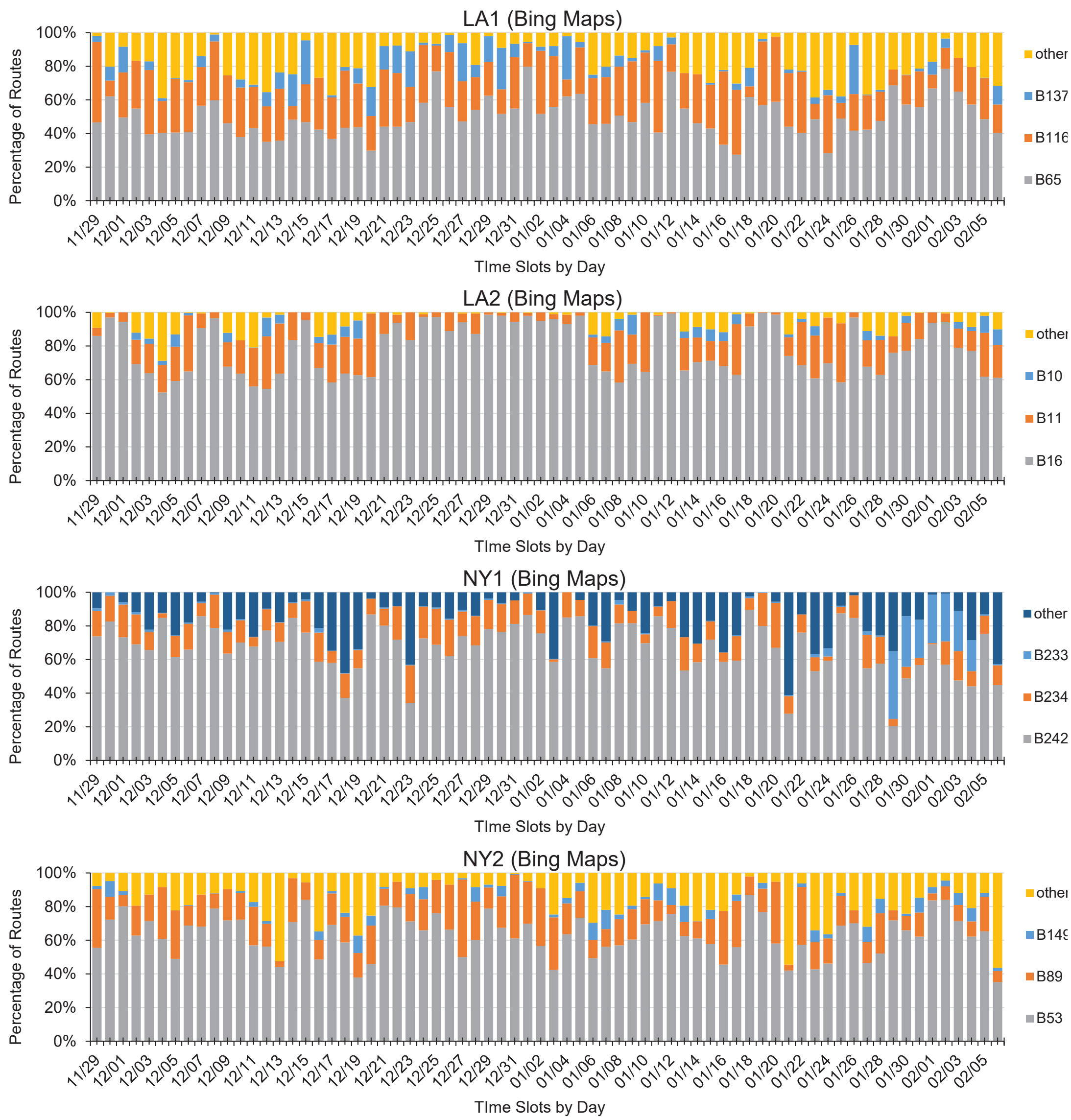


Figure A.4: Routes viewed by date for Bing Maps.

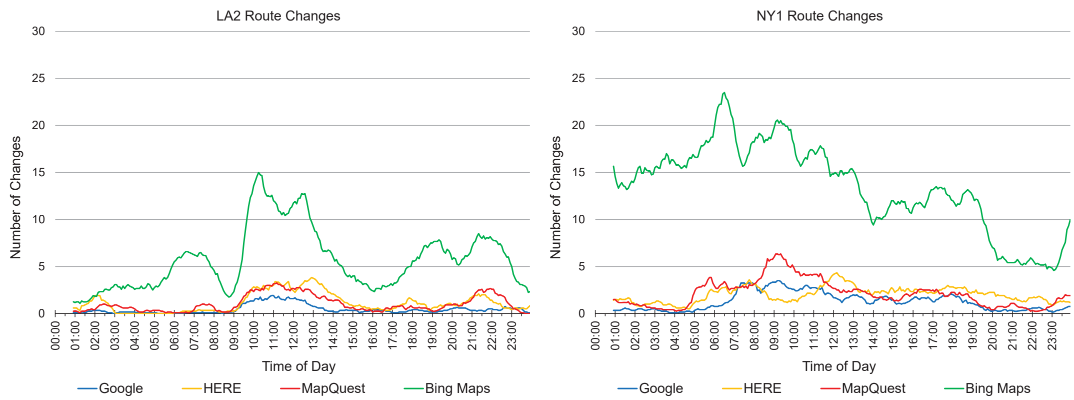


Figure A.5: Frequency of best-route changes for LA2 and NY1.

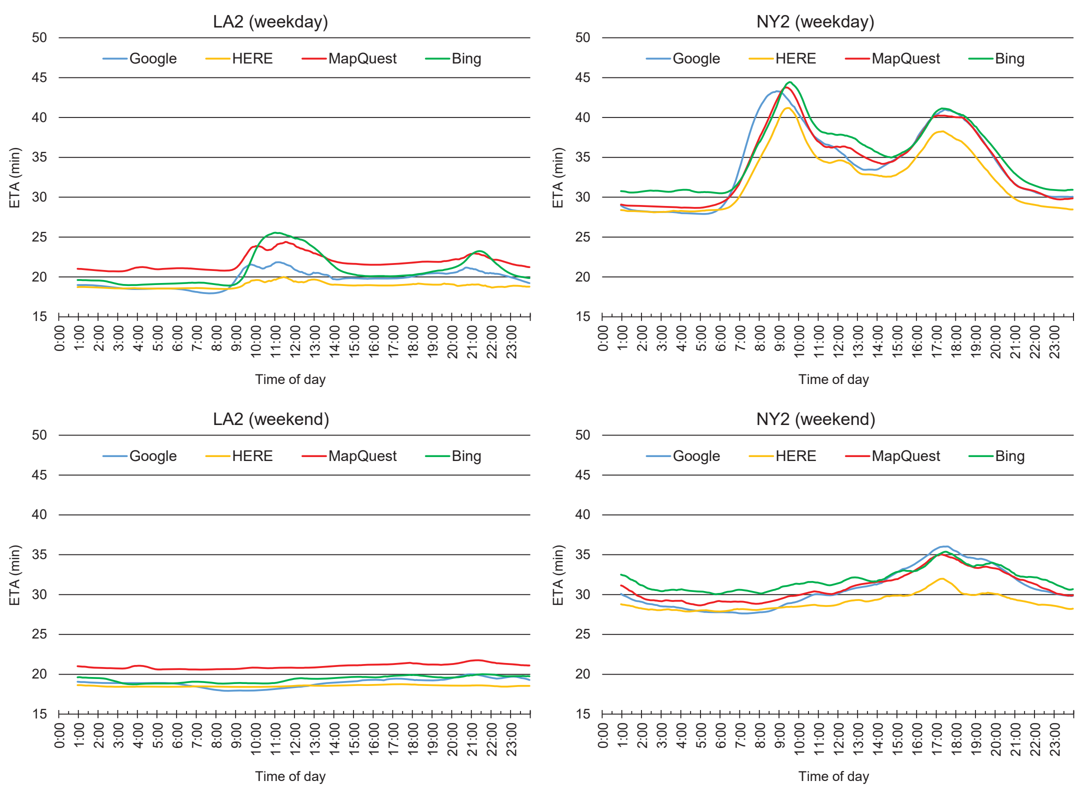


Figure A.6: Weekday and weekend average ETAs for LA2 and NY2.

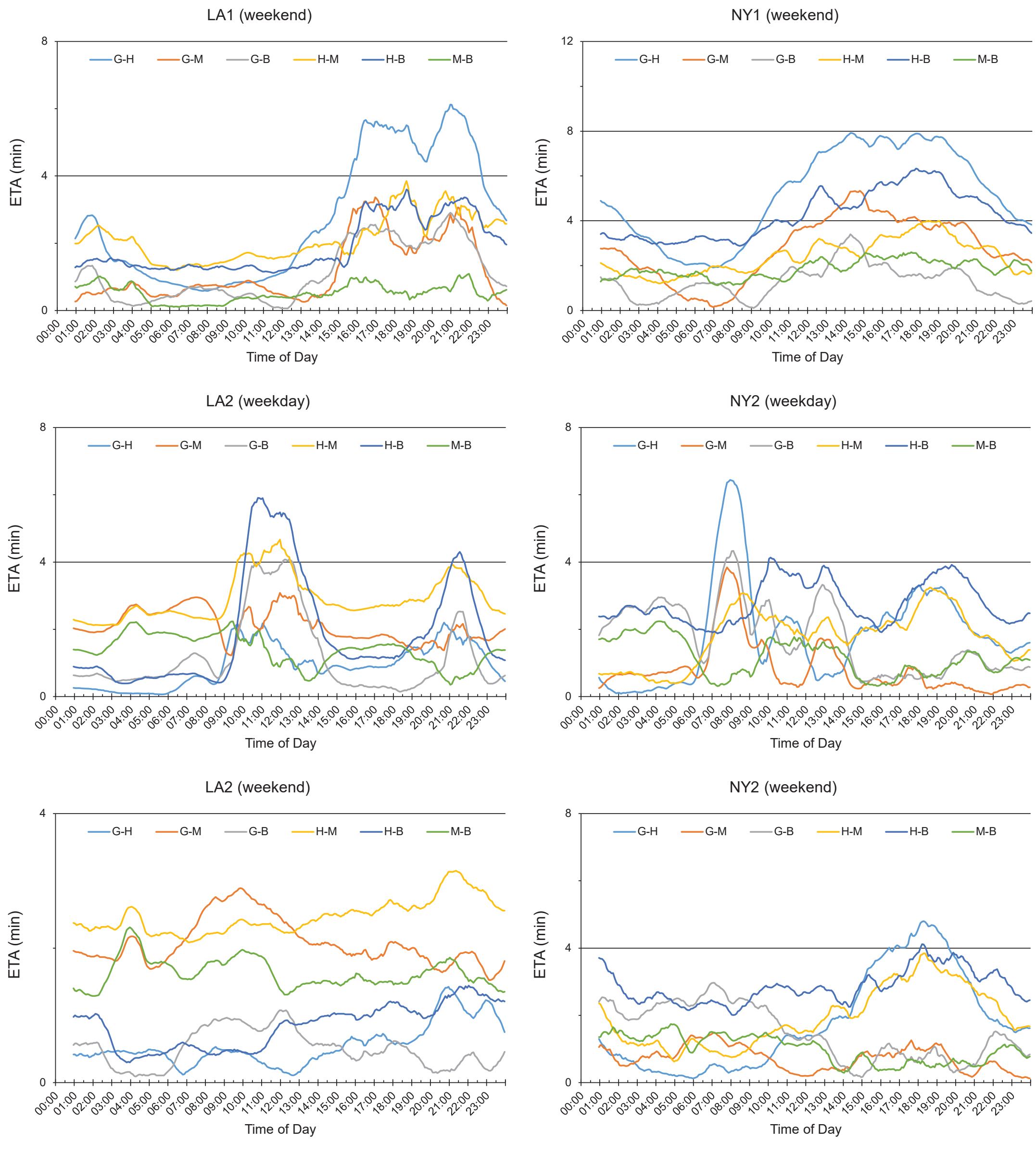


Figure A.7: Absolute difference for best ETAs (remaining figures).

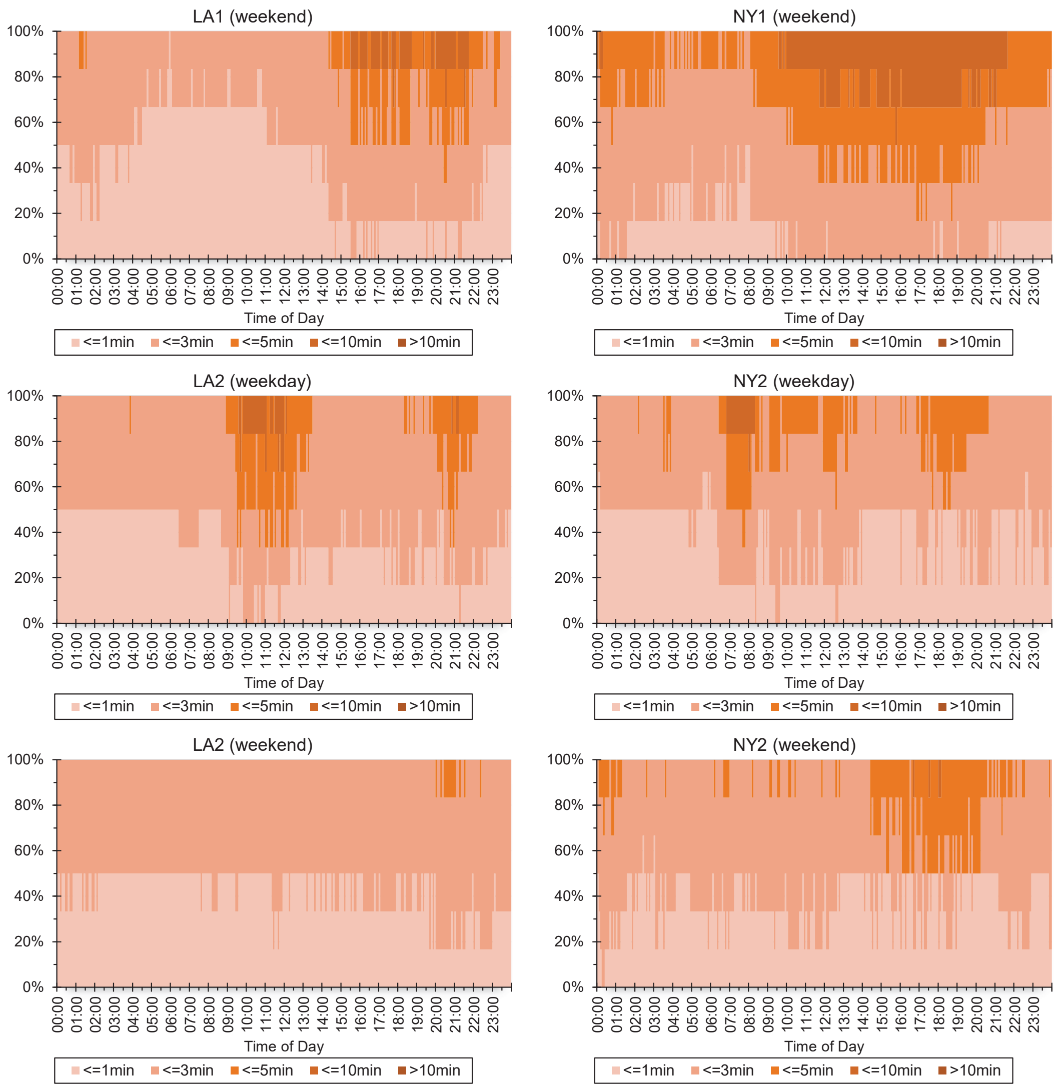


Figure A.8: Heat map on the difference on best ETAs for all online map pairs (remaining figures).

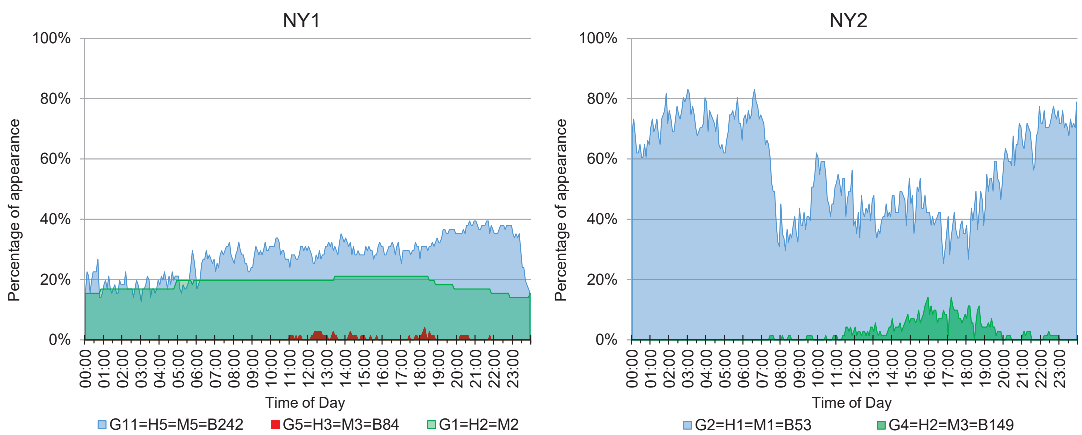


Figure A.9: Percentage of identical routes that appeared simultaneously for NY.

Appendix B

Appendix for Questionnaire Survey

This appendix contains the Survey Questionnaire for Chapter 3.

Profile

What is your gender?

- ○ Male
- ○ Female

Please select your age range.

- ○ 18-30
- ○ 31-40
- ○ 41-50
- ○ 51-60
- ○ 61-70
- ○ 71+

Driving experience.

- ○ less than 1 year
- ○ 1 to 3 years
- ○ 4 to 5 years
- ○ 5 to 10 years
- ○ more than 10 years

Part 1

In general, rate the source of traffic information you typically consider prior to heading to your destination.

	Strongly Don't Consider	Don't Consider	Neutral	Consider	Strongly Consider
Radio or TV Traffic News	○	○	○	○	○
Social Network Websites (e.g. Facebook, Twitter etc)	○	○	○	○	○
State Government Websites (e.g. Department of Transportation or 511 websites)	○	○	○	○	○
Online Maps (e.g. Google Maps, Mapquest, HERE Maps etc.)	○	○	○	○	○
Smartphone apps/mobile web (ex. Waze, Inrix)	○	○	○	○	○
GPS Navigation Systems e.g. TomTom, Garmin etc.)	○	○	○	○	○
Traffic Camera Images (e.g. CCTV)	○	○	○	○	○
Other (Please Specify):	○	○	○	○	○

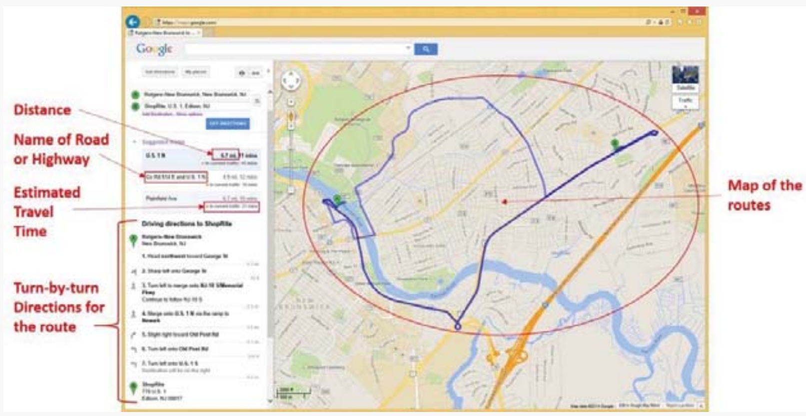
In general, what do you mostly rely on when choosing a route to a destination in terms of percentage? (Total must sum to 100)

Based on Experience	0
Recommendations from a navigation	0
Others: (Please Specify)	0
Total	0

What type of navigation system or app do you use? (you may select multiple answers)

- ☐ In-car navigation systems (e.g. TomTom, Garmin)
- ☐ Smartphone Navigation Apps (e.g. Google Maps, Map Quest, Waze, Apple Maps, HERE Maps)
- ☐ None
- ☐ Others: (Please Specify)

Rate the criteria you consider when making your decision to plan for a route using a navigation system/app or online maps.



	Strongly Don't Consider	Don't Consider	Neutral	Consider	Strongly Consider
Distance	○	○	○	○	○
Estimated travel time	○	○	○	○	○
Name of road or highway	○	○	○	○	○
Map of the routes	○	○	○	○	○
Turn-by-turn directions for the route	○	○	○	○	○

How likely do you consider to use a navigation system/app for the following instances?

	Strongly Don't Consider	Don't Consider	Neutral	Consider	Strongly Consider
Daily commute to and from work	○	○	○	○	○
Time constraint occasions	○	○	○	○	○
Daily activities	○	○	○	○	○
Long distance travel	○	○	○	○	○
Special occasions (e.g. Traveling for leisure)	○	○	○	○	○
When delay occurs and want to find an alternative route	○	○	○	○	○
Familiar routes	○	○	○	○	○
Unfamiliar routes	○	○	○	○	○

In general, how accurate do you think the Estimated Time of Arrival (ETA) is calculated by your navigation/app?

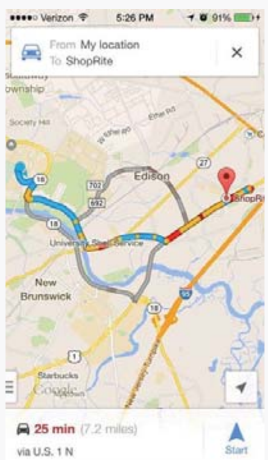
- ○ Not at all accurate
- ○ Not accurate
- ○ Somewhat not accurate
- ○ Moderate
- ○ Somewhat accurate
- ○ Accurate
- ○ Very accurate

What other secondary level of detailed traffic information may be useful in your decision to select a route? (select all that apply)

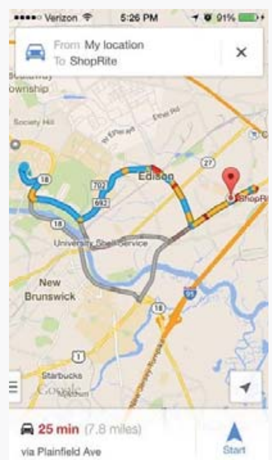
- ☐ Short video clips showing the current traffic on interested locations (e.g. CCTV)
- ☐ Images showing the current traffic on interested locations
- ☐ Voice messages describing the current traffic on interested locations
- ☐ Text descriptions that are tagged on the map
- ☐ Others: (Please Specify)

Part 2-1

Imagine you are ***unfamiliar*** with the area you are traveling to and you need to choose from 2 recommended routes. Which route would you consider to take? (Note: From below, time in "min" represent the estimated travel time based on current traffic flow)



○



○

Rate the criteria that you used to make your decision.

	Strongly Don't Consider	Don't Consider	Neutral	Consider	Strongly Consider
Road or highway name	○	○	○	○	○
Distance	○	○	○	○	○
Estimate travel time (ETA)	○	○	○	○	○
Colors coded on roads indicating of the	○	○	○	○	○

severity of the traffic

Part 2-2

With the same recommended 2 routes in the previous question, now imagine you were given additional information (real-time image traffic reports) as shown in the figures below. Which route would you consider to take: **Route 1** or **Route 2**? (Note that the images were taken from real trips)

○ Route 1

○ Route 2

Rate the criteria that you used to make your decision.

	Strongly Don't Consider	Don't Consider	Neutral	Consider	Strongly Consider
Image traffic reports	○	○	○	○	○
Road or highway name	○	○	○	○	○
Distance	○	○	○	○	○
Estimate travel time (ETA)	○	○	○	○	○
Colors coded on roads indicating of the severity of the traffic	○	○	○	○	○

Part 3

If you were provided with an application that provided real-time image traffic reports on relevant locations, how important would the images be in the decision of route choice?

- ○ Essential
- ○ Important
- ○ Neither
- ○ Not important
- ○ Not at all important

If you were provided with such app, why would you look at real-time traffic images before you set out on a trip? (select all that apply)

- ☐ To confirm the traffic condition when there is doubt in the reliability of other criteria such as ETA
- ☐ To see traffic conditions for myself
- ☐ To check particular intersections
- ☐ To get an extra level of detail
- ☐ To prepare for the conditions ahead
- ☐ To estimate travel time
- ☐ Would not use traffic images before setting out
- ☐ For curiosity / entertainment
- ☐ Others: (Please Specify)

What influence would real-time traffic images have on your trip? (select all that apply)

- ☐ Images help to determine route choice
- ☐ Prepare me for the conditions ahead
- ☐ Re-estimate travel time
- ☐ Set off earlier or later than planned
- ☐ Travel on public transportation
- ☐ I would not be influenced by the images
- ☐ Others: (Please Specify)

References

[1] 511 org websites. Available online at NJ:www.511nj.org/,NY:www.511ny.org/ ,VA:www.511virginia.org/,MN:www.511mn.org/SF:www.511.org/.

[2] Apple maps. Available online at https://www.apple.com/ios/maps/.

[3] Arcgis. Available online at https://www.arcgis.com/.

[4] Bing maps. Available online at https://www.bing.com/maps/.

[5] Facebook. Available online at https://www.facebook.com/.

[6] Google maps. Available online at https://developers.google.com/maps/.

[7] Here. Available online at https://here.com/.

[8] Inrix traffic. Available online at http://www.inrixtraffic.com/.

[9] Instragram. Available online at http://instagram.com/.

[10] Mapquest. Available online at https://www.mapquest.com/.

[11] Osmand (openstreetmap automated navigation directions). Available online at http://osmand.net/.

[12] Qualtrics. Available online at http://www.qualtrics.com/.

[13] R project for statistical computing. Available online at https://www.r-project.org/.

[14] Self-driving car test. Available online at https://waymo.com/.

[15] Trapster. Available online at http://www.trapster.com/.

[16] Twitter. Available online at https://twitter.com/.

[17] Waze. Available online at http://www.waze.com/.

[18] T. Adam, I. Atkinson, and M. Dixon. Navigation device displaying traffic information, May 24 2007. US Patent App. 11/561,484.

[19] E. D' Andrea, P. Ducange, B. Lazzerini, and F. Marcelloni. Real-time detection of traffic from twitter stream analysis. *IEEE Transactions on Intelligent Transportation Systems*, 16(4):2269–2283, 2015.

[20] Erel Avineri and Joseph N. Prashker. The impact of travel time information on travelers' learning under uncertainty. *Transportation*, 33(4):393–408, 2006.

[21] Eran Ben-Elia, Roberta Di Pace, Gennaro N. Bifulco, and Yoram Shiftan. The impact of travel information's accuracy on route-choice. *Transportation Research Part C: Emerging Technologies*, 26:146–159, 2013.

[22] Eran Ben-Elia, Ido Erev, and Yoram Shiftan. The combined effect of information and experience on drivers' route-choice behavior. *Transportation*, 35(2):165–177, 2008.

[23] Eran Ben-Elia, Robert Ishaq, and Yoram Shiftan. "if only i had taken the other road...": Regret, risk and reinforced learning in informed route-choice. *Transportation*, 40(2):269–293, 2013.

[24] Eran Ben-Elia and Yoram Shiftan. Which road do i take? a learning-based model of route-choice behavior with real-time information. *Transportation Research Part A: Policy and Practice*, 44(4):249–264, 2010.

[25] RF Benekohal and Joseph Treiterer. Carsim: Car-following model for simulation of traffic in normal and stop-and-go conditions. *Transportation research record*, (1194), 1988.

[26] Margrit Betke, Esin Haritaoglu, and Larry S. Davis. Real-time multiple vehicle detection and tracking from a moving vehicle. *Machine Vision and Applications*, 12(2):69–83, 2000.

[27] Transportation Research Board. Highway capacity manual. *Washington, DC*, 2000.

[28] Enide Bogers, Francesco Viti, and Serge Hoogendoorn. Joint modeling of advanced travel information service, habit, and learning impacts on route choice by laboratory simulator experiments. *Transportation Research Record: Journal of the Transportation Research Board*, 1926:189–197, 2005.

[29] V. Ceikute and C. S. Jensen. Routing service quality - local driver behavior versus routing services. In *2013 IEEE 14th International Conference on Mobile Data Management (MDM)*, volume 1, pages 97–106.

[30] Peter Chen, Karthik Srinivasan, and Hani Mahmassani. Effect of information quality on compliance behavior of commuters under real-time traffic information. *Transportation Research Record: Journal of the Transportation Research Board*, 1676:53–60, 1999.

[31] Caspar G. Chorus. Acquisition of ex-post travel information: A matter of balancing regrets. *Transportation Science*, 48(2):243–255, 2014.

[32] Catherine Clifford. 10 u.s. cities with the longest commute times, 2015. Available online at `http://www.entrepreneur.com/article/246563`.

[33] E. Coelingh and S. Solyom. All aboard the robotic road train. *IEEE Spectrum*, 49(11):34–39, 2012.

[34] Giselle de Moraes Ramos, Winnie Daamen, and Serge Hoogendoorn. Modelling travellers' heterogeneous route choice behaviour as prospect maximizers. *Journal of Choice Modelling*, 6:17–33, 2013.

[35] Kit Eaton. Map apps to bypass the rivalry of google and apple, 2012. Available online at `http://www.nytimes.com/2012/10/04/technology/personaltech/navigon-and-waze-map-apps-to-bypass-apple-and-google-review.html?_r=0`.

[36] Paul A. Eisenstein. In-vehicle navigation sales will quadruple by 2019, 2012. Available online at `http://bottomline.nbcnews.com/_news/2012/07/11/12683063-in-vehicle-navigation-sales-will-quadruple-by-2019?lite`.

[37] D. Fagan and R. Meier. Intelligent time of arrival estimation. In *2011 IEEE Forum on Integrated and Sustainable Transportation Systems*, pages 60–66.

[38] New Cities Foundation. Connected commuting: Research and analysis on the new cities foundation task force in san jose, 2012. Available online at `http://www.newcitiesfoundation.org/wp-content/uploads/New-Cities-Foundation-Connected-Commuting-Full-Report.pdf`.

[39] Hong-Cheng Gan, Yang Bai, and June Wei. Why do people change routes? impact of information services. *Industrial Management and Data Systems*, 113(3):403–422, 2013.

[40] Raghu K. Ganti, Nam Pham, Hossein Ahmadi, Saurabh Nangia, and Tarek F. Abdelzaher. Greengps: A participatory sensing fuel-efficient maps application. In *Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services*, MobiSys '10, pages 151–164, New York, NY, USA, 2010. ACM.

[41] Shravan Gaonkar, Jack Li, Romit Roy Choudhury, Landon Cox, and Al Schmidt. Micro-blog: Sharing and querying content through mobile phones and social participation. In *Proceedings of the 6th International Conference on Mobile Systems, Applications, and Services*, MobiSys '08, pages 174–186, New York, NY, USA, 2008. ACM.

[42] D. M. Gavrila and V. Philomin. Real-time object detection for smart vehicles. In *Proceedings of the Seventh IEEE International Conference on Computer Vision (ICCV)*, volume 1, pages 87–93 vol.1.

[43] R. Gerber and N. H. Nagel. Knowledge representation for the generation of quantified natural language descriptions of vehicle traffic in image sequences. In *Proceedings of 3rd IEEE International Conference on Image Processing (ICIP)*, volume 1, pages 805–808 vol.2.

[44] M. Gerla, J. T. Weng, and G. Pau. Pics-on-wheels: Photo surveillance in the vehicular cloud. In *2013 International Conference on Computing, Networking and Communications (ICNC)*, pages 1123–1127.

[45] Mario Gerla and Leonard Kleinrock. Vehicular networks and the future of the mobile internet. *Computer Networks*, 55(2):457–469, 2011.

[46] Dan Goldstein. The counterfactual gps! Available online at `https://www.r-project.org/`.

[47] Reginald G Golledge. Defining the criteria used in path selection. *University of California Transportation Center*, 1995.

[48] Feng Guo, Hesham Rakha, and Sangjun Park. Multistate model for travel time reliability. *Transportation Research Record: Journal of the Transportation Research Board*, 2188:46–54, 2010.

[49] Ali Haghani, Xuechi Zhang, and Masoud Hamedi. Validation and augmentation of inrix arterial travel time data using independent sources. Report, 2015.

[50] B.L. Hanchett. Traffic condition information system, March 7 1995. US Patent 5,396,429.

[51] Dirk Helbing, Martin Schönhof, Hans-Ulrich Stark, and Janusz A. Hołyst. How individuals learn to take turns: Emergence of alternating cooperation in a congestion game and the prisoner's dilemma. *Advances in Complex Systems*, 08(01):87–116, 2005.

[52] Alexander E.M. Hess and Samuel Weigley. Ten cities with the worst traffic, 2013. Available online at `https://www.usatoday.com/story/money/cars/2013/05/04/worst-traffic-cities/2127661/`.

[53] Laura Higgins, Alicia Nelson, Tina Geiselbrecht, and Brooke Ullman. Understanding the decision-making process for drivers faced with lane restrictions or closures on wisconsin highways. Report, 2013.

[54] C. Humphrey and A. Jennings. Get the picture: How digital camera images are changing the way we travel. In *Road Transport Information and Control Conference and the ITS United Kingdom Members' Conference (RTIC 2010) - Better transport through technology, IET*, pages 1–6, May 2010.

[55] R. Hussain, F. Abbas, J. Son, D. Kim, S. Kim, and H. Oh. Vehicle witnesses as a service: Leveraging vehicles as witnesses on the road in vanet clouds. In *2013 IEEE 5th International Conference on Cloud Computing Technology and Science (CloudCom)*, volume 1, pages 439–444.

[56] Andreas Janecek, Karin A. Hummel, Danilo Valerio, Fabio Ricciato, and Helmut Hlavacs. Cellular data meet vehicular traffic theory: Location area updates and cell transitions for travel time estimation. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, UbiComp '12, pages 361–370, New York, NY, USA, 2012. ACM.

[57] S. Kamijo, Y. Matsushita, K. Ikeuchi, and M. Sakauchi. Traffic monitoring and accident detection at intersections. *IEEE Transactions on Intelligent Transportation Systems*, 1(2):108–118, 2000.

[58] Adrian G. Kudler. La doing better than average on carpooling, public transit, 2013. Available online at `http://la.curbed.com/archives/2013/03/la_doing_better_than_average_on_carpooling_public_transit.php`.

[59] D. Kwak, D. Kim, R. Liu, L. Iftode, and B. Nath. Tweeting traffic image reports on the road. In *6th International Conference on Mobile Computing, Applications and Services*, MobiCASE '14, pages 40–48.

[60] D. Kwak, D. Kim, R. Liu, B. Nath, and L. Iftode. Doppeldriver: Counterfactual actual travel times for alternative routes. In *2015 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, pages 178–185.

[61] D. Kwak, R. Liu, D. Kim, B. Nath, and L. Iftode. Seeing is believing: Sharing real-time visual traffic information via vehicular clouds. *IEEE Access*, 4:3617–3631, 2016.

[62] Gary Langer. Poll: Traffic in the united states, 2005. Available online at `http://abcnews.go.com/Technology/Traffic/story?id=485098&page=1&singlePage=true`.

[63] David M. Levinson. Understanding the irrational commuter, 2013. Available online at `http://www.symposium-magazine.com/understanding-the-irrational-commuter/`.

[64] Hainan Li, Randall Guensler, and Jennifer Ogle. Analysis of morning commute route choice patterns using global positioning system-based vehicle activity data. *Transportation Research Record: Journal of the Transportation Research Board*, 1926:162–170, 2005.

[65] Janne Lindqvist and Jason Hong. Undistracted driving: A mobile phone that doesn't distract. In *Proceedings of the 12th Workshop on Mobile Computing Systems and Applications*, HotMobile '11, pages 70–75, New York, NY, USA, 2011. ACM.

[66] R. Liu, H. Liu, D. Kwak, Y. Xiang, C. Borcea, B. Nath, and L. Iftode. Themis: A participatory navigation system for balanced traffic routing. In *2014 IEEE Vehicular Networking Conference (VNC)*, pages 159–166.

[67] Ruilin Liu, Hongzhang Liu, Daehan Kwak, Yong Xiang, Cristian Borcea, Badri Nath, and Liviu Iftode. Balanced traffic routing: Design, implementation, and evaluation. *Ad Hoc Networks*, 37, Part 1:14–28, 2016.

[68] Angélica Lozano, Giuseppe Manfredi, and Luciano Nieddu. An algorithm for the recognition of levels of congestion in road traffic problems. *Mathematics and Computers in Simulation*, 79(6):1926–1934, 2009.

[69] Xuan Lu, Song Gao, and Eran Ben-Elia. Information impacts on route choice and learning behavior in a congested network. *Transportation Research Record: Journal of the Transportation Research Board*, 2243:89–98, 2011.

[70] Peter McMeekin, Jo Gray, Gary A Ford, Jay Duckett, and Christopher I Price. A comparison of actual versus predicted emergency ambulance journey times using generic geographic information system software. *Emergency Medicine Journal*, 2013.

[71] Ender Morgul, Hong Yang, Abdullah Kurkcu, Kaan Ozbay, Bekir Bartin, Camille Kamga, and Richard Salloum. Virtual sensors. *Transportation Research Record: Journal of the Transportation Research Board*, 2442:106–116, 2014.

[72] Paul Newson and John Krumm. Hidden markov map matching through noise and sparseness. In *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, GIS '09, pages 336–343, New York, NY, USA, 2009. ACM.

[73] Robert B. Noland and John W. Polak. Travel time variability: A review of theoretical and empirical issues. *Transport Reviews*, 22(1):39–54, 2002.

[74] K. A. Olsen and E. Fagerlie. Adaptive systems - a case for calculating estimated time of arrival. *IEEE Potentials*, 30(2):15–19, 2011.

[75] D. J. Parkyns and M. Bozzo. Cctv camera sharing for improved traffic monitoring. In *Road Transport Information and Control - RTIC 2008 and ITS United Kingdom Members' Conference, IET*, pages 1–6, May 2008.

[76] Brad Plumer. Commuting in the u.s. is hellish — but at least it's not getting worse, 2013.

[77] J.D. Power and Associates. Navigation usage and satisfaction study, 2013. Available online at `http://www.jdpower.com/content/press-release/O4ycHMo/2012-u-s-navigation-usage-and-satisfaction-study.htm`.

[78] M. Rahmani, E. Jenelius, and H. N. Koutsopoulos. Route travel time estimation using low-frequency floating car data. In *16th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pages 2292–2297.

[79] M Ramos, Emma Frejinger, W Daamen, and Serge Hoogendoorn. A revealed preference study on route choices in a congested network with real-time information. In *The 13th International Conference on Travel Behavior Research, Toronto, Ontario, Canada, July 15-20, 2012*.

[80] TomTom Press Release. Study finds los angeles most congested city in north america, 2013. Available online at `http://corporate.tomtom.com/releasedetail.cfm?releaseid=732208`.

[81] Anna Izabel João Tostes Ribeiro, Thiago Henrique Silva, Fátima Duarte-Figueiredo, and Antonio A.F. Loureiro. Studying traffic conditions by analyzing foursquare and instagram data. In *Proceedings of the 11th ACM Symposium on Performance Evaluation of Wireless Ad Hoc, Sensor, & Ubiquitous Networks*, PE-WASUN '14, pages 17–24, New York, NY, USA, 2014. ACM.

[82] AJ Richardson and MAP Taylor. Travel time variability on commuter journeys. *High Speed Ground Transportation Journal*, 12(1), 1978.

[83] Neal J. Roese. The functional basis of counterfactual thinking. *Journal of Personality and Social Psychology*, 66(5):805–818, 1994.

[84] Neal J. Roese and James M. Olson. *Counterfactual Thinking*. John Wiley & Sons, Ltd, 2006.

[85] T. Sakaki, Y. Matsuo, T. Yanagihara, N. P. Chandrasiri, and K. Nawa. Real-time event extraction for driving information from social sensors. In *2012 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER)*, pages 221–226.

[86] M. Satyanarayanan, P. Bahl, R. Caceres, and N. Davies. The case for vm-based cloudlets in mobile computing. *IEEE Pervasive Computing*, 8(4):14–23, 2009.

[87] Axel Schulz, Petar Ristoski, and Heiko Paulheim. *I See a Car Crash: Real-Time Detection of Small Scale Incidents in Microblogs*, pages 22–33. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.

[88] R. Selten, T. Chmura, T. Pitz, S. Kube, and M. Schreckenberg. Commuters route choice behaviour. *Games and Economic Behavior*, 58(2):394–406, 2007.

[89] Wenjie Sha, Daehan Kwak, Badri Nath, and Liviu Iftode. Social vehicle navigation: Integrating shared driving experience into vehicle navigation. In *Proceedings of the 14th Workshop on Mobile Computing Systems and Applications*, HotMobile '13, pages 16:1–16:6, New York, NY, USA, 2013. ACM.

[90] Y. Shiftan, S. Bekhor, and G. Albert. Route choice behaviour with pre-trip travel time information. *IET Intelligent Transport Systems*, 5(3):183–189, September 2011.

[91] Nikhil Sikka. *Understanding travelers' route choice behavior under uncertainty*. Thesis, 2012.

[92] Stephen Smaldone, Lu Han, Pravin Shankar, and Liviu Iftode. Roadspeak: Enabling voice chat on roadways using vehicular social networks. In *Proceedings of the 1st Workshop on Social Network Systems*, SocialNets '08, pages 43–48, New York, NY, USA, 2008. ACM.

[93] Robert Speer and Catherine Havasi. Representing general relational knowledge in conceptnet 5. In *LREC*, pages 3679–3686.

[94] S. Speirs and P. Whitehead. Impact of images from traffic cameras on journey planning. In *Road Transport Information and Control - RTIC 2008 and ITS United Kingdom Members' Conference, IET*, pages 1–8, May 2008.

[95] Karthik K. Srinivasan and Hani S. Mahmassani. Analyzing heterogeneity and unobserved structural effects in route-switching behavior under atis: a dynamic kernel logit formulation. *Transportation Research Part B: Methodological*, 37(9):793–814, 2003.

[96] Hans-Ulrich Stark, Dirk Helbing, Martin Schönhof, JA Hołyst, A Innocenti, and P Sbriglia. Alternating cooperation strategies in a route choice game: Theory, experiments, and effects of a learning scenario. *Games, Rationality and Behaviour. Houndmills and New York: Palgrave McMillan*, pages 256–273, 2008.

[97] S Susilawati, M Taylor, and S Somenahalli. Travel time variability and reliability: the south road corridor study case. In *Conference of Australian Institutes of Transport Research (CAITR), 30th, 2008, Perth, Western Australia, Australia*.

[98] Susilawati Susilawati, Michael A. P. Taylor, and Sekhar V. C. Somenahalli. Travel time reliability measurement for selected corridors in the adelaide metropolitan area. *Journal of the Eastern Asia Society for Transportation Studies*, 8:86–102, 2010.

[99] A. M. Tawfik, H. A. Rakha, and S. D. Miller. Driver route choice behavior: Experiences, perceptions, and choices. In *2010 IEEE Intelligent Vehicles Symposium (IV)*, pages 1195–1200.

[100] Robert Thomson. I-95 drivers get to compare travel times, 2013. Available online at `http://www.washingtonpost.com/blogs/dr-gridlock/wp/2013/09/09/i-95-drivers-get-to-compare-travel-times/`.

[101] T. Uchida, Y. Iida, and M. Nakahara. Panel survey on drivers' route choice behavior under travel time information. In *Proceedings Vehicle Navigation and Information Systems Conference (VNIS), 1994*, pages 383–388.

[102] Eric van Dijk and Marcel Zeelenberg. When curiosity killed regret: Avoiding or seeking the unknown in decision-making under uncertainty. *Journal of Experimental Social Psychology*, 43(4):656–662, 2007.

[103] David J. Wallace, Jeremy M. Kahn, Derek C. Angus, Christian Martin-Gill, Clifton W. Callaway, Thomas D. Rea, Jagpreet Chhatwal, Kristen Kurland, and Christopher W. Seymour. Accuracy of prehospital transport time estimation. *Academic Emergency Medicine*, 21(1):9–16, 2014.

[104] Fahui Wang and Yanqing Xu. Estimating o–d travel time matrix by google maps api: implementation, advantages, and implications. *Annals of GIS*, 17(4):199–209, 2011.

[105] Lin Wang and Da Young Ju. Concurrent use of an in-vehicle navigation system and a smartphone navigation application. *Social Behavior and Personality*, 43(10):1629–1640, 2015.

[106] N. Wanichayapong, W. Pruthipunyaskul, W. Pattara-Atikom, and P. Chaovalit. Social-based traffic information extraction and classification. In *2011 11th International Conference on ITS Telecommunications (ITST)*, pages 107–112.

[107] Yun Wei, Qing Shao, and Yin Han. Modeling of dynamic game-based route guidance system. In *Applied Mechanics and Materials*, volume 130, pages 1867–1871. Trans Tech Publ.

[108] Md Whaiduzzaman, Mehdi Sookhak, Abdullah Gani, and Rajkumar Buyya. A survey on vehicular cloud computing. *Journal of Network and Computer Applications*, 40:325–344, 2014.

[109] Hwasoo Yeo and Alexander Skabardonis. *Understanding Stop-and-go Traffic in View of Asymmetric Traffic Theory*, pages 99–115. Springer US, Boston, MA, 2009.

[110] Yu Zheng, Lizhu Zhang, Xing Xie, and Wei-Ying Ma. Mining interesting locations and travel sequences from gps trajectories. In *Proceedings of the 18th International Conference on World Wide Web*, WWW '09, pages 791–800, New York, NY, USA, 2009. ACM.

[111] P. Zhou, Y. Zheng, and M. Li. How long to wait? predicting bus arrival time with mobile phone based participatory sensing. *IEEE Transactions on Mobile Computing*, 13(6):1228–1241, June 2014.

[112] Shanjiang Zhu and David Levinson. Do people use the shortest path? an empirical test of wardrop's first principle. *PLOS ONE*, 10(8):e0134322, 2015.