AN ASSESSMENT OF IMPACTS OF CONNECTED VEHICLES ON TRAFFIC

OPERATION AND FUEL CONSUMPTION

A Dissertation

by

AREZOO SAMIMI ABIANEH

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Mark Burris
Committee Members,	Kumares Sinha
	Josias Zietsman
	Alireza Talebpour
Head of Department,	Robin Autenrieth

August 2020

Major Subject: Civil Engineering

ABSTRACT

Connected vehicle technology has the potential to improve the performance of the transportation system. The real-time traffic information communicated between connected vehicles and infrastructure enables a more efficient management and use of the transportation system. This new technology equips vehicles to receive real-time information from their surroundings, which can warn them of approaching congestion. The emergence of connected vehicles that facilitate data exchange among vehicles and infrastructure has the potential to improve mobility, increase safety, and reduce the harmful environmental impacts of travel. Unfortunately, most of the studies that evaluate the impacts of connected vehicle technology focus only on highway performance. Moreover, these studies often ignore the impacts of uncertainties and unpredictable factors in evaluating the connected environment performance. One of these uncertainties is the behavior of travelers. Unlike the autonomous vehicles' user, the driver of a connected vehicle has decision making power. This technology helps the users to make more informed decisions by providing real-time traffic information of the roads on their path. However, the users' acceptance rate in response to the given information is still a significant parameter affecting the overall performance of the transportation system.

In order to address the questions concerning the impacts of travelers' willingness to comply with the provided information, this research examined travelers' responses to the real-time rerouting information provided through connected vehicle technology and its impact. An internet-based survey was employed to investigate the impacts of different factors such as socio-economic characteristics, time saving, visibility of congestion, and the length of the trip on drivers'

willingness to follow the rerouting information. A total of 1881 complete responses were collected from a random nationwide sample of people living in small and medium-sized metropolitan areas (SMMAs). The results of the study indicated the importance of trip and socio-economic characteristics of users on their willingness to follow the rerouting advice. Comparison of the probability of the results of the study demonstrated a higher percentage of acceptance to the received information from connected vehicles compared to dynamic message signs and other similar technologies.

The research goal was to investigate the impacts of communication technology on the transportation system. There are three types of simulation frameworks in evaluating the network performance: microscopic, mesoscopic and macroscopic simulation models. Due to the significant effects of individual travelers on network performance, the microscopic simulation model is more appropriate. Most of the studies in this area focus on the highways or a small network. However, an accurate microsimulation model with different types of roads is required to be able to generalize the results for the real-world applications. To this end, Simulation of Urban Mobility (SUMO) is combined with Traffic Control Interface (TraCI) to create the network and conduct the study on the impacts of connected vehicles on traffic operation and fuel consumption in a large portion of the city of El Paso, Texas. The main objective of this study was to investigate the effects of providing the real-time information to the connected vehicles from two perspectives: traffic operation and fuel consumption. Different rerouting algorithms including rerouting based on 1. The real travel time information, 2. Energy (fuel) consumption, 3. Average travel time and 4. The modified travel time with its variance were used to determine the effectiveness of different methodologies for rerouting and to investigate the network performance under these scenarios.

Moreover, sensitivity analysis was carried out to assess the effects of important factors in the connected environment such as market penetration rate, driver's acceptance rate, congestion levels, the update interval for rerouting etc. The impacts of incidents were also considered as an important factor affecting traffic operation and fuel consumption at different market penetration rates (MPR) of connectivity. Several scenarios were implemented and tested on the simulation model to determine the impacts of incidents on the network performance in the urban area. The scenarios were defined by changing the duration of the incidents and the number of lanes closed. Finally, fuel-flow, flow-density and flow-speed diagrams were developed to characterize the macroscopic impacts of connected vehicles technology on the network performance.

The results of the study demonstrated the effectiveness of rerouting in improving traffic operations and reducing fuel consumption in a connected environment. The comparison between the four methods of rerouting demonstrated that the approaches involving the average travel time of individuals (real-time travel time and average travel time) resulted in higher efficiency of the network. The highest performance in terms of total travel time reduction and total decrease in fuel consumption over the network was observed a 20 percent of rerouting rate. The other two approaches (including the fuel consumption and the modified travel time which involve the variance of travelers' speed) resulted in smaller improvements to travel. For these two approaches higher rerouting rates (40% and 60% correspondingly) resulted in higher performance in these two models. Various scenarios including different duration and number of lanes closed were then simulated. The impacts of update interval on the performance of the network were also investigated. Several cases of incidents including different duration and one and two-lane closures on the freeway were examined. For the incidents, three durations of 900, 1500 and 2400 seconds and for the update interval three values of 150, 300, and 600 seconds were simulated and analyzed. As expected, an increase in the duration of incidents and number of lanes closed worsens the operation of the network. The study also showed that the update interval of 150 seconds resulted in higher performance of the network. The ability of connected vehicles to receive real-time traffic data helps the network perform more efficiently. The real-time traffic data provided by connected vehicle technology informs drivers of the routes with lower congestion and distributes travelers in the main and alternative paths which increase the throughput and efficiency of the network. This was demonstrated in the study using the macroscopic variables including flow, density and average speed. Overall, the ability of the connected vehicles in receiving the real-time traffic data improves the overall performance of the transportation network.

DEDICATION

To my beloved parents,
my brother, sister and sister-in-law
for their love, encouragement and support.

ACKNOWLEDGEMENTS

I would like to express my special appreciation and thanks to my advisor, Prof. Mark Burris for his excellent guidance, patience and support throughout my PhD studies. He has been a tremendous mentor for me who provided me technical guidance, scientific intuition and valuable comments that helped me to grow as a research scientist.

I am also grateful to the members of advisory committee Prof. Sinha, Dr. Zietsman, and Dr. Talebpour. I am so thankful to Prof. Sinha for conveying a spirit of adventure in regard to research, to Dr. Zietsman for his continuous support and providing opportunities for me to be involved in various research projects and to Dr. Talebpour for his invaluable insights and feedback.

I would also greatly thank my supervisors Dr. MohammadReza Farzaneh and Dr. Andrew Birt at Texas A&M Transportation Institute, and Dr. Gene Hawkins at Texas A&M university for providing me the opportunity to work with them.

Many thanks also go to my friends Elham, Parisa, Yalda, Mitra, Farinoush, Neda, Soheil, Mohammad, Sara, Niloufar, Ali, Hananeh, Ali, Maryam, Ashkan, Alireza and list goes on, for making my life so pleasant during the graduate studies.

Lastly, but most importantly, I would like to thank my beloved parents, Zahra and Hossein, my brother, Omid, sister, Elham and sister-in-law, Mehrnaz for their unconditional love. Without their encouragement and support, this dissertation would have never been completed.

CONTRIBUTORS AND FUNDING SOURCES

This work was supervised by a dissertation committee consisting of Professor Mark Burris (advisor) and Professor Kumare Sinha of the Civil and Environmental Engineering Department Dr. Josias Zietsman of the Urban Planning Department and Dr. Alireza Talebpour of the Civil and Environmental Engineering Department.

The data analyzed for Chapter II was provided by Professors Mark Burris and Kumares Sinha.

All other work conducted for the dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by a fellowship from Hagler Institute of Advanced Studies at Texas A&M University and Assistantships at Texas A&M Transportation Institute and Civil and Environmental Engineering Department at Texas A&M University.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

The transportation system is responsible for approximately 27% of the greenhouse gases (GHG) emitted in United States in 2015 (*1*). Fossil fuels utilized in transportation are the primary sources of GHG of which carbon dioxide is the main component. Carbon dioxide represents approximately 82.2% of the GHG emissions from human activities. The transportation sector was responsible for 34.5% of the total carbon dioxide production in 2015, where the largest source of carbon dioxide was the highway sector including passenger cars (42.3% of produced carbon dioxide), medium and heavy-duty trucks (23.6%), and light-duty trucks (17.1%). Overall, there has been an increasing trend in GHG emission production from 1990 to 2015 due to the increase in travel demand (*2*). Vehicular fuel consumption thus continues to play an essential role in urban GHG emissions production as long as fossil fuels remain the primary source of energy used in highway transportation.

More efficient and reliable use of the transportation system may reduce these emissions while also provide improved travel. To this end, the advances in wireless communication technology provide the potential to reach the goal of an interconnected network of vehicles and infrastructure in which the users can make better decisions in their use of the transportation system. Communication technology can enable vehicles to receive real-time information from their surroundings, which can warn them of approaching congestion. The emergence of connected vehicles that facilitate data exchange among vehicles and infrastructure has the potential to improve mobility, increase safety, and reduce the harmful environmental impacts from transportation systems.

In addition, connected vehicle technology is likely to be deployed in the short term and may provide the transportation system a transition path to reach the more efficient state of autonomous vehicles. The study and experiments of connected vehicle technology provides valuable input towards this transition state. In this research, the potential impacts of rerouting guidance strategy using connected vehicle technology on mobility and fuel consumption were examined. In a connected transportation environment, the individual vehicles communicate with other vehicles and the infrastructure systems including roadside devices and traffic management centers (TMC), to make more efficient trip decisions. This communication technology could both improve the mobility of the network and reduce the harmful environmental impacts of travel. This may lead to a safer system by enhancing driver's awareness. From the mobility perspective, traffic can be managed more efficiently by enabling vehicles to be informed about travel time and routes, which in turn helps travelers to select the best route for traveling. This could result in less congestion and improved traffic operation. From the environmental perspective, the communication technology provides the opportunity to manage the transportation system more efficiently and reduce emissions. Route guidance, dynamic signal timing, and more efficient driving cycles are possible with the help of the real-time information provided by connected vehicle technology. With the route guidance strategy, selecting less congested routes leads to less stop-and-go traffic, which in turn can reduce emissions.

In order to evaluate the network performance, three types of simulation frameworks have been used in the literature; microscopic, mesoscopic and macroscopic simulation models. The selection between these models depends on the level of detail required to achieve the objective of the study. Since individual vehicles have significant impact on the performance of the network, the microscopic simulation model was most appropriate for this research. Although many studies have been done investigating the impacts of connected vehicles on the transportation system, most of the studies focused on highways or a small network. However, an accurate microsimulation model with different types of roads is required to be able to generalize the results for the real-world applications. To this end, Simulation of Urban Mobility (SUMO) was combined with Traffic Control Interface (TraCI) to conduct this study on the impacts of connected vehicles on traffic operation and fuel consumption in the city of El Paso, Texas.

The eastern part of the city of El Paso was selected for several reasons. First, the area suffers from traffic congestion and travel improvements could therefore have significant benefits. Second, it encompasses different types of roads including freeway, arterial, and local roads. This provides operational information on a complex network. Third, this area is in a non-attainment area (an area with air quality worse than the National Ambient Air Quality Standards (NAAQS)) and evaluation of the potential use of CVs in this area can be beneficial for future decision making (93). Finally, El Paso is a medium sized metropolitan area. 273 out of 382 metropolitan areas in the United States are classified as small and medium-sized areas (SMMAs) (17). These areas are struggling with air quality problems more because automobiles are the dominant modes of transportation. These areas also have more flexibility to prepare the built environment to accommodate future changes in transportation as they are not as densly developed as the large cities that get much of the research focus. The information provided to travelers using connected vehicle technology might be effective and used in the short term to improve the transportation system performance and the air quality in these regions.

SUMO is an open-source microscopic traffic simulator, which is widely used among researchers. The effects of providing the real time information to the connected vehicles with the goal of selecting better routes to bypass congestion is assessed. Moreover, sensitivity analysis was carried out to assess the effects of important factors in the connected environment such as market penetration rate, driver's willingness to follow the rerouting advice, and congestion levels. In addition, the simulation of events on the roads that cause congestion, such as incidents and lane closures, were modeled. The ultimate goal was to examine the potential influence of connected vehicles on traffic operations and fuel consumption.

CHAPTER II

RESEARCH IMPACT

This research study provides guidance to transportation policy makers, such as officials from US Department of Transportation or local Metropolitan Planning Organization (MPOs). First, this study investigates the travelers' response to received information from the transportation system. Next, the impacts of the travelers' decisions on the networkwide performance of the transportation system may help policy makers to optimize the communicated information, with an objective of an efficient transportation system. Third, the results from the simulation models of connected vehicles provide an extensive view of the transportation network performance under several scenarios for policy makers. Based on this result, policy makers can decide to modify and improve the transportation system, such as increasing the capacity of some of the roads or changing the traffic control devices to prepare the network for this emerging technology. Finally, the developed microsimulation framework can be used to examine different scenarios in transportation system to detect limitations and potential improvements to the system. The developed microsimulation framework can also be used to implement new algorithms and test them to optimize the system before real-world implementation, which will reduce the cost of implementation. This study also provides a transition path to the more efficient transportation system of autonomous vehicles. The study simulates part of the city of El Paso in Texas. The city of El Paso is the medium sized city. About 75% of the population of the United States lives in Small and Medium-sized metropolitan areas (SMMAs). Most of the SMMAs are considered as non-attainment areas in terms of the air quality according to National Ambient Air Quality Standards (NAAQS). Therefore, an efficient management of the transportation systems might reduce the emissions produced and improve the air quality in these areas. In addition, the selected network includes various types of roads including highway, arterial, and local street that are helpful in generalizing the results to other networks as well. Finally, the developed survey which collected the willingness of travelers to follow the advice communicated by connected vehicles in these areas, can be used for other SMMAs for deployment of connected autonomous vehicles (CAVs).

CHAPTER III

RESEARCH OBJECTIVES

The primary research objective was to quantify the impacts of connected vehicles on mobility and energy consumption in a small and medium-sized metropolitan areas (SMMAs). More specifically, scenarios illustrating the potential impacts on traffic operations and fuel consumption on roads with the help of communication technology were defined and evaluated. This study addressed some of the limitations of the studies on connected vehicles found in the literature. First the data used in the study, including the survey, the results of the survey and data used in the model are unique. Previous studies focused on large metropolitan areas. However, 75% of the population of the United States lives in SMMAs. These areas also have a great opportunity for realignment to meet the infrastructure needed for the emergence of connected and autonomous vehicles. This study provides an insight to the acceptance rates of drivers to reroute in response to information received from connected vehicles in SMMAs. Then the methodology used for evaluating the impacts of connected vehicle technology at the network level was provided and evaluated. Sensitivity analysis was done to investigate the impacts of the parameters used in the study on mobility and fuel consumption. Finally, the impacts of the connected vehicles at the macroscopic level was evaluated to demonstrate the overall performance of the network. This study provides a comprehensive analysis of the impacts of rerouting strategies and traveler response to the guidance in a connected environment that can be used for future analysis in the area of connected and automated vehicles (CAVs).

CHAPTER IV

LITERATURE REVIEW

Travelers' Response to Rerouting Advice

Surveys of travelers have shown that drivers would like to get route guidance information. However, the reported rerouting (to the suggested routes) rarely exceeded 40%. Cummings (*3*) found that geographic and traffic conditions affect the travelers' response to variable message signs (VMS); only 4 to 7% of the travelers typically switch their routes due to the received information from the VMS. Ramsay and Luk (*4*) investigated the route choice behavior of travelers using real-time traffic information. They estimated that up to 30% of travelers will reroute if the traffic congestion information is provided. A graphical traffic information and control software system was employed by Davidson and Taylor (*5*) to investigate travelers' choices from a set of alternative messaging and controlled strategies. The software enabled the operator to locate the incident and also any blockage on the major roads. The authors claimed that 6 to 41% of travelers switched to alternative routes to bypass congestion in Sweden. Tsirimpa and Polydoropoulou (*6*) found that 54.3% of travelers switched their routes in response to the congestion reported by VMS in Athens. Based on these studies, there are many potential reasons why people do not adjust their travel due to the en-route suggestions. For some ITS technologies like VMS, drivers may overlook the messages (*7*), may not trust the messages, may be uncertain about travel time on the alternative route, or may not even understand the messages (*8*).

The impacts of different factors on the travelers' behavior in response to traffic information were also investigated, e.g., (*6–9*). Bonsall (*10*) found that the total trip duration, toll roads, familiarity with the network, congestion, safety, security, delays and costs affect individuals' behavior in response to route guidance information. Bonsall and Palmer (*11*) found that the wording of the messages had a significant impact on drivers' willingness to follow the rerouting advice. Lai et al. (*12*) also demonstrated that socioeconomic factors such as gender, age, education and also the display characteristic of the messages shown on ITS devices were important parameters in drivers' willingness to reroute. Li et al. (*13*) found that drivers with more risk-based driving styles were more likely switch their route in case of congestion compared to drivers with more conservative driving styles. Dia and Panwai (*9*) demonstrated that factors such as commuters' socioeconomic characteristics, the degree of familiarity with the network, and the expectation of travel time savings influence the drives' response to route guidance information.

The above literature on travel behavior studies has focused on metropolitan areas with populations greater than 500,000, e.g., (*14*, *15*). However, the regional characteristics and city type are important factors that also influence travelers' responses to rerouting advice (*16*). Moreover, 273 out of 382 metropolitan areas in the United States are classified as small and medium-sized areas (SMMAs) (*17*). Hence, the findings of this study, with a focus on SMMAs, provides additional insight to the implementation of connected vehicle (CV) technology. In the rest of this section, the literature on modeling travelers' behavior or choice models were reviewed.

Logit models have been used for decades to model traveler behavior, e.g., (*18*, *19*, *19–23*). The advantages of these models include a well-defined mathematical structure of the models and also the possibility of interpreting the estimated parameters to provide significant insight into the independent variables influence over the result. The downsides include assumptions that restrict the applicability of the models to various analyses. New advancements in computational science have allowed for some machine learning models to be used in traveler behavior studies. The nonlinearity in the problems and the ability of the machine learning techniques to identify the complicated boundaries for classification, reveal that machine learning techniques can sometimes be superior to the traditional statistical models. In addition, many machine learning techniques can quickly analyze much larger datasets (*24*). Moreover, the traditional statistical models use rule-based programming that result in deterministic mathematical equations to predict the output given the input. However, machine learning methods do not necessarily follow an explicit program and can learn from the data.

A review of the types of models used in the travel behavior/mode choice literature shows two major categories of ordinal and nominal logit models. The ordinal models are classified to three sub-categories including ordered logit model, ordered probit model, and ordered mixed logit models. On the other hand, there are three common nominal discrete response models that have been used in the literature including multinomial logit models, nested logit models and mixed logit models. In general, the multinomial logit model, mixed logit model and ordered probit model are the most frequently used models for travel behavior studies, e.g., (*22, 23*). Each of these models has its own benefits and limitations. The ordered models have the benefit of providing one coefficient in the model for each parameter for all of the response categories. However, the proportional odds assumption and parallel regression assumption restrict the usage of these models in statistical analysis (*25*). Although multinomial logit models do not have the issues of the ordered models, they bring two main challenges to its applicability. The first issue is the major assumption of Independence of Irrelevant Alternative (IIA). This property claims that in the selection among a set of alternatives, the odds of selecting one alternative (A) over another (B) should not depend on the existence of another irrelevant alternative (C). In addition, the multinomial logit model is not capable of modeling randomness or variations among individuals. The mixed logit model does not have those limitations mentioned for the ordered and multinomial logit models. It can also accommodate preference heterogeneity which improves the realism of the developed model (*26*).

With advances in computational science, there is growing interest in using machine learning methods for travel behavior studies. Recent studies using the machine learning algorithms found that these techniques were effective in modeling travel behaviour (*27–33*). In the area of travel mode choice modeling, a number of studies compared the accuracy of discrete choice models with machine learning methods. For instance, Xie et al. (*32*) compared the performance of two data mining methods including Decision Tree and Artificial Neural Network with multinomial logit model for work travel mode choice modeling. The authors found that the two data mining approaches provide better results compared to the multinomial logit model. They presented that Decision Tree demonstrated highest efficiency among the applied models and artificial neural network provides better prediction performance. Hagenauer and Helbich (*31*) compared seven machine learning classifiers for travel mode choice modeling. They claimed that the random forest performed better compared to other selected machine learning techniques. The authors believed that since the importance of variables are different from method to method, the analysis of variables is important for effective travel behavior modeling. A comparison of the machine learning techniques including Support Vector Machine (SVM) and Artificial Neural Network (ANN) with multinomial logit model demonstrated a better performance of the machine learning algorithms in modeling travelers' behaviour (*28*). However, machine learning was not superior in all studies. Zao et al. (*34*) found that the random forest method produced lower accuracy than the multinomial logit model in travel mode choice modeling. Travelers' responses to rerouting advice suggested by connected vehicle technology can be categorized as a classification problem and statistical models and supervised machine learning techniques can be used to train a response model for studies evaluating the effectiveness of connected vehicle technology.

In this study, the travelers' responses to rerouting advice were also investigated using stated preference data. Due to the successful results of both logit models and machine learning techniques in the previous studies, both techniques were used to develop models of the travelers' behaviour. Logit models including Ordered Logit and Probit, Multinomial Logit and Mixed Logit models and machine learning techniques including Decision Tree, Support Vector Machine and Multilayer Perceptron were used to develop models of travelers responses to the rerouting suggestions communicated by connected vehicles technology based on the survey response (see chapter V).

Connected Vehicles

Communication technology enables traffic data exchange among vehicles and infrastructure (Vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I)). This data exchange has the potential to improve traffic operation, air quality, and safety (*35–39*). Recently, many researchers have investigated various rerouting strategies with the help of communication technologies to improve the network performance. To this end, the impacts of traveler information systems and rerouting strategies on the performance of the network have been investigated in many studies that will be described in this section.

Oh and Jayakrishnan (*40*) investigated the impacts of advanced traveler information system (ATIS) on travel time and the simulation model's output showed that the average network travel time (NTT) was reduced by 25% when 40% of the drivers received real time information and all of them rerouted. Abdulhai & Look (*41*) evaluated the impacts of dynamic route guidance systems and safety-enhanced route guidance systems on the network-wide safety and travel time. A microsimulation model integrated with some accident prediction models were developed. The study concluded that the increase in the dynamic route guidance market penetration rate generally reduced the average travel time of travelers during an incident and the safety route guidance strategy performs better from the safety perspective by suggesting routes with lower crash risk to travelers. Lee and Park (*42*) conducted a study evaluating the impacts of route guidance strategies with the existence of incidents. A microsimulation model and two incident scenarios on two directions of the freeway were modeled in VISSIM and the impacts of route guidance strategies on travel time, vehicle miles traveled, and average speed were investigated. The authors conducted sensitivity analysis on the factors including market penetration rate (MPR), acceptance rate, congestion level and update intervals of route guidance strategies. The results of the study demonstrated the effectiveness of route guidance strategies on network performance. The environmental impacts of the same model of rerouting guidance strategies in a connected environment were then investigated in another study (*43*). The simulation results showed that this rerouting system that was found to be effective for enhancing traffic operation is also effective in improving air quality and fuel consumption. The impacts of communication technology on traffic operation were evaluated by implementing dynamic route diversion strategies and variable speed limit control during severe congestion (*44*). The results of the study demonstrated some evidence of sensitivity to the MPR and the implemented control strategies in the model. Yeo et al. (*45*) investigated the network traffic operation in response to the lane blocking freeway with and without communication technologies. The conclusion of the study confirms the effectiveness of deploying the connected vehicles technology in reducing delays and improving traffic operation.

IntelliDrive vehicles (Vehicle-to-infrastructure) were developed in Paramics microsimulation model by Dion et al. (*46*) to simulate the dissemination of messages between vehicles and roadside devices. The results demonstrate the dependency of the quality of the data collection on the market penetration rates. A decentralized ATIS was developed in the connected environment by Kim (*47*) with the use of Automatic Incident Detection (AID). The results of the study demonstrated that ATIS using vehicle-to-vehicle communication reduced travel time and that AID had an important role in making the system efficient.

Another study that evaluated the impacts of communication technology during an incident was done by Kattan et al. (*48*). Two application programming interfaces (API) were developed for simulating incidents and warning drivers of incidents to increase their awareness and reduce their aggressiveness. The impacts of congestion level and market penetration rates (MPR) on travel time savings were evaluated. The results support the conclusion of the effectiveness of connected vehicles in improving safety and reducing travel time for moderate and high congestion levels. Paikari et al. (*49*, *50*) evaluated the benefits of deploying connected vehicles using the Paramics microsimulation model . These studies demonstrated the effectiveness of connected vehicle technologies in reducing travel time, improving traffic operation and incident occurrence rate. Olia et al. (*39*) investigated the impacts of real-time routing guidance strategies and warning messages on network performance from different aspects of safety, environment and traffic operation. In their study, they assume that nonconnected vehicles also reroute using the information they received from other control devices like dynamic message signs, GPS or seeing the congestion ahead of them.

Xiong et al. (*51*) evaluated the impact of information provision and en-route decisions with the use of variable message signs. Using an agent-based model integrated with a dynamic traffic assignment model, an incident scenario was designed along with ITS devices, including variable message signs to demonstrate the en-route diversion behavior of the drivers on network performance. The network fundamental diagram was used to represent traffic dynamics.

The rerouting strategies have been investigated extensively in the literature. Many different algorithms of rerouting using information among vehicles and infrastructure devices were developed and tested using microsimulation models to assess their ability to improve the efficiency of the network (*38, 52–54*). Most of the models selected an urban freeway or a small network (*55*). However, the congestion caused by incidents affect the performance of other routes as well. In addition, assessing the effects of different parameters like incident duration and update interval of connected vehicles on the operation of the network is fundamental to have a better view of how frequent these data exchanges should be done to improve the performance of the transportation system. The present study attempts to fill these gaps.

Network Fundamental Diagram

The connected vehicle technology affects the performance of individual vehicles. Therefore, in order to evaluate the impacts of connected vehicle technology on the transportation system performance, the information from individual vehicles is important. However, the overall performance of the network requires investigating traffic flow, density and speed. These can be used as the parameters demonstrating the efficiency of large networks. Therefore, in this study the network fundamental diagram was used as a graphical method to evaluate the impacts of communication technology and congestion warnings on the overall performance of the network.

Daganzo (*56*) reintroduced the concept of Macroscopic Fundamental Diagram (MFD) or

Network Fundamental Diagram (NFD). Speed, density and flow are the most common variables to characterize a traffic stream. For networks with homogenous, well-connected, and uniformly distributed demands, the existence of a network fundamental diagram has been proved in theory (*56*, *57*), simulation (*58*) and in the real world (*59*, *60*). One of the first studies in a large scale onsite experiment to demonstrate this relationship was done by Daganzo and Geroliminis (*57*). In addition, the network fundamental diagram was illustrated using a large simulation model by Ji et al. (*58*). Traditionally, traffic variables at a network level were mostly computed using link measurements with the data given by loop detectors. With the use of GPS data, and more recently connected vehicle data, vehicle trajectory data is more widely available and can be used to estimate traffic variables. By estimating traffic variables based on The Eddie's methodology (*61*) using vehicle trajectories, we can derive the network fundamental diagram for complex networks.

Courbon and Leclercq (*62*) mentioned three methods for estimating a network fundamental diagram: the analytical method, the trajectory based method and the loop detector data. The first method was applied on simple networks (*57, 63*) as well as on more complex networks (*64*, *65*) to characterize the shape of the network fundamental diagram. In order to estimate the average flow, density and speed from link measurements, equations 1, 2 and 3 can be used (*66*).

$$Q = \frac{\sum_{i=1}^{M} l_i q_i}{\sum_{i=1}^{M} l_i} \tag{1}$$

$$K = \frac{\sum_{i=1}^{M} l_i k_i}{\sum_{i=1}^{M} l_i} \tag{2}$$

$$V = \frac{\sum_{i=1}^{M} l_i v_i}{\sum_{i=1}^{M} l_i} \tag{3}$$

Where:

Q, K, V : Networkwide average flow, density, and speed respectively

q_i, k_i, v_i : Individual link average flow, density, and speed respectively, for observation period

l_i : Length of lane link i

M : Total number of lane links

The easiest way to derive the NFD using the trajectory method is to get the vehicle trajectories from the simulation tools (*62*). Although having the trajectories of all vehicles in the network can simplify calculating traffic states, in the real world, acquiring such data is almost impossible. Some studies used a vehicle's probe data to derive the NFD (*67*, *68*). Saberi et al. (*69*) has extended the Eddie's method for estimating macroscopic traffic variables by three-dimensional vehicle trajectories in order to construct the fundamental diagram.

Emission and fuel consumption models

There are several available emissions models in the literature, including Motor Vehicle Emission Simulator (MOVES) (*70*), Virginia Tech Comprehensive Power-based Fuel Consumption Model , Comprehensive Modal Emission Model (CMEM) (*71*, *72*), Virginia Tech microscopic (VT-Micro) emission model (*73–75*), Handbook emission factors for road transport (HBEFA) (*76*), Passenger Car and Heavy Duty Emission Model (PHEM) etc. The first four models are commonly used in the United States and would provide a better estimate of the emissions production and fuel consumption in El Paso because the models are developed based on common vehicles in the United States. These models are described in the following paragraphs.

MOVES is the U.S. Environmental Protection Agency's emission modeling system (*70*). Four sets of outputs from a microsimulation traffic model are needed to estimate the emissions production of vehicles simulated in MOVES. The first one is the link average speed during the time of simulation. The second one is link instantaneous speed based on a second-by-second evaluation. The third one is the vehicle trajectory data including length, speed, acceleration, and location on a second-by-second basis. The last one is the overall average speed and volume during the entire hour. More information like the current age distribution of vehicles on the road and meteorology data of the study area can help to estimate a more accurate emission production. Since it is hard to estimate the total emissions at the network level using MOVES, a simplified version of MOVES was developed based on the emission rates of several base driving cycles and some modification factors. The model was then validated by the results of MOVES (*77–79*). The formulation of the model is as follows:

$$CE_{p,c} = \sum_v \{[\sum_a (EF_{p,b,a,v} \times CCF_{p,c,a,v} \times f_{a,v})] \times f_v\} \qquad (4)$$

Where:

type v

c: cycle c

b : base cycle

p : Pollutant

$$CCF_{p,c,a,v} = \left(\frac{(\sum_m f_m^c \times ER_{p,a,v,m})}{(\sum_m f_m^b \times ER_{p,a,v,m})}\right)\left(\frac{v^b}{v^c}\right) \qquad (5)$$

Where:

$ER_{p,a,v,m}$: default emission rate for pollutant p, age a, vehicle type v and in operating mode bin m

f_m^c : fraction of time in operation mode bin m in cycle c

f_m^b : fraction of time in operation mode bin m in cycle b

v^c : cycle average speed for cycle c

v^b : cycle average speed for cycle b

The second model, VT-CPFM estimates fuel consumption based on the driving cycles and the vehicle features. This model has the merits of (1) not switching abruptly between two states of fuel consumption and zero consumption without any middle state and (2) it has more flexibility for calibrating the parameters using publicly available highway and city data (*80*). Two equations for estimating fuel consumption with the similar structure were presented. Although both of them need vehicle related features for calibrating the parameters in the model, one of them requires fewer vehicle specific parameters with the cost of lower accuracy. The following formulation is used to estimate fuel consumption. The model parameters are calibrated based on the vehicle's feature, which increase the accuracy of the results due to considering the vehicles specific characteristics.

$$FC(t) = \begin{Bmatrix} \alpha_0 + \alpha_1 \times P(t) + \alpha_2 \times P(t)^2 & P(t) \geq 0 \\ \propto_0 & P(t) < 0 \end{Bmatrix} \quad (6)$$

$$P(t) = \left(\frac{R(t) + 1.04 \times m \times a(t)}{3600 \times \eta_d}\right) \times v(t) \quad (7)$$

$$\propto_0 = max\left(\frac{P_{mfo} \times \omega_{idle} \times d}{22164 \times QN}, \frac{\left(F_{city} - F_{hwy}\frac{P_{city}}{P_{hwy}}\right) - \varepsilon\left(P_{city}^2 - P_{hwy}^2 \times \frac{P_{city}}{P_{hwy}}\right)}{T_{city} - T_{hwy}\frac{P_{ciy}}{P_{hwy}}}\right) \quad (8)$$

$$R(t) = \frac{\rho}{25.92} C_D C_H A_f v(t)^2 + 9.8066m \frac{C_r}{1000}(C_1 v(t) + C_2) + 9.8066mG(t) \quad (9)$$

$$F_{city} = T_{city}\alpha_0 + P_{city}\alpha_1 + {P_{city}}^2\alpha_2 \quad (10)$$

$$F_{hwy} = T_{hwy}\alpha_0 + P_{hwy}\alpha_1 + {P_{hwy}}^2\alpha_2 \quad (11)$$

Where:

FC *: Fuel consumption at time t* ($^{l}/_{sec}$)

$P(t)$ *: Power exerted at any instant t*

m *: Vehicle mass (kg)*

$a(t)$*: Acceleration of the vehicle at time t* $\left({}^{m}/_{s^2}\right)$

η_d *: Driveline efficiency or the efficiency with which the vehicle transfers its power from the motor to the wheels (%)*

P_{mfo} : $Idling\ fuel\ mean\ pressure(400{,}000\ Pa)$

ω_{idle} : $Idling\ engine\ speed\ (rpm)$

d : $Engine\ displacement\ (L)$

Q : $Fuel\ lower\ heating\ value(43{,}000{,}000\,\frac{J}{kg}\ for\ gasoline\ fuel)$

N : $Number\ of\ engine\ cylinders$

$F_{city}\ and\ F_{hwy}$: $Fuel\ consumed\ for\ the\ EPA\ city\ and\ highway\ drive\ cycles\ (liters)$

$P_{city}\ and\ P_{city}^2$: $Sum\ of\ power\ and\ power\ squared\ exerted\ each\ second\ over\ the\ entire\ cycle$

$P_{hwy}\ and\ P_{hwy}^2$: $Sum\ of\ power\ and\ power\ squared\ exerted\ each\ second\ for\ the\ highway\ cycle$

ε : $ensures\ that\ the\ second\ order\ parameter(\alpha_2)\ is\ greater\ than\ zero$

$v(t)$*: The velocity at time t* $({}^{m}/_{s})$

$R(t)$ *: Resistance function determined by rolling resistance*

ρ*: Density of air at sea level at temperature of* $15\,^{\circ}C$

C_D*: Vehicle drag coefficient*

C_H*: Correction factor for altitude*

A_f *: Frontal area of the vehicle* (m^2)

$C_r, C_1, and\ C_2$ *: Coefficients associated with rolling*

$G(t)$ *: Grade at time t*

$\alpha_0, \alpha_1, \alpha_2 : vehicle\ specific\ model\ constant$

$T_{city}\ and\ T_{hwy}: Durations\ of\ the\ city\ and\ highway\ cycle$

The third model, CMEM, is a microscopic emission and fuel consumption model, which employs a physical, power-demand approach based on a parameterized analytical representation of fuel consumption and emissions production (*71*, *72*). The input data of CMEM includes vehicle activity (second-by-second speed trace), and fleet composition of traffic. The VT-Micro software is a microscopic vehicle emission-modeling tool, which is used to estimate the amount of emissions of various pollutants including CO2, NOx, CO, HC and fuel consumption. The input of the model is the second-by-second vehicle's speed profile (*73–75*). An image of the software is shown in Figure 1.

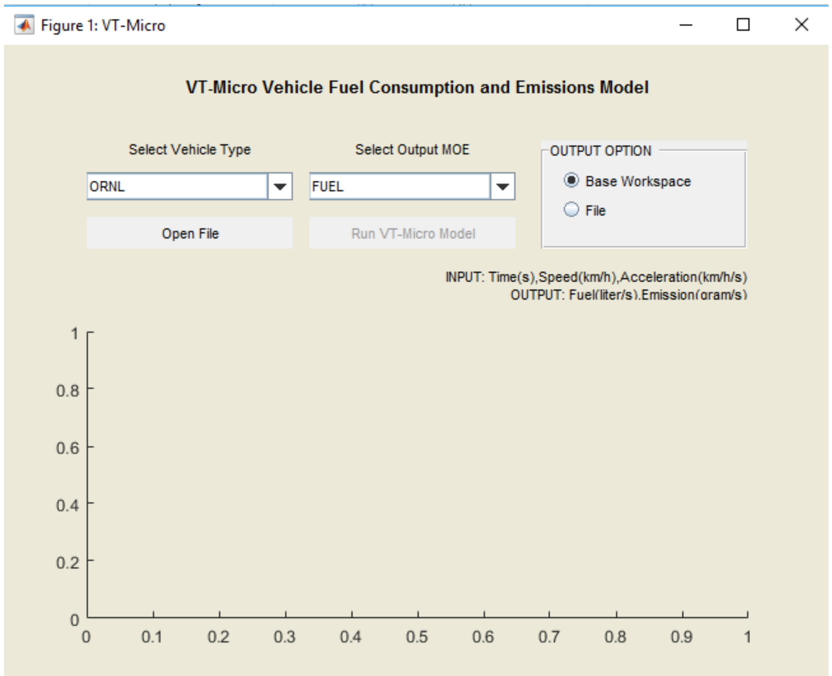


Figure 1. VT micro emission modeling software

Since VT-micro model was developed based on the data from the older vehicles years. Therefore, the emission's models which can be updated based on the vehicles years and adapted to more recent vehicles are more appropriate for the fuel consumption and emissions estimation. In this study, the VT-CPFM model was used to estimate fuel consumption (*81*). It provides the capability of calibrating the parameters based on the vehicle characteristics which has the merit of additional model parameter calibration compared to the MOVES model. This model gives the estimated fuel consumption.

CHAPTER V

DATA COLLECTION

Traveler behavior

To evaluate the effectiveness of connected vehicle technology on the overall performance of the network, questions concerning the travelers' willingness to reroute to the provided information must be addressed. To this end, an internet-based questionnaire organized by the research team and administered by a commercial firm, LightSpeed Research, was used to gather potential traveler responses to rerouting advice. The target population was randomly selected from small and medium size metropolitan areas (SMMAs) in the United Sates. These areas were selected since as approximately 75% of the population of the United States lives in these areas (*17*) and commuters are highly dependent on automobiles in these areas. Second, SMMAs have more opportunities in urban growth and realignment for preparing the infrastructure for the emergence of connected vehicle technology. Third, El Paso is a medium sized metropolitan area and this research examines travel in El Paso.

A draft questionnaire was developed and tested twice on a group of 100 respondents with various levels of education, age, gender, and profession. Based on feedback from the respondents, the questions were modified to ensure the survey was easily understood. The survey was also kept fairly short (less than 10 minutes for the majority of respondents) and thus survey fatigue should not be a factor in the quality of the responses. The final version of the survey was then distributed among the selected target population. 4625 participants were introduced to the questionnaire. Two screening questions were asked at the beginning of the survey to ensure the eligibility of the respondents: (1) Do you own/lease/have access to a vehicle? and (2) Do you use a vehicle for travel to work or school?

A positive response to both of these questions confirmed the eligibility of the participant. The survey firm also ensured only adults over 18 years old living in small to medium size communities took the survey. Based on these criteria, 2111 of the potential 4625 participants were allowed to take the survey. After the survey was completed, several steps were performed to remove any invalid responses. First, incomplete responses were removed from the final set of data. Second, the amount of time spent by each respondent to answer the questions on each section of the survey were recorded to identify the potential inaccurate responses. Surveys completed too quickly were also removed from the results, yielding 1881 valid responses.

As mentioned earlier, one application of the connected vehicle technology is to provide information regarding congestion on routes and to suggest alternative route(s) to bypass the congestion. The survey included two potential scenarios that could take place when receiving the congestion alert: visible and not yet visible congestion on the road ahead. For both scenarios, the respondents were asked to assume that they were notified of congestion on the road ahead by the communication technologies while driving. In the first scenario, the respondents were also asked to imagine that they can see some congestion on the road ahead. Then the travel time on the current route (a random number between 30 and 50 minutes) and the time saved (a random number between 10 and 15 minutes) by following the alternative route were shown to respondents. The second scenario was almost identical to the first scenario except that the congestion was not yet visible to the respondent. The travel time (a random number between 30 and 50 minutes) and the time saved (a random number between 5 and 10 minutes) were shown to the respondents. There are five possible responses to each scenario from the participants: (1) I would take the alternative route, (2) I would likely take the alternative route, (3) I am not sure, (4) I would probably not take the alternative route, and (5) I would definitely not take the alternative route. Other survey questions included questions regarding the respondents' current travel behavior and the socio-demographic characteristics. The results are presented in the following sections. The stated preference question was as follows:

For the recent trip you made (described in the early part of the survey), imagine you are driving a connected vehicle that is receiving traffic information from other vehicles on the road. What would you do if:

You see no congestion on the road but your vehicle warns that there is congestion ahead in 2 miles. It estimates a xx minute trip if you stay on your current road, or a yy minute trip on a different road. How likely would you be to switch to the different road? (xx and yy were generated randomly such that xx was between 5 to 10 minutes greater than yy.)

You see some congestion on the road ahead and your vehicle warns you that there is congestion ahead. It estimates a AA minute trip if you stay on your current road, or a BB minute trip on a different road (AA and BB were generated randomly such that AA was between 5 to 10 minutes greater than BB).

Now that you can see some congestion, how likely would you be to switch to the different road? It should also be noted that the connected vehicles were defined in the survey for the respondents. (see Appendix A)

The responses of travelers to their willingness to reroute with the information provided by connected vehicle technologies for two cases of not seeing and seeing the congestion on the road ahead are displayed in Figure 1. In the case where respondents could not see the congestion, around

70% of travelers would accept the rerouting suggestion over all of the assumed saving times (5-10 minutes). Only around 16% of travelers would likely ignore the rerouting suggestion.

As shown in Figure 2, for the case of visible congestion on the road ahead, a higher percentage of positive responses to the rerouting suggestion was observed. As demonstrated, for this case around 85% of travelers accepted the rerouting suggestion over the assumed saving times (10-15 minutes). Only around 5% preferred to stay on their original route.

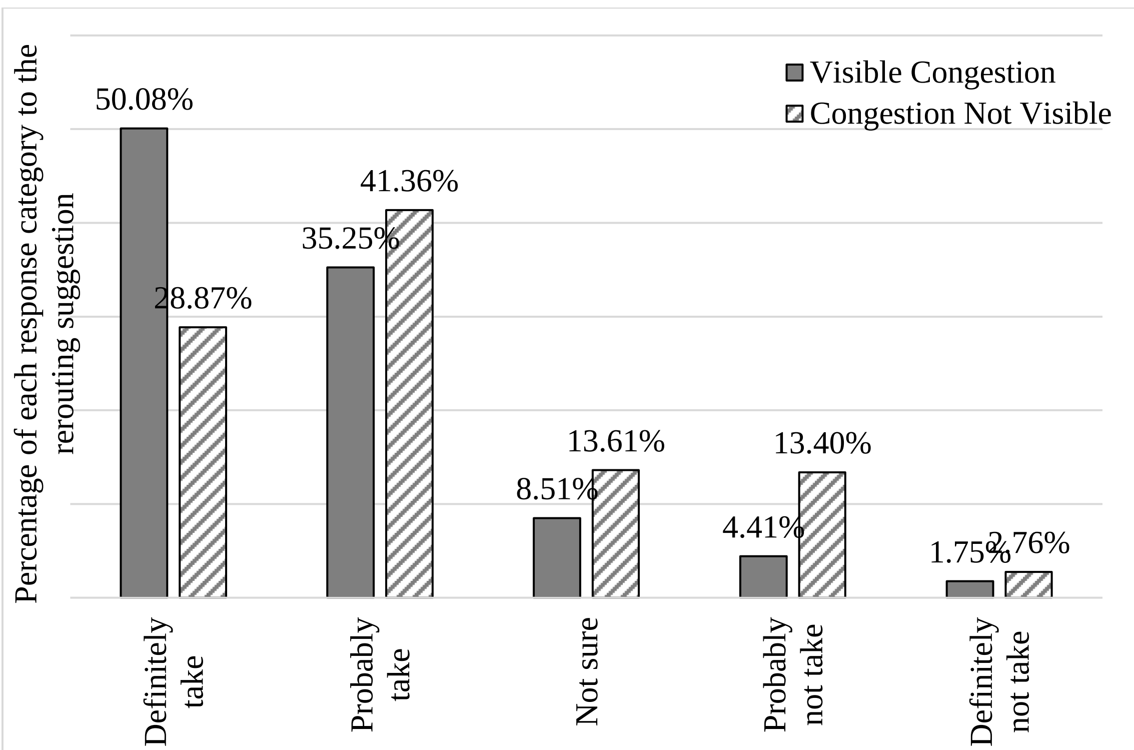


Figure 2. Acceptance Rate (a) congestion not visible (b) visible congestion

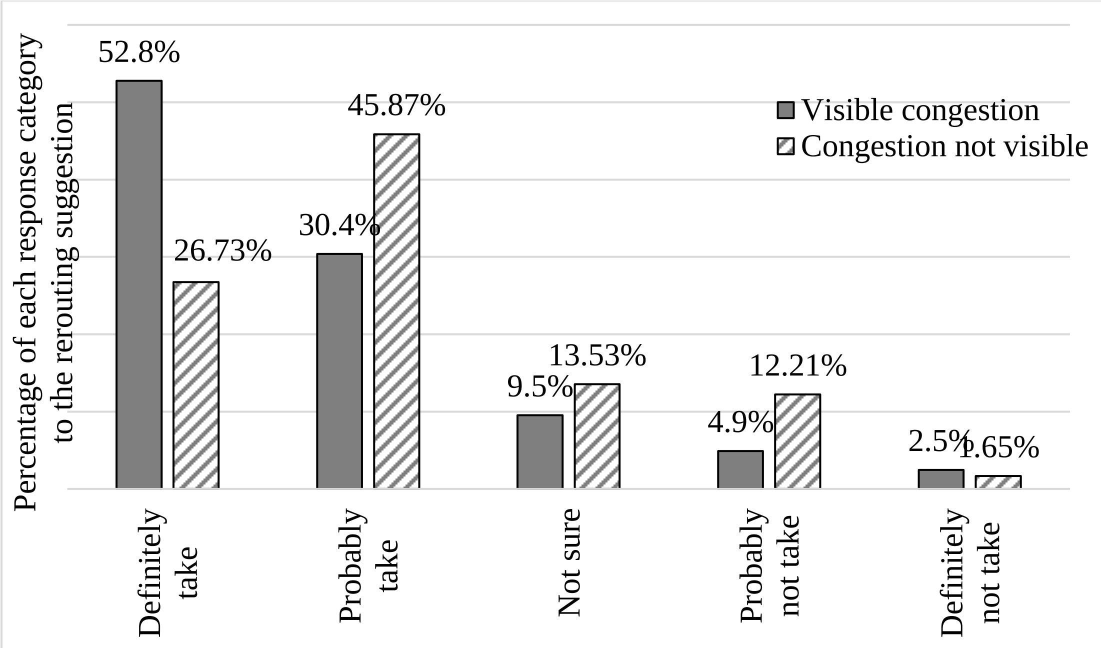


Figure 3. Acceptance Rate (Travel time saving of 10 minutes)

Since for the visible congestion, the travel time saving was in the range of 10 to 15 minutes, and for the congestion not visible this range was 5 to 10 minutes, the percentage of travelers accepting the rerouting suggestion for both cases for the travel time saving of 10 minutes was compared. Figure 3 demonstrates the percentage of travelers in each category of the responses for two cases of visible and not yet visible congestion. As shown in the figure, the percentage of travelers who accepted the rerouting suggestion in the case of visible congestion was around 83% and the percentage of travelers who accepted the rerouting suggestion in the case of congestion not visible was around 72% which demonstrates the importance of visibility of congestion in travelers' responses to the traffic information.

The list of attributes and their associated statistics in the study are shown in Table 1. It should be noted that not all the respondents answered every question, for example the income question, Also, for some questions, the respondents could select more than one answer, for instance employment. Thus, not all number of responses totaled 1881.

Table 1. The list of explanatory variables in the model

Variables	Choices	Number of Responses	Percentage
Respondents	Driver	1718	91%
	Passenger	163	9%
Distance to work place	Trip miles in [0, 1)	72	4%
	Trip miles in [1-5)	490	26%
	Trip miles in [5,10)	456	24%
	Trip miles in [10,15)	272	15%
	Trip miles in [15,20)	174	9%
	Trip miles in [20,25)	109	6%
	Trip miles in [25,50)	200	10%
	Trip miles ≥ 50	108	6%
Smartphone	using a smart phone	1635	87%
	not using	246	13%
Enjoy Driving	Yes	1671	89%
	No	210	11%
Gender	Male	1098	59%
	Female	783	41%
Employment Status	Full time	990	53%
	part-time	417	22%
	Not employed	159	8%
	Retired	303	16%
	Student	66	4%
Household size	1	447	24%
	2	786	42%
	3	299	16%
	4	238	13%
	≥ 5	111	4%
Number of households' members under 18 years old	0	1445	77%
	1	232	12%
	2	149	8%
	3	41	2%
	4	9	0%
	≥ 5	5	0%
Education	<High School	13	1%
	High School	313	17%
	College	613	33%
	BS/BA	553	29%
	Professional	53	3%

Table 1 (continued). The list of explanatory variables in the model

Variables	Choices	Number of Responses	Percentage
Education	Professional	53	3%
	MS/MA	286	15%
	PhD	50	3%
Income	[0, 25000)	228	12%
	[25000, 50000)	468	25%
	[50000, 75000)	413	22%
	[75000, 100000)	307	16%
	[100000, 200000)	318	17%
	≥ 200000	61	3%
	No Answer	86	5%
Place of living	City center	160	9%
	Urban outside of the city center	368	20%
	Suburban area	1004	53%
	Rural area	349	19%
Age group	[18, 25)	126	7%
	[25, 35)	204	11%
	[35, 45)	278	15%
	[45, 55)	342	18%
	[55, 65)	521	28%
	≥ 65	410	22%
Trip Duration	<20 min	850	45%
	[20, 40)	662	35%
	[40, 60)	175	9%
	[60, 90)	106	6%
	≥ 90	88	5%
Heard about connected vehicles	Yes	742	39%
	No	1139	61%
Travelling during peak hour period	Yes	859	46%
	No	1022	54%

Modeling Methodology

To investigate travelers' willingness to change route due to the traffic information provided by connected vehicle technology, several models were built based on the survey data. Both driver and passenger data were included since passengers often have significant navigation duties while the driver focuses on the driving task. For this purpose, three discrete choice models and three machine learning techniques were used. The ordered probit model, multinomial logit model and mixed logit model were the three discrete choice models selected for this study. The three machine learning techniques included Decision Trees, Support Vector Machine (SVM) and Artificial Neural Network (ANN). All models attempted to predict the respondents' response to the rerouting advice. These three techniques are the most popular strategies for supervised machine learning and classification. In this section, a brief explanation of each model is provided.

Discrete Choice models

Ordered Models: As mentioned earlier, the responses to the questionnaire followed a five-point likert scale. Therefore, the ordered logit and probit models were examined first due to the ordered nature of the responses. In order to develop an ordered logit model, an unobserved variable (z) was defined based on a linear function for each observation (*82*):

$$z = \beta X + \varepsilon \tag{12}$$

The y responses are determined using the following relationships:

$$y = 1 \qquad if\ \ z \le \mu_0 \tag{13}$$

$$y = 2 \qquad if\ \mu_0 < z \le \mu_1 \tag{14}$$

$$y = 3 \qquad if\ \ \mu_1 < z \le \mu_2 \tag{15}$$

$$y = 4 \qquad if\ \ \mu_2 < z \le \mu_3 \tag{16}$$

$$y = 5 \qquad if\ \ z \ge \mu_3 \tag{17}$$

Where X is the vector of variables for estimating the discrete ordering of the observations, β is the vector of parameters, ε is the random error assumed to follow the Normal distribution with mean 0 and variance 1 for probit model and logistic distribution with mean 0 and variance 1 for logit model, and μs are thresholds for defining y variables which are estimated jointly with β. The problem then becomes the estimation of the probability for each response of each observation. In equation 5, μ_0 can be set to zero without loss of generality; Therefore, three thresholds should be calculated. For both the ordered probit and logit models, these probabilities are estimated as follows:

$$P[y=1] = \Phi[-\beta X] \tag{18}$$

$$P[y=2] = \Phi[\mu_1 - \beta X] - \Phi[-\beta X] \tag{19}$$

$$P[y=3] = \Phi[\mu_2 - \beta X] - \Phi[\mu_1 - \beta X] \tag{20}$$

$$P[y=4] = \Phi[\mu_3 - \beta X] - \Phi[\mu_2 - \beta X] \tag{21}$$

$$P[y=5] = 1 - \Phi[\mu_3 - \beta X] \tag{22}$$

Where $\Phi(\mu) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\mu} EXP[-\frac{1}{2}\omega^2]\, d\omega$, is the cumulative Normal distribution of a variate ω.

To estimate the values of the parameters in the ordered models, the maximum of log-likelihood function over all observations should be calculated. The equation for the log-likelihood function is as follows:

$$\mathrm{LL} = \sum_{n=1}^{N} \sum_{i=1}^{I} \delta_{in} \mathrm{LN}[\,\Phi(\mu_i - \beta X_n) - \Phi(\mu_{i+1} - \beta X_n)] \tag{23}$$

S.t. $\quad 0 \le \mu_1 \le \mu_2 \le \ldots \le \mu_{I-1}$

Where N is the number of observations, I is the highest integer ordered response and δ_{in} is calculated based on the following equation:

$$\delta_{in} = \begin{cases} 1 & \text{The observed response is i} \\ 0 & \text{Otherwise} \end{cases} \tag{24}$$

The explanatory variables (Table 1) were examined for possible inclusion in the model.

Several factors were considered, including coefficient size, engineering judgment and significance level. Another issue was the assumption of having fixed parameters for all of the parameters in the model. This assumption limits the individual specific disturbances and can cause an erroneous parameter estimate. To tackle this problem, random effects models were used, similar to previous studies (*82*, *83*). By using random effects models, two terms for the disturbances will be considered in equation (4); one is the traditional disturbance term (ε) and the other one is the individual specific random disturbances (φ_n). This variance is computed as part of the random effects model and demonstrates the significance of the random effects model compared to the standard ordered probit model. This also helps to account for reduced variance due to a single respondent answering multiple survey questions. With this assumption, equation 1 was modified as follows:

$$z = \beta X + \varepsilon + \varphi_n \tag{25}$$

There were four main assumptions for using the ordered logit and probit models. The first assumption was that the response variable should be an ordered item (Likert item). The second assumption was that the ordinal predictor variables were treated as continuous or categorical not ordinal. The third assumption was the assumption of not having multicollinearity among independent variables. Finally, two specifications that restrict the use of the ordered models are proportional odds assumption and parallel regression assumption (*25*). The proportional odds assumption is only applied to the ordered logit model. This assumption supports the idea that the log-odds of each outcome differs with any other by a constant. Since this issue is raised only for the ordered logit model, most researchers prefer to use an ordered probit model which does not have this restrictive issue. The parallel regression assumption is applied to both ordered logit and probit models. One of the well-known tests for determining the validity of this assumption is the Brant test. Brant test is a popular test for reviewing the parallel regression assumption (*84*, *85*). The test considers J-1 binary models constructed by defining a variable $z_{ij} = 1\ if\ y_i > j$. The model restricts the probabilities of the binary models to have the same coefficient vector β while having a different constant term. This Brant test can be adopted to both ordered logit (logistic distribution) and ordered probit (Normal distribution). If the results of the Brant test demonstrate that the parallel regression assumption is valid, one set of coefficients can define the model effectively. Otherwise, different coefficients should be used to model each outcome group (UCLA, 2018). In this case the positive or negative values for the β coefficients do not necessarily represent the increase or decrease in the likelihood of agreeing/disagreeing to the choices defined. To this end, the marginal effects of each category should be determined to have an accurate sense of the impacts of the assumed variables on the interior categories. For indicator variables, the marginal effects are calculated as the difference in the estimated probabilities while value of the variable changes from 0 to 1. For computing the marginal impacts of the continuous variables, the partial derivatives are used (*82*):

$$\frac{\partial P(y=1)}{\partial X} = -\phi(-\beta X)\beta' \quad (26)$$

$$\frac{\partial P(y=2)}{\partial X} = [\phi(\mu_0 - \beta X) - \phi(\mu_1 - \beta X)]\beta' \quad (27)$$

$$\frac{\partial P(y=3)}{\partial X} = [\phi(\mu_1 - \beta X) - \phi(\mu_2 - \beta X)]\beta' \quad (28)$$

$$\frac{\partial P(y=4)}{\partial X} = [\phi(\mu_2 - \beta X) - \phi(\mu_3 - \beta X)]\beta' \quad (29)$$

$$\frac{\partial P(y=5)}{\partial X} = -\phi(\mu_3 - \beta X)]\beta' \quad (30)$$

Where P(y=i) is the probability of response category i, and ϕ is the standard Normal density. The marginal effect is defined as the change in the probability of the responses for each threshold category given a unit change in the explanatory variables. Therefore, a positive marginal effect demonstrates an increase in the probability for the corresponding response and a negative value for the marginal effect represents a decrease in the probability of the response for the unit increase of the explanatory variable. Furthermore, a large value for the marginal effect represents large effects on the users' response and the small value of it, demonstrates relatively small impacts on the users' choice (*82*).

Multinomial Logit Model: The other model developed in this study is the multinomial logit model. The multinomial logit model is a well-known model when the response variables are categorical with more than two options. The multinomial logit model is developed based on a stochastic utility function consisting of two systematic and random parts. The systematic portion of the utility function is estimated using a linear function of the predictor variables and the random part is estimated using a logistic distribution. The general equation for the multinomial logit model is:

$$p(y = j) = \frac{\exp(\beta_j \, x_t)}{\sum_{m=1}^{J} \exp(\beta_j \, x_t)} \, , j = 0, \dots , J \tag{31}$$

Where p(y=j) is the probability of response category j, β_j is the vector of estimable parameters for discrete outcome I, and x_t is the vector of the explanatory variables.

Mixed Logit Model: The mixed logit model has the ability to capture randomness in parameters in the model. In addition, the mixed logit model provides more flexibility compared to the multinomial logit model which is restricted by the assumption of independence from irrelevant alternatives (IIA) (82). The features of the mixed logit model enabling the unobserved factors to follow any distribution, make the mixed logit model applicable in any discrete model estimation. Another weakness of the standard multinomial logit model that is addressed with the mixed logit model is the ability of this model to allow the parameters to vary across observations (82). The general form of the mixed logit model, which is the weighted average of the standard multinomial logit model with the weights determined by the density function, is as follows:

$$P_n^m(i) = \int P_n(i) f(\beta|\varphi) d\beta \qquad (32)$$

Where $P_n^m(i)$ is the mixed logit model probabilities of observation n with discrete outcome I, $P_n(i)$ is the probability of observation n having discrete outcome of I, $f(\beta|\varphi)$: The density function of β with φ as the vector of parameters of the density function (mean and variance).

The choice models were developed using Nlogit 5. First, the ordered probit (Table 2) and logit (Table 3) models were developed based on the simulation-based maximum likelihood method using 500 Halton draws for parameter estimation. Numerous combinations of fixed and random parameters were selected and tested. A t-test was used to verify the randomness of the parameters. A comparison between the two ordered probit and logit models demonstrated similar results in terms of significant parameters and goodness of fit test. However, the number of significant parameters in the ordered probit model were higher compared to the ordered logit model. The marginal effects of the parameters of ordered porbit model is reported in Table 4. As mentioned earlier, in order to verify the accuracy of the parameter estimates, the parallel regression assumption should be tested. For this purpose, we used the Brant test and the results of the Brant test rejected the validity of the parallel regression assumption (Table 5).

Table 2. Likelihood of changing route, from 5=Definitely Change to 1= no change using Ordered Probit Model

Explanatory variables	Parameter estimates	t-statistic
Non-random parameters		
Constant	1.22***	0.00
Trip miles (1 if travel to workplace takes less than 5 miles, 0 otherwise)	.11**	0.01
Smartphone (1 if the respondent uses a smartphone, 0 otherwise)	-.18***	0.00
Number of persons in households under 18	-.10***	0.00
Income (1 if higher than 25,000, 0 otherwise)	-.23***	0.00
Random parameters		
See the congestion ahead (1=visible congestion, 0=otherwise)	-.63***	0.00
Enjoy driving (1 if enjoy driving, 0 otherwise)	-.20***	0.00
Location of living (1 if location of living is city center, 0 otherwise)	-.20***	0.00
Employment (1 if employed full-time, 0 otherwise)	-.09**	0.02
Age group (1 if aged between 18 and 34 years, 0 otherwise)	-0.18***	0.00
Threshold 1	1.20***	0.00
Threshold 2	1.73***	0.00
Threshold 3	2.61***	0.00
AIC	9313.4	
Log-likelihood function at convergence	-4638.68	

***, **==> Significance at 1%, 5% level

Table 3. Likelihood of changing route, from 5=Definitely Change to 1= no change using Ordered Logit Model

Explanatory variables	Parameter estimates	t-statistic
Non-random parameters		
Constant	1.99***	0.00
Trip miles (1 if travel to workplace takes less than 5 miles, 0 otherwise)	.18**	0.02
Smartphone (1 if the respondent uses a smartphone, 0 otherwise)	-.29***	0.00
Number of persons in households under 18	-.16***	0.00
Income (1 if higher than 25,000, 0 otherwise)	-.38***	0.00

Table 3 (continued). Likelihood of changing route, from 5=Definitely Change to 1= no change using Ordered Logit Model

Explanatory variables	Parameter estimates	t-statistic
Random parameters		
See the congestion ahead (1=visible congestion, 0=otherwise)	-1.01***	0.00
Enjoy driving (1 if enjoy driving, 0 otherwise)	-.35***	0.00
Location of living (1 if location of living is city center, 0 otherwise)	-.32***	0.01
Employment (1 if employed full-time, 0 otherwise)	-.13**	0.00
Age group (1 if aged between 18 and 34 years, 0 otherwise)	-0.29***	0.04
Threshold 1	1.93***	0.00
Threshold 2	2.82***	0.00
Threshold 3	4.58***	0.00
AIC	9312.7	
Log-likelihood function at convergence	-4638.3	

***, **==> Significance at 1% and 5% level

Table 4. Marginal effects of the explanatory variables in the ordered probit model

Explanatory Variables	Marginal effects of taking the alternative routes				
	Definitely take	Probably take	Not sure	Probably not take	Definitely not take
Non-random parameters					
Trip miles (1 if travel to workplace takes less than 5 miles, 0 otherwise)	-0.0425	0.0118	0.01137	0.0137	0.0034
Smartphone (1 if the respondent uses a smartphone, 0 otherwise)	0.0668	-0.0165	-0.0219	-0.0226	-0.0058
Number of persons in households under 18	0.0388	-0.0119	-0.0123	-0.0119	-0.0028
Income (1 if higher than 25,000, 0 otherwise)	0.0837	-0.0200	0.0275	-0.0288	-0.0075
Random parameters					
See the congestion ahead (1=visible congestion, 0=otherwise)	0.2360	-0.0702	-0.0736	-0.0736	-0.0186
Enjoy driving (1 if enjoy driving, 0 otherwise)	0.0792	-0.0207	-0.0257	-0.0262	-0.0066
Location of living (1 if location of living is city center, 0 otherwise)	0.0766	-0.0283	-0.0230	-0.0207	-0.0045
Employment status (1 if being full-time employment, 0 otherwise)	0.033	-0.0101	-0.0105	-0.0102	-0.0024
Age group (1 if aged between 18 and 34 years, 0 otherwise)	0.0703	-0.0247	-0.02141	-0.0197	-0.0044

Table 5. Brant Specification Test for equal coefficient vectors in the ordered probit model

	Brant Test		Coefficients in implied model Prob(y>j)			
Explanatory variables	**ChiSquare**	**Pvalue**	**0**	**1**	**2**	**3**
See the congestion ahead (1=visible congestion, 0=otherwise)	10.78	.01	-.93	-.93	-1.08	-.47
Trip miles (1 if travel to workplace takes less than 5 miles, 0 otherwise)	1.25	.74	.16	.20	.11	.27
Smartphone (1 if the respondent uses a smartphone, 0 otherwise)	6.00	.11	-.20	-.42	-.24	-.49
Enjoy driving (1 if enjoy driving, 0 otherwise)	6.61	.09	-.30	-.43	-.23	-.38
Number of persons in households under 18	1.15	.77	-.14	-.17	-.22	-.22
Income (1 if higher than 25,000, 0 otherwise)	23.26	.00	-.27	-.53	-.12	-.65
Location of living (1 if location of living is city center, 0 otherwise)	3.51	.32	-.28	-.21	-.41	-.95
Age group (1 if aged between 18 and 34 years, 0 otherwise)	1.42	.70	-.26	-.27	-.17	-.43
Employment (1 if employed full-time, 0 otherwise)	5.86	.12	-.15	-.03	-.12	.24
Chi squared test statistic	62.241					

The multinomial logit model (Table 6) and mixed logit model (Table 7) were then developed in Nlogit 5. As discussed in the previous section, the mixed logit model not only has the ability to capture randomness in the parameters' estimation, but also does not suffer from the major assumption of IIA as in the multinomial logit model. The comparison between the results of these two models (using the R-squared values) demonstrated a higher performance of the mixed logit model compared to the discrete choice models. Table 7 contains the results of the mixed logit model. The values in parenthesis are the standard deviation of the estimated random parameters with Normal distribution.

Table 6. Multinomial logit model

Explanatory variables	Parameter estimates	t-statistic
Definitely take the new route		
Constant		
Trip miles (1 if travel to workplace takes less than 5 miles, 0 otherwise)	-0.19**	0.0283
Saving time (minutes)	0.20***	0.0000
Enjoy driving (1 if enjoy driving, 0 otherwise)	0.55***	0.0000
Number of households under 18	0.17***	0.0001
Income (1 if higher than 200,000, 0 otherwise)	0.38***	0.0459
Education (1 if the education level is master's degree, 0 otherwise)	0.25**	0.0295
Employment status (1 if being full-time employment, 0 otherwise)	0.18**	0.0085
Location of living (1 if City center, 0 otherwise)	0.32***	0.0076
Probably take the new route		
Constant	0.47**	0.0136
See the congestion ahead	-0.67***	0.0000
Saving time (minutes)	0.22***	0.0000
Enjoy driving (1 if enjoy driving, 0 otherwise)	0.38***	0.0001
Education (1 if the education level is PhD, 0 otherwise)	0.32***	0.0001
Not sure if taking the new route or not		
Constant	0.59**	0.0420
See the congestion ahead	-0.73***	0.0002
Saving time (minutes)	0.16***	0.0000
Smartphone (1 if the respondent uses a smartphone, 0 otherwise)	-0.60***	0.0000
Education (1 if the education level is less than high school, 0 otherwise)	1.46***	0.0016
Probably not take the new route		
Constant	1.33***	0.0000
Smartphone (1 if the respondent uses a smartphone, 0 otherwise)	-0.35**	0.0320
Age group (1 if aged between 35 and 45 years, 0 otherwise)	0.34**	0.0264
Education (1 if the education level is less than high school, 0 otherwise)	1.23**	0.0234
Definitely not take the new route		
Constant	0.95***	0.0101
Smartphone (1 if the respondent uses a smartphone, 0 otherwise)	-0.60**	0.0273
Education (1 if the education level is PhD, 0 otherwise)	-0.75**	0.0490
Household size	-0.29**	0.0097
AIC	9353.1	
Log-likelihood function at convergence	-4811.58	
R-squared (R^2)	0.034	

***, **==> Significance at 1%, 5% level

Table 7. Mixed logit model

Explanatory variables	Parameter estimates	t-statistic
Definitely take the new route		
Trip Duration	-1.11*** (0.71)	0.00
Trip miles (1 if travel to workplace takes less than 5 miles, 0 otherwise)	-0.19***	0.03
Saving time (minutes)	0.48***	0.00
Enjoy driving (1 if enjoy driving, 0 otherwise)	0.66***	0.00
Number of households under 18	0.15***	0.00
Income (1 if higher than 200,000, 0 otherwise)	0.54**	0.01
Education (1 if the education level is master's degree, 0 otherwise)	0.35***	0.01
Employment status (1 if being full-time employment, 0 otherwise)	0.17***	0.02
Location of living (1 if City center, 0 otherwise)	0.31***	0.01
Probably take the new route		
Alternative Specific Coefficient	0.27	0.16
See the congestion ahead	-0.92***	0.00
Saving time (minutes)	0.53***	0.00
Enjoy driving (1 if enjoy driving, 0 otherwise)	0.48***	0.00
Education (1 if the education level is PhD, 0 otherwise)	0.43***	0.00
Not sure if taking the new route or not		
Alternative Specific Coefficient	0.45	0.13
See the congestion ahead	-0.99***	0.00
Saving time (minutes)	0.48***	0.00
Smartphone (1 if the respondent uses a smartphone, 0 otherwise)	-0.64***	0.00
Income (1 if higher than 200,000, 0 otherwise)	0.74**	0.01
Education (1 if the education level is less than high school, 0 otherwise)	1.56***	0.00
Probably not take the new route		
Alternative Specific Coefficient	3.25***	0.00
Driver/Passenger (1 if the respondent is driver)	-9.38*** (3.88)	0.00
Smartphone (1 if the respondent uses a smartphone, 0 otherwise)	-7.53*** (6.94)	0.00
Education (1 if the education level is less than high school, 0 otherwise)	2.79***	0.02
Definitely not take the new route		
Alternative Specific Coefficient	9.75***	0.00
Smartphone (1 if the respondent uses a smartphone, 0 otherwise)	-2.77***	0.00
Household size	-2.44***	0.00
AIC	9390.4	
Log-likelihood function at convergence	-4665.22	
R-squared (R^2)	0.23	

***, **==> Significance at 1%, 5% level

Machine Learning Techniques

Decision Tree: Decision trees are non-parametric supervised learning methods applicable to both classification and regression analysis. This approach is one of the most practical methods in supervised learning problems. It uses an algorithm to determine a pattern in the data to split it into multiple parts. The objective of this method is to build a model to predict the value of the response variables by learning the rules achieved from the given attributes. The basic algorithm which is used in the decision tree is called ID3. This algorithm creates the decision tree using a greedy approach. The first step for this approach is to find the best attribute. This is measured using a statistical property called 'information gain' which demonstrates how well the attributes separate the data into groups. To define this property, a commonly used measure in information theory representing the impurity in a group of datasets, is used. This measure is known as entropy (*86*).

$$Entropy(T) = -p_+ \, log_2 \, p_+ - p_- \, log_2 \, p_- \quad (33)$$

Where T is a sample of training set, p_+ is the proportion of positive examples in T and p_- is the proportion of negative examples in T.

The information gain can then be measured using the Entropy:

$$Gain(T, A) = Entropy(T) - \sum_{v \in Values(A)} rac|T_b||T| \times Entropy(T_b) \quad (34)$$

Where A is the selected attribute, Gain (T, A) is the information gain of attribute A relative to the training sample T, Values(A) is the set of all values of attribute A in the given dataset, T_b is the subset of the sample of training that have the value of b and $rac|T_b||T|$ represents the fraction of examples belong to T_b.

Equation 26 indicates that the information gain is measured as the difference between the entropy of the parent node and the average entropy of the children nodes. In this way, the information gain for each attribute is calculated and the attribute with highest information gain is selected for further analysis. This selection is done at each node of the tree and continued until the algorithm classified all the training data and also all the attributes were used.

Support Vector Machine: Support vector machine (SVM) is another supervised learning algorithm which can be used for both classification and also regression, though more promising for classification. SVM uses a separating hyperplane in an N-dimensional space to classify the data. The dimension of the hyperplane depends on the number of features in the data and the location of it is selected with the aim of maximizing the distance to the nearest training data. There are two types: linear and non-linear SVM. SVM found the optimal hyperplane for linearly separable patterns and extend it for the data that are not linearly separable by transforming the original data into the new space. This method takes the low dimensional input space and transforms it to a higher dimensional space using kernel functions and then find out a way to separate the new transformed data. Reducing the dimensionality were used to be common for analyzing the data. However, for SVM transforming the data to a higher dimensional space is beneficial in creating a larger space providing the opportunity of finding a separating hyperplane. This method is used for non-linear separation problem.

The ultimate goal of this approach is to determine the hyperplane represented in equation (25) while maximizing the margin between the linear decision boundaries. The parameter of a Normal vector (w) and bias (b) that define the decision boundaries, are estimated through a learning process on the training dataset.

$$y(x) = w^T x + b = 0 \tag{35}$$

The support vectors are the data points that are close to the hyperplane and have significant impacts on the position of the hyperplane. The optimal hyperplane can be found using the objective function demonstrated in equation (26):

$$minimze\ \lambda||\omega||^2 + C\sum_i \max\big(0,1 - y_i(\omega^T x_i + b)\big) \quad (36)$$

Where
λ is the regularization parameter

ω is the weight vector (feature vector)

y_i is the actual response

$\omega^T x_i + b$: function representing the predicted response

The first term in the loss function (equation 26) which is called as the hard margin SVM tries to minimize the $||\omega||^2$ which is equivalent to maximizing the margin. The second term which is called the soft margin, penalizes the misclassification. The term λ as the regularization was added to the loss function to avoid overfitting by penalizing the large coefficients in the vector of solution.

Artificial Neural Network: The third approach selected for this study is the Artificial neural network (ANN). ANNs are inspired by the structure of the biological neural systems and originates from the neuro and computer science fields and currently are being used in many other disciplines (*87*). The basic elements of the ANNs are the neurons which are arranged in layers. The neurons of each layer are interconnected to all the neurons in the next layer. The input layer of the ANN model consists of the explanatory variables as a set of neurons. The output neurons are the dependent variables or the response variables. The difference of the ANN model with the logistic regression model is with the hidden layers that handle the non-linearity relationship of the data. Each neuron in the hidden layer created by the weighted linear summation of the neurons in the previous layer followed by a non-linear activation function applied to the calculated sum. These nodes are connected using the links with weights that are learned through the model development from the given data. The advantages of the ANN models are: (1) ability in learning non-linear models, and (2) ability to perform online learning. The disadvantages of these models are: (1) non-convex loss function (2) requiring tuning a number of hyperparameters, and (3) sensitivity to feature scaling (*88–91*).

A comparison of these models suggested that the decision tree techniques have the advantage that no real hyperparameters need to be tuned except the number of trees (*92*). On the other hand, there are many parameters need hyper-tuning for SVM and ANN. Comparing the artificial neural network and support vector machine revealed that there is not any clear differences between these two methods from the accuracy perspectives. On some datasets, ANN works better and, on some others, SVM results in a more accurate classification model. In this study, the aforementioned machine learning methods were developed to train a model that classifies human responses falling into 5 categories of "definitely take a new route", "probably take a new route", "Unsure if taking a new route or not", "probably not taking a new route" and "definitely not taking a new route". Given the set of features collected using the survey, the proposed models learned travelers' decision and classified them into the five categories mentioned earlier.

Three machine learning techniques described earlier were implemented in Python to investigate the travelers' responses to rerouting advice communicated by connected vehicle technology. As mentioned in the methodology section, Decision Tree does not have any real hyperparameter tuning except the maximum depth. The maximum depth in the decision tree model demonstrated how deep the tree can be. As the tree gets deeper, more attributes will be provided for the model development and more splits the tree will have. In order to determine the maximum depth of the tree, we examine the range from 1 to 32 for the maximum depth and 10_fold cross validation was used to estimate the average accuracy of the model over the validation dataset. The maximum accuracy of 0.59 achieved at the maximum depth of 23.

SVM model were then implemented in Python. The first parameter for Hyper-tuning in the SVM model is Kernel. Kernel parameter defines the type of hyperplane which is used to separate the data. Linear and non-linear hyperplane can be used for the model development. In this study we tested three types of hyper-plane including the linear, rbf and poly. Gamma is another parameter for non-linear hyperplanes. The higher the gamma value is, the more the model tries to fit the training data. The values of {0.1, 1, 10, 100} were examined. When the hyperplane is set to poly, another hyperparameter need to be considered is called degree. This parameter determines the degree of polynomial for finding the hyperplane for splitting the data. Regularization term in another important parameter that need to be determined in classification using SVM. One of the major aspects of training the data using machine learning technique is to assure that the model does not overfit the training dataset. If overfitting happened, the data will have a very low accuracy on the test data. Regularization is one method of avoiding overfitting by regularizing the coefficients estimates toward the zero. The regularization parameter in the model makes a tradeoff between having a smooth decision boundary and classifying the training data accurately. In this study, we used the values of {0.1, 1, 10, 100, 1000} for tuning the regularization parameter.

The data was then classified using the Artificial Neural Network. Similar to what has been done for the SVM, there are many hyperparameters in the ANN model that need to be tuned. The following categories were examined for tuning each hyperparameters: hidden layer size: {(10,10,10), (10,20,10), (10,), (30,30,30,30), (20,20,20),(20,10,10), (20,20,10,10)}, activation function for the hidden layer: {'tanh' , 'relu'}, solver: {'lbfgs', 'sgd', 'adam'}, alpha (regularization term): {0.0001,0.01, 0.05, 0.1,1}, learning rate: {constant, adaptive}. The best parameters found results in having a neural network with two hidden layers of sizes (30,30,30,30), with alpha 0.05, and Adam as the solver of the model with the activation function of Relu and adaptive learning rate.

In order to examine how well these models performed on the data, precision, recall and F1-score were estimated. The formulations of these evaluation techniques were provided in equations 27, 28 and 29. True Positive is the number of correctly predicting a class in the problem. For example, for the first category of "definitely take", the true positive is the number of observations where the actual response and predicted response are both "definitely take". The False Positive for the first category means the number of observations where the predicted value of the observation is "definitely take", but the actual value is not associated to the first category of response ("Definitely Take"). The false negative of the first category of response also occurred when the actual category of the response is "Definitely Take" and the model classified it as the other categories of response. For instance, the actual value of the class is "definitely take" and the predicted value assigned the observation to other classes of response. Figure 4 demonstrates the precision and recall of the response classes using the three machine learning techniques. It should be noted that the F1-score value provided in equation 29 represents a weighted average of the precision and recall.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \tag{37}$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{38}$$

$$F1 - score = \frac{2\times(Precision\times Recall)}{Precision+Recall} \quad (39)$$

The results demonstrated that SVM and ANN performed better compared to DT for most of the categories. In order to have a trade-off between the precision and recall, F1 scores were also estimated. The results are shown in Figure 5. Based on F1 score, it was found that the F1-score was higher for SVM in 4 categories of the response variable.

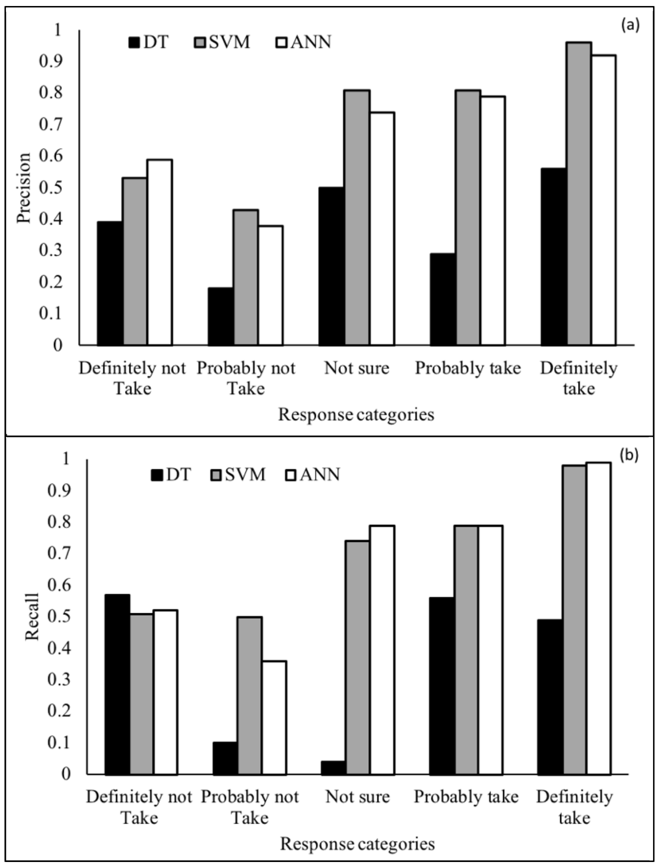


Figure 4. Precision (a) and Recall (b) of the three machine learning models

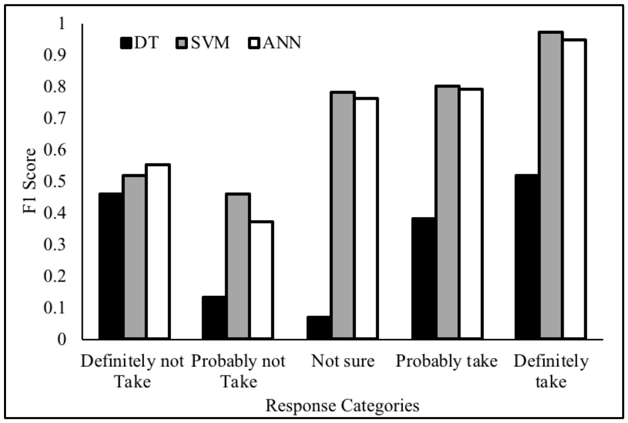


Figure 5. Comparison of the accuracy of the machine learning models using F1-score

Comparison of Machine Learning Techniques and Discrete Choice Models

In this section the predictive accuracy of the discrete choice models applied in this study including the multinomial logit model and mixed logit model were compared to the machine learning algorithms. To compare the models in terms of minimizing overall prediction error, an evaluation criterion was identified. This criterion is often theoretical measures such as adjusted R2, AIC and BIC or resampling based measures such as cross validation and bootstrapping techniques. Generally, resampling based measures are more common to compare different approaches (*34*). To this end, the K-Fold Cross Validation approach was used to compare the results of the traditional statistical analysis with the machine learning techniques.

Cross validation is a resampling technique used to evaluate the machine learning models. One parameter K is defined in the model which refers to the number of groups the data is divided into. In this method, the data is divided into K subsets and each time k-1 subsets are used for training the model and one is used for the test. The overall error is estimated using the average of all the k model estimations. This technique reduces the bias and variance of the model. It reduces the bias since it used most of the data for fitting and also it reduces the variance because it also used most of the data for validation as well. As a rule of thumb, 5 or 10 categories are commonly used for K-Fold Cross Validation.

Here a 10-Fold Cross Validation was applied to the multinomial logit model, mixed logit model, Decision Tree, Support Vector Machine and Artificial Neural Network. The accuracies of the models are shown in Figure 6. The standard deviation of the accuracy of each model is represented by lines. Based on the results of the cross validation, both multinomial logit model and mixed logit model have similar performance and the accuracy of the models developed using these techniques is high. However, the Rho value of mixed logit model is higher than the multinomial logit model in this study (Table 6 and Table 7). The machine learning techniques demonstrated a relatively lower performance compared to the discrete choice models. The results of the comparison among these machine learning models demonstrated that support vector machine performed better compared to decision tree and artificial neural network.

Discussion on the impacts of explanatory variables on Travelers Acceptance Rate

As demonstrated in the previous section, the multinomial and mixed logit models performed better in predicting the travelers' responses to rerouting advice. Seeing the congestion ahead was a significant parameter for two response categories, including: probably take the new route and not sure if taking a new route (see Table 7). The negative value of this parameter in the aforementioned categories may be due to travelers who see congestion selecting they would definitely take the new route. The short distance trips (trips less than 5 miles) was also a significant parameter for the response category of definitely take the new route. The negative value of this parameter confirms that when the distance of travel is short, the traveler rarely reroutes to the alternative path for saving time. Saving time, resulting from changing to the new route and enjoy driving were also significant parameters for two categories of definitely take the new route and also probably take the new route. The positive value of these two parameters demonstrates that the increase in the saving time results in higher tendency for changing the route and also the probability of rerouting for travelers who enjoy driving is higher. The use of a smart phone was also a significant parameter in the three categories of definitely not take the new route, probably not take the new route and not sure if taking the new route. The negative value of this parameter demonstrates that the respondents who do not use smart phone have higher likelihood of not changing the route. This can be explained by the fact that respondents who do not use smartphones probably feel new technology is not as beneficial or useful as those that own smart phones. Therefore, it does not seem that they are as accepting of the information communicated by connected vehicles. Different categories of education were also significant in the mixed logit model. The estimated coefficients for education represent that the higher level of education results in the higher probability of rerouting to the alternative routes. Another significant variable in the model is whether the respondent was a driver or passenger. This variable was significant for probably not take the new route. The negative sign of the coefficient states that the passengers have higher tendency to change the route which seems to be reasonable. Passengers are more able to navigate the new route and check maps than the drivers.

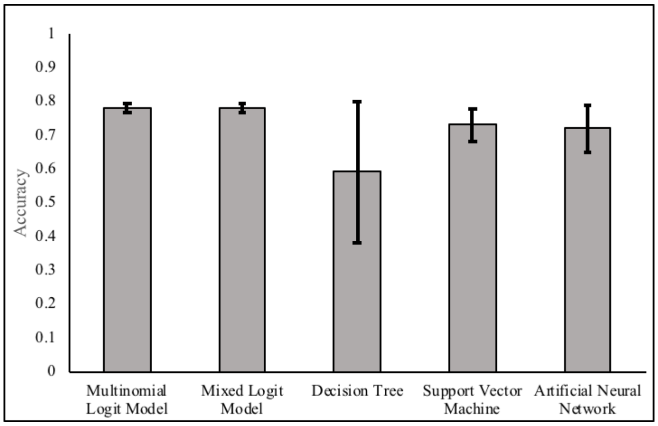


Figure 6. Comparison of the accuracy of the models using cross validation

The outcome of this study showed that in the case where respondents could not see the congestion ahead, around 70% of travelers would accept the rerouting suggestion over all of the assumed saving times (5-10 minutes). Only around 16% of travelers would likely ignore the rerouting suggestion. For the case of visible congestion on the road ahead, around 85% of travelers accepted the rerouting suggestion over the assumed saving times (10-15 minutes). Only around 5% preferred to stay on their original route.

The results of this study provide an insight into the travelers' acceptance rate. As mentioned earlier, the connected vehicle technology provides travelers with the real-time traffic information. Therefore, the drivers of connected vehicles can make the decision to follow the advice and reroute or just continue their last path. This study helps to estimate the range of travelers acceptance rate to guidance information provided by connected vehicle technology. A random distribution for travelers' response was then developed using the results of the study (see chapter VI).

Transportation Network Description

The study area is the east part of El Paso in Texas. There are several reasons why this area was selected for the study. First it is a congested region. Figure 7 and Figure 8 Show the traffic during the peak hour period on I-10. Secondly, it encompasses different types of roads including freeway, arterial, and local roads. This provides operational information on a complex network. Third, this area is in a non-attainment area (an area with air quality worse than the National Ambient Air Quality Standards (NAAQS)) and evaluation of the potential improvement methods in this area can be beneficial for future decision making (*93*). For a non-attainment area, there must be a plan for the future to reduce the pollutant produced to meet the NAAQS defined in the Clean Air Act Amendments. Figure 9 shows the conditions of the city due to high levels of smog in the city. Therefore, evaluating emission productions and understanding the reasons of the current conditions is helpful for future decisions.

Figure 7. Use of ITS devices for controlling traffic during the peak hour period (Photo courtesy of Texas A&M Transportation Institute)

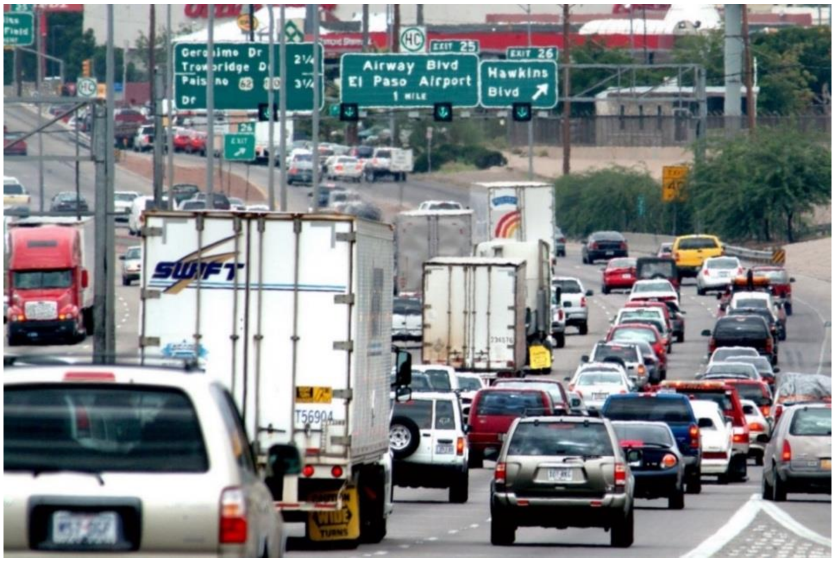


Figure 8. Traffic during the peak hour period (Photo courtesy of Texas A&M Transportation Institute)

Figure 9. An illustration of the level of smog and emissions (Photo courtesy of Texas A&M Transportation Institute)

The network was modeled in Simulation of Urban Mobility (SUMO) software for the morning peak period (6 am to 8 am). Simulation of Urban Mobility (SUMO) could simulate the traffic of the selected area and assess the impacts of strategies in mitigating congestion and also conduct a sensitivity analysis on the required parameters in the model on network performance and en-route decision effectiveness.

SUMO is used in this study for three reasons. First, it provides detailed output of operational characteristics of vehicles such as speed, acceleration, position of vehicles, etc. Secondly, it has application programming interface (API) called TraCI which enables extending the basic functionality of the tool and investigate different research objectives in a microsimulation model (*94–97*). Third, since this tool is open source, accessing all the codes enabled us to modify the operational functionalities based on the desired functional features (*94*, *96*, *98–100*). This software has been widely used in research including modeling of connected and automated vehicles (*97*, *101*, *102*).

The east part of the network of El Paso, Texas was modeled in SUMO. This network consists of the roads between and including I10 and Montana Avenue from Chelsea street to McRae Blvd. The simulated network is shown in Figure 10. The area includes 5.6 miles of interstate I10 and 4.8 miles of arterial Montana Ave. as well as the major roads and the local streets between these two roads allowing for more comprehensive impact analysis than previous efforts. I10 includes 4 lanes in each direction and Montana Avenue, known to be congested especially during the peak period, has 3 lanes in each direction. The maximum speed limit of the freeway in the selected part is 60 mph on I10 and the maximum speed limit of the arterial on this network is 45 mph on some parts of Montana Avenue and the frontage road on I10. The area modeled in the microscopic simulation model is a much larger model and covered different types of roads compared to the models in the literature. These properties of the model provides the opportunity to conduct a more comprehensive study than most studies that look at the small areas with limited types of roads.

Overall, the network includes 584 nodes and 712 edges with the total length of 135.96 km and the total lane length of 313.74 km. There are 48 traffic signals in the network, which are designed based on the real phasing of traffic signals in El Paso. The total number of vehicles entering the network is 50,872 from 6 to 8 am on a typical weekday. The map and a snapshot of the network of El Paso modeled in SUMO is demonstrated in Figure 10.

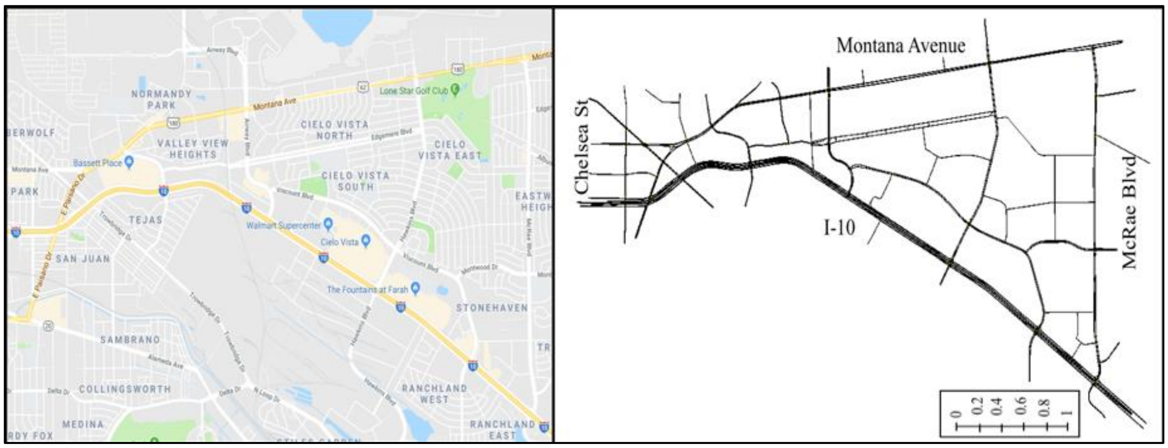


Figure 10. Map of the network with street names

The origin-destination (OD) data given by Texas A&M Transportation Institute El Paso office was used to model the traffic in the area. In order to convert the given origin-destination data to the route file, the developed dynamic traffic assignment model known as DuaIterate was used. The DuaIterate algorithm in SUMO is an iterative process to find the shortest path and try to minimize the individuals travel time based on the updated cost function. Then, the simulation model was run, and the volume on each link was then compared and adjusted based on the camera traffic counts of TxDOT. An iterative process was performed to calibrate the routes for achieving the desired accuracy (Figure 11). In Figure 11, an acceptable change was defined such that a reduction was observed in the total differences in the traffic volumes and the simulated traffic. The algorithm stopped when no improvement was observed for 5 iterations.

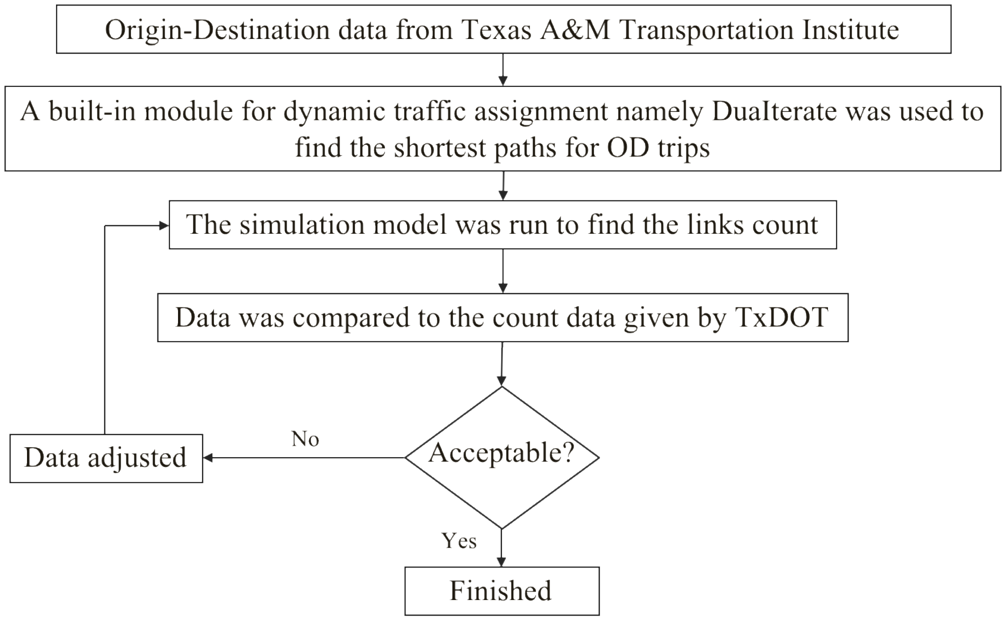


Figure 11. The flow chart of adjusting traffic data

Based on the presented flow chart (Figure 11), first DuaIterate built-in function in SUMO was used with the origin-destination data as the input to find the shortest path. Then the simulation model was running, and the links volume was generated after the simulation completed. The links volumes were compared with the links volume given by TXDOT. Then links with high difference between the real volume and simulated was selected. If the real volume is higher than the simulated, some vehicles traveling on the alternative route with high volume were forced to use this link during the path generation using DuaIterate. If the real value is lower than the simulated, then some vehicles were forced to use an alternative path in the process of generating paths with DuaIterate. A comparison of the 15-minute link volumes of the simulation model and the real traffic on the road for part of the network of El Paso for 7:00 AM to 8:00 AM is demonstrated in

Figure .

The network-wide flow-density diagram is demonstrated in Figure 13. The comparison of this network-wide flow density diagram with the literature represents the high congestion level on the network. The network-wide maximum flow rate for the high demand level of two cities of Chicago and Salt Lake City were around 300 and 350 veh/hr/ln (103). In the current study the maximum flow rate over the network was about 470 veh/hr/ln.

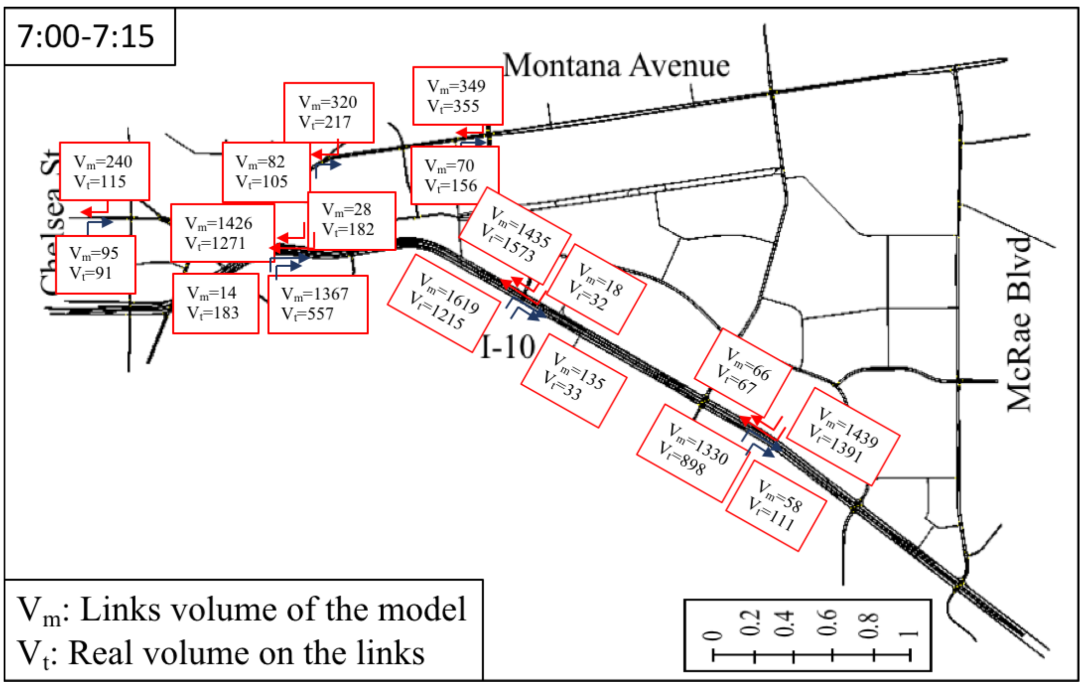
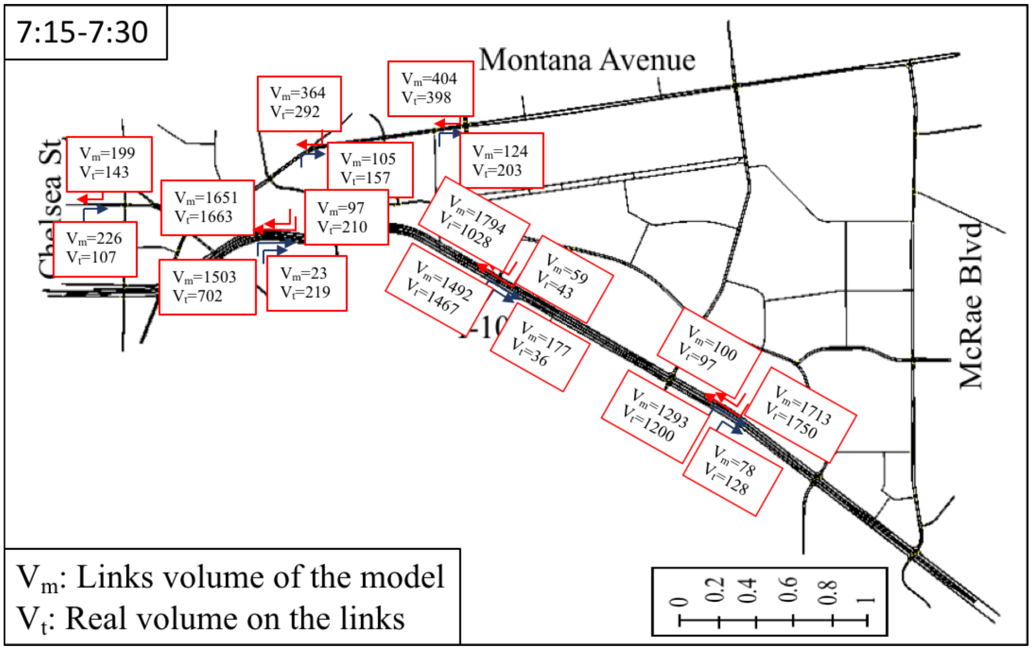


Figure 12. 15-minute links volume of the simulation model and real-traffic on part of the network from 7:00 AM to 8:00 AM



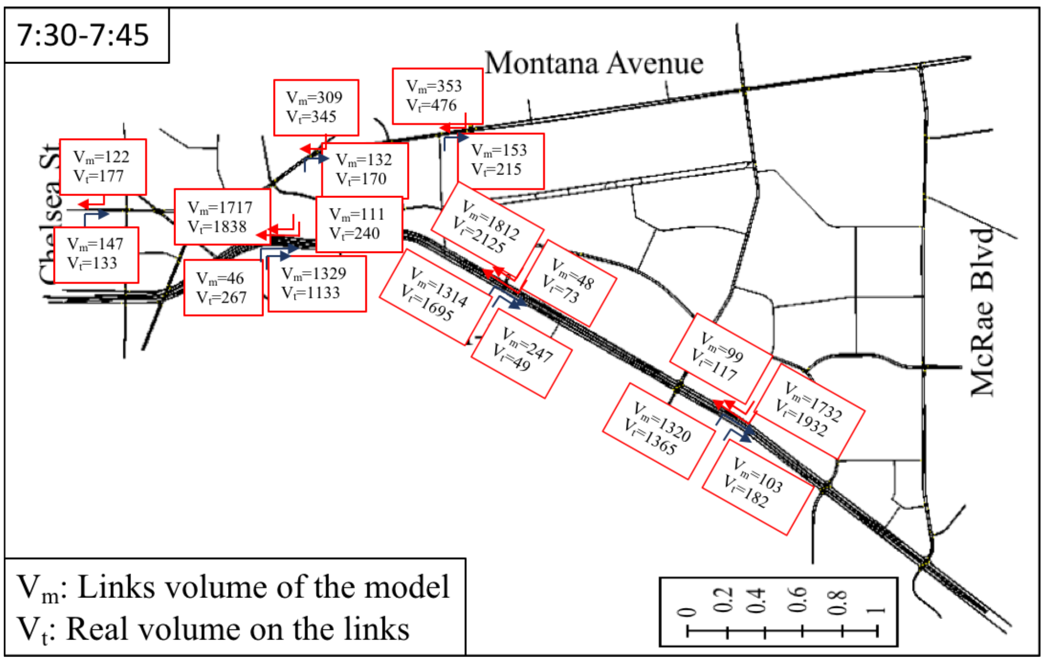


Figure 12 (continued). 15-minute links volume of the simulation model and real-traffic on part of the network from 7:00 AM to 8:00 AM

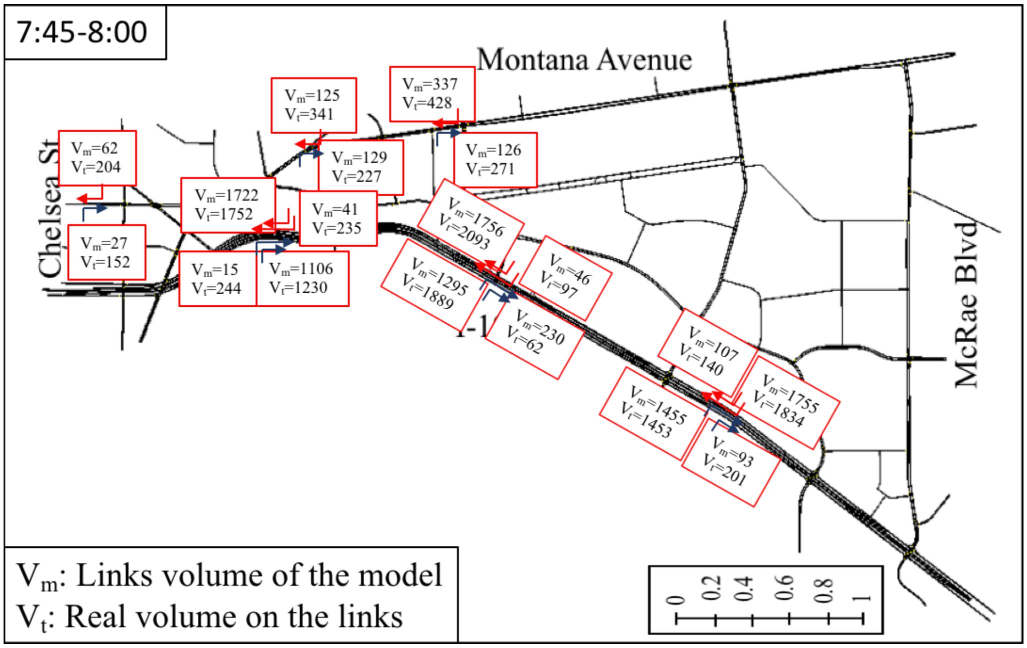


Figure 12 (continued). 15-minute links volume of the simulation model and real-traffic on part of the network from 7:00 AM to 8:00 AM

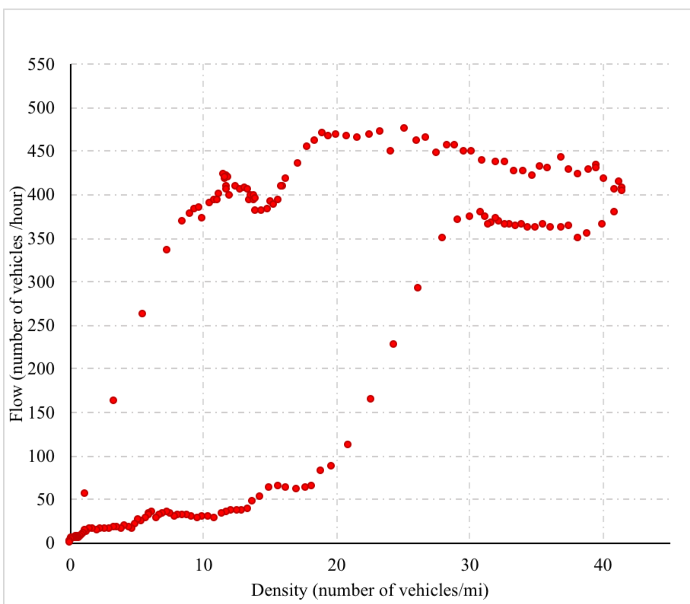


Figure 13. Flow-Density diagram for the base case scenario without any incidents

In this section, stated preference data were analyzed to characterize travelers' behavior in response to rerouting information communicated by connected vehicle technology. Then a simulation model of the east part of El Paso, Texas was developed in SUMO to evaluate the network-wide impacts of connected vehicles on traffic operation and fuel consumption. In the next section, the rerouting scenarios in the connected environment were defined and the parameters affecting the performance of the rerouting approaches on the efficiency of the transportation system were determined.

CHAPTER VI

METHODOLOGY

A microsimulation model of the network of El Paso, Texas was developed to assess the impacts of en-route decisions made by travelers using the information communicated by connected vehicle technology. Simulation of Urban Mobility (SUMO) was utilized to develop the microsimulation model of the selected network. SUMO is an open source traffic simulation package which is used along with Traffic Control Interface (TraCI) to develop the connectivity scenarios and investigating different research objectives in a microsimulation model (*16–19*). This simulation model has many functionalities that allow it to adapt to the future of the transportation system.

The developed network consists of the roads between and including I10 and Montana Avenue from Chelsea street to McRae Blvd. The simulated network is shown in Figure 10. The data used for developing this model was discussed in the previous section.

The default car-following model in SUMO, which is the modified version of the Krauss model, was used to characterize the behaviour of the regular and connected vehicles. However, it was assumed that regular vehicles do not receive and send any information. So, they follow the route that they were originally assigned to using the dynamic traffic assignment. On the other hand, it was assumed that the connected vehicles are able to send and receive information to and from other connected vehicles and also roadside devices which can be used by drivers to make en-route decisions. Congestion warning and rerouting algorithms were developed to assess the effectiveness of the deployment of connected vehicle technology. The developed congestion warning and rerouting algorithm is shown in a flowchart in Figure 14 (104, 105).

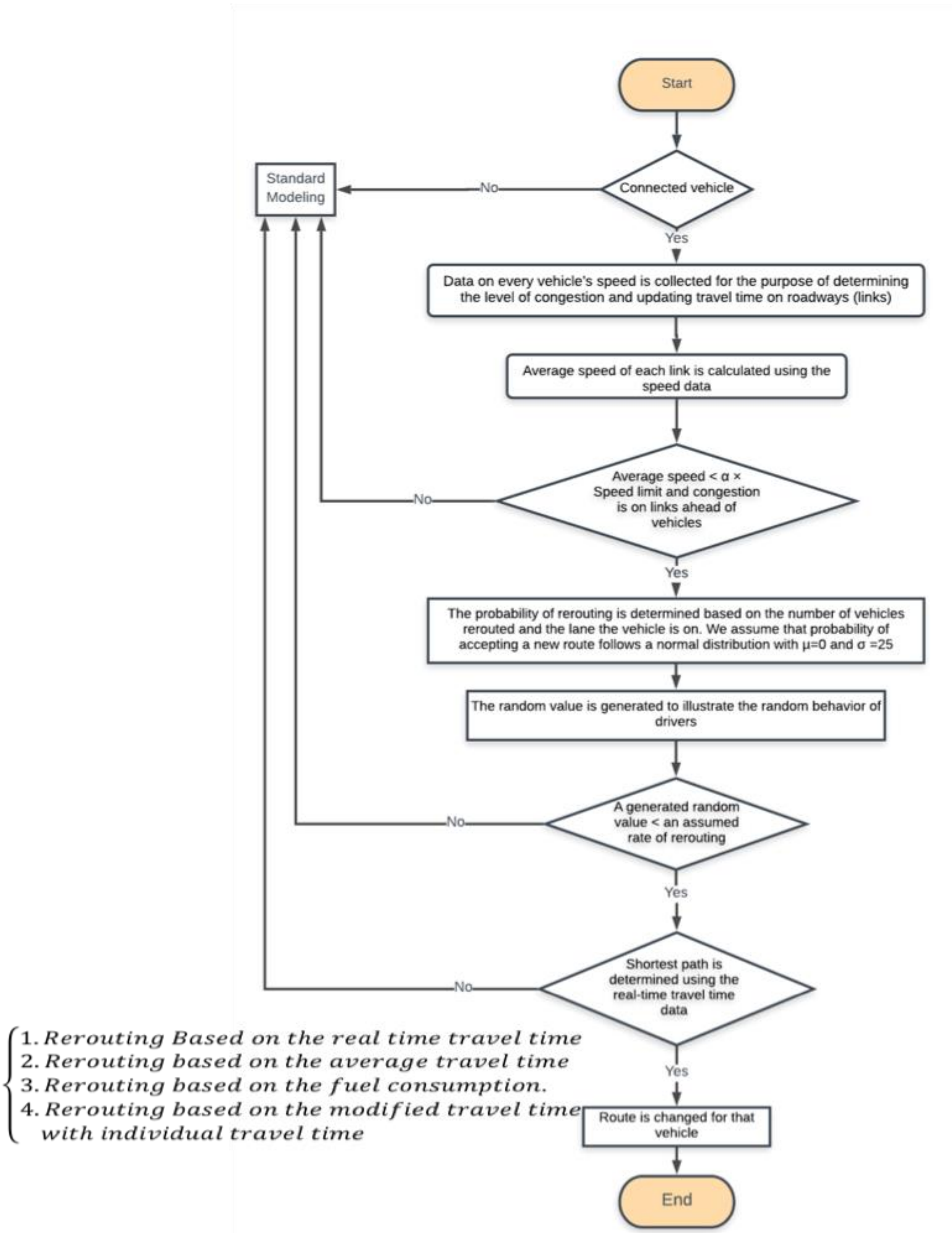


Figure 14. Flowchart of the rerouting algorithms and congestion warning

This algorithm takes advantage of the real time traffic information disseminated by connected vehicles and roadside devices. The algorithm informs CV drivers of any significant congestion on their route using warning messages. Then a new route will be suggested to the CV drivers. This new route uses the real time traffic data collected by roadside devices from connected vehicles to determine the current links traffic information and travel conditions to the vehicle's destination.

In order to define if there is significant congestion on any links on the vehicles' routes, the ratio of current average speed to the speed limit of the link was used. It was assumed that only connected vehicles provide information for roadside devices. Therefore, for calculating the average speed utilized for the congestion warning, the speed data of connected vehicles were collected during the simulation. If the average speed over the link is less than a specified percentage of the speed limit (for instance 40% which is used in most of the cases in this study), then the link would be marked as congested. The information of congestion is then transferred to the connected vehicles. After informing the drivers of the connected vehicles of the upcoming congestion, drivers will make the decision to reroute to a parallel path or not. This was defined in this study by two factors acceptance rate and rerouting rate (combination of MPR and acceptance rate).

There are three factors used in this study for determining what percentage of vehicles would reroute due to the received congestion alert and suggested path. These include the market penetration rate (MPR) of connected vehicles, the acceptance rate, and the rerouting rate. As discussed earlier, the MPR is the percentage of vehicles which received the congestion information. The acceptance rate (sometimes referred to as the compliance rate) is the percentage of the vehicles that used the new, suggested, route for bypassing congestion out of all the vehicles which received the information. The rerouting rate is the multiplication of the two factors (acceptance rate and MPR) to represent the percentage of vehicles rerouted. For example, if the MPR was 60% (percentage of connected vehicles) and the acceptance rate was 40%, the rate of rerouting would be 24%. In other words, 24% of all vehicles would use the new route.

This study examined four potential scenarios for why a vehicle may be given a suggested alternate route. These four scenarios are discussed below.

The first scenario assumed that the new route is found using real-time travel time data. For this purpose, the average travel time of connected vehicles on each link was used as the real-time travel time of the links. In the case of no connected vehicles on the links, the speed limit of the road was used to estimate the real-time travel time on the links. Then the shortest route was computed by assigning the estimated travel time to the links as the cost of traveling on the links. A list of connected vehicles aimed to reroute is provided by a comparison of a random probability assigned to the vehicle and acceptance rate assumed for the simulation. Then the route for each connected vehicle provided by the aforementioned list was updated based on the new estimated shortest path. The connected vehicles are following the new route till they received new information of congestion.

The second approach for finding the new route was to estimate the shortest path based on the average travel time on the links in the network. The difference between this method and the previous one is that, in this scenario, the average speed of connected vehicles for an assumed interval of time (compared to the last time step of the simulation which corresponds to the real-time traffic data) was considered for travel time estimation. For instance, for a link the speeds of connected vehicles for an interval (values between 30 and 150 seconds) collected (for the real-time travel time data, the speed of the last time step was used). Then the average of the collected speeds was calculated, the travel time was estimated using the average of speeds and assigned to the links as the cost of traveling the link. Different interval of time was tested to assess the impacts of them on the overall performance of the network. A similar process was then implemented to characterize the traffic for the simulation model for the upcoming time steps. To this end, first the current average speed was calculated using the speed data of the connected vehicles. This value was then compared to the speed limit of the links to prepare a list of congested links on the network. A random number between 0 and 1 is then generated and assigned to the connected vehicles. This number is compared with the predicted willingness of drivers to reroute following the rerouting advice or not. Shortest paths from the current location of connected vehicles who aimed to follow the warning advice to their destination are also estimated using the new costs (average travel time) assigned to the links. The paths of the connected vehicles willing to reroute are updated with the new information and rerouting is applied.

Looking for effective factors in the performance of the network, the third scenario of rerouting was dedicated to the fuel consumption on the links of the network. To this end, the real-time fuel consumption was used on the links as the link travel cost for finding the shortest path. In order to estimate fuel consumption, the trajectories of vehicles including speed and acceleration for each time step were recorded. Vehicle trajectories can be measured in the real world with the help of equipped vehicles or from the microsimulation model as was done in this research. Fuel consumption was calculated using the Virginia Tech Comprehensive Power-based Fuel Consumption Model (CPFM) (*81*, *106*). This method was described in the literature review section.

A 2012 Toyota Camry was used as a typical vehicle for fuel consumption model calibration. The 2012 Toyota Camry represents the family sedan and was selected because it was reported as the bestselling car (*107*) Table 8 shows the calibrated parameters for estimating fuel consumption using the Comprehensive Power-based Fuel Consumption Model (*108*).

Table 8. Vehicle parameters for CPFM calibration

Parameter	Toyota Camry-2012
Vehicle Mass(m) [kg]	1446.96 (*109*)
Drag Coefficient (C_D)	0.28 (*109*)
Frontal Area (A_f)	2.276 (*109*)
Rolling Coefficient (C_r)	1.75 (*110*)
C_1	0.0328 (*110*)
C_2	4.575 (*81*)
Drive Efficiency (η_d)	0.9 (*81*)
Number of Cylinders (N)	4 (*109*)
Engine Size [L]	2.5 (*109*)
Altitude (H) [km]	0
P_{mfo} [Pa]	400000 (*81*)
Q[J/kg]	43000000 (*81*)
Engine Idle Speed (ω_{idle}) [rpm]	675 (*81*)
Engine Displacement(d)[L]	2.494 (*109*)
Fuel Economy City (FE_{city}) [mi/gal]	25 (*109*)
Fuel Economy Highway (FE_{hwy})	35 (*109*)
ε	1×10^{-6} (*81*)

For estimating the fuel consumption for the links without any connected vehicles the average fuel consumption was assumed to be 0.04347 * length of the link which corresponds to the average 23 MPG reported by Bureau of Transportation Statistics (*111*). Following the estimation of the links cost using the fuel consumption during the last time step of the simulation, the average speed is also calculated using the speed data collected from connected vehicles technology. The comparison of the average speed on the links with the speed limit defines the need for warning to the connected vehicles having this link as their upcoming path. After determining the congested links, similar process is performed. At the end, the new route which was found based on the real-time information of links fuel consumption was assigned to the connected vehicles willing to reroute.

For the last rerouting scenario, a scenario using the travel time was developed. The modified version of travel time (link cost) was defined such if people set aside that much time for travel they are 95% confident they will arrive on time or early (equation 32).

$$Link\ Cost = \overline{TT} + 1.96\ \frac{\sigma}{\sqrt{n}} \quad (40)$$

In which $\overline{TT}$ is the average travel time, σ is the standard deviation of the travel time of vehicles on the links and n is the number of vehicles in the estimation. Moreover, with the use of standard deviation in calculating the link cost we are considering the shockwave formation (*112*) in our calculations. Following the estimation of links cost, the new shortest path for connected vehicles is calculated and suggested to these vehicles. As explained earlier, the connected vehicles then can reroute or not to the new path based on their willingness.

The rerouting algorithms were developed in python and Traffic Control Interface (TraCI) was used to run the code in SUMO. Two traffic scenarios were examined; one with typical (baseline) traffic and one with a major incident on I-10 during the simulation to make a high congestion level on part of the network. The vehicles can still use the link that the incident occurred. However, the affected lanes is closed to traffic.

The impact assessment was divided to three main parts. In the first part of the study, the impacts of rerouting using the typical cost function of real-time travel time data is used on the base simulation model without lane closures. Different market penetration rates of connectivity assuming that all of the drivers of connected vehicles follow the suggested route are examined. In the second part of study, a major incident causing lane closures is simulated and the four discussed scenarios of rerouting were investigated. The impacts of willingness to rerouting and also market penetration of connected vehicles are applied together in this section. Finally, different scenarios of incidents with various duration and number of closed lanes were studied. A sensitivity analysis on the update interval of information is also performed to understand the impacts of frequent traffic information communication on the overall performance of the network. In order to take into account the random behaviour of travelers in response to congestion and warning messages, a probability distribution was used. We assumed travelers' willingness to change route followed a Normal distribution. To this end, the equation (3) was used as the to find the probability of a CV traveller adjusting their route.

$$\text{Probability of route change} = \alpha \times \frac{\exp\left(\frac{-n^2}{2\times\sigma^2}\right)}{2\sqrt{\pi}\times\sigma} \qquad (41)$$

Where α is a constant to adjust the probability function, n is the number of rerouted vehicles due to congestion on each link, and σ is the standard deviation of the Normal distribution. Different values of α were considered to simulate the behaviour of travelers on each lane of the roads. These values are defined such that the vehicles on the inner lane (further away from the exit ramp) have a lower probability of choosing the suggested new route. Another assumption in this equation is that as the number of vehicles rerouted to the alternative routes increases, the probability that the next vehicle accepts the alternative route decreases.

As discussed in the section of input data (chapter V), an internet-based questionnaire was prepared. Almost 1900 completed responses were collected to investigate the travelers' reaction to the rerouting information. Based on the results of the study, 87-92% of travelers would accept the rerouting suggestion if they see the congestion ahead of them and 75-79% would accept in case the congestion is not visible. On the other hand, 8-13% of travelers would not accept the rerouting advice with visible congestion and 21-25% would not accept rerouting suggestion without visible congestion. These probabilities were used to develop the distribution representing the random behaviour of drivers of connected vehicles. To develop the distribution using these probabilities, the parameters are defined such that the travelers following the advices on all the lanes, the percentage of rerouting is about the sum of the probabilities of definitely take the new route and half of the probabilities of travelers who are not sure if take the new route or not.

After calculating the rate for willingness of drivers to reroute, a random value will be assigned to the traveler and compared with the willingness to reroute to determine whether the vehicle will change its route or not. Figure 15 demonstrates the probability distribution of travelers' willingness to reroute. Figure 16 shows a sample of the lane numbering used for freeways in Figure 15.

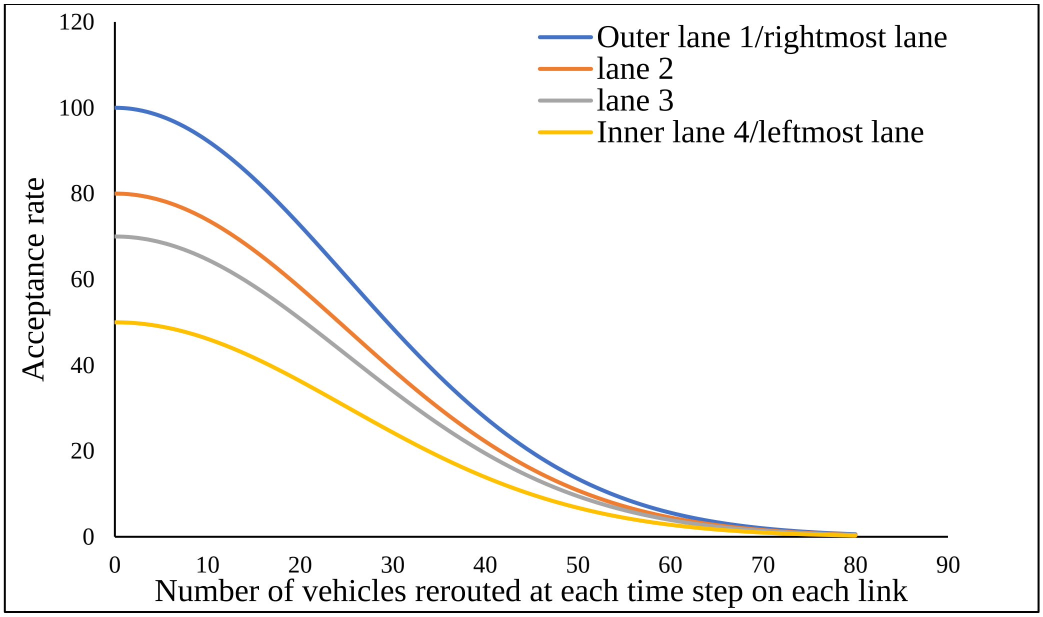


Figure 15. The probability distribution of tavelers choosing the suggested route

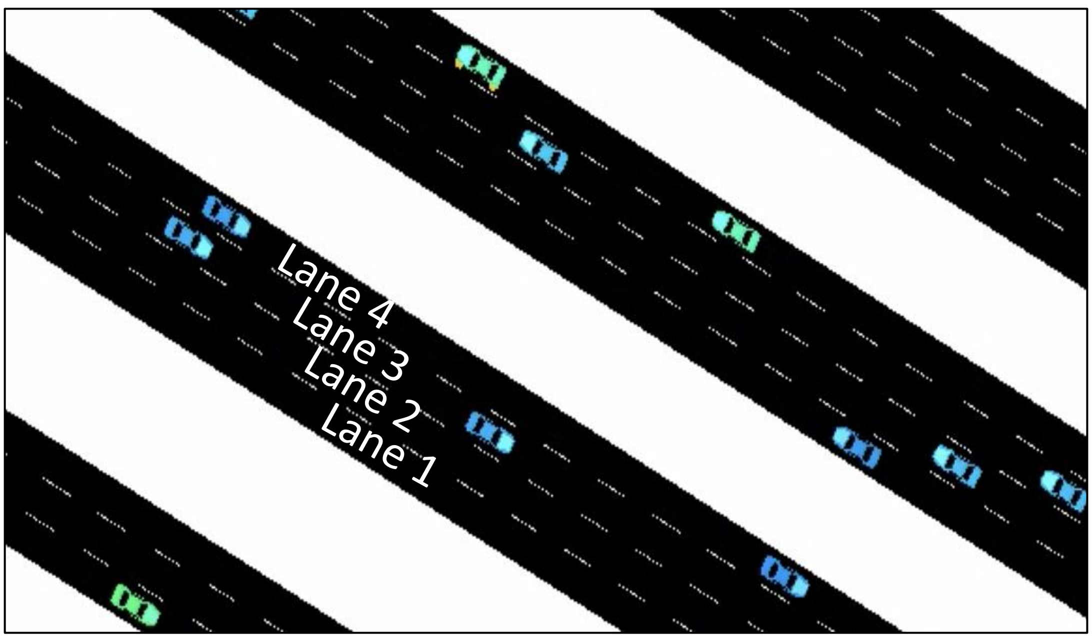


Figure 16. Lane labeling for the rerouting acceptance

As presented in Figure 15, the same value was considered for various update intervals. It should be noted that acceptance rate is the percentage of connected vehicles accept the congestion alert and reroute to the suggested path for traveling. It was assumed that it does not matter how frequent the information is provided (every 150, 300 or 600 seconds). The driver's willingness to accept the new suggested route or not would not change by the frequency of the information. This provides an increased impact of the update interval to more easily see the impact of update interval on the performance of the system.

In the next section, the rerouting models described here were evaluated. First the four rerouting strategies were modeled and tested. A sensitivity analysis was also performed on the congestion threshold, update interval of rerouting, and interval for estimating the average travel time for the second approach of rerouting.

Next, the best rerouting approach was used to analyze the model during various incident scenarios (duration and number of lane closures). In this part, a similar sensitivity analysis on the update interval was also performed. In this part, instead of using a fixed value for rerouting rates, the Normal distribution developed in this section was used to mimic random behavior of travelers. Finally, the best approach of rerouting was used to visualize the fuel-flow, flow-density and flow-speed diagrams. These diagrams demonstrated the efficiency of the model at the macroscopic model.

CHAPTER VII

ANALYSIS AND RESULTS*

To conduct a comprehensive evaluation on the implemented rerouting algorithms, factors involving the rerouting should be concisely defined. The important factors in the study include: the market penetration rate of connectivity, the willingness of drivers to reroute, the update interval of information, the interval of time for the average travel time rerouting algorithm, the factors associated with incidents including the duration and number of lanes closed. This section is divided into three parts. In the first part, the four scenarios of rerouting were analyzed and sensitivity analysis was performed on some of the parameters in the simulation model. These parameters include the rerouting rates (combination of MPR and acceptance rate), the update interval of travel time and the interval for estimating the average travel time. As the acceptance rate and market penetration rate of connectivity both determine the rate of accepting a suggested route, in this study, these two parameters were considered together as one parameter called rerouting rate. The best rerouting approach was selected based on the results of this section. This rerouting algorithm was then used for the rest of the study.

In the second part, the focus was on the incident scenarios in the mixed environment of connected and non-connected vehicles. To this end, the analysis was performed on some parameters including incident duration, number of lanes closed and the update interval of travel time. Unlike the first section, in which static acceptance rates were considered for travelers response to the information, in this section, the Normal distribution explained in the previous section was used to take into account the random behaviour of travelers which is affected by other

*Reprinted with permission from "The impacts of connected vehicle technology on network-wide traffic operation and fuel consumption under various incident scenarios" by Samimi Abianeh, A., Burris, M., Talebpour, A., & Sinha, K., 2020. Transportation Planning and Technology, 43(3), 293-312, Copyright 2020 by Routledge.

travelers on the network (Figure 15). This would help in having variable rerouting rates during the simulation (in previous part, the model was analyzed using static rerouting rates).

Finally, in the last section, the goal was to see how the rerouting in the connected environment affects the network performance at the macroscopic level. Similarly, the best rerouting algorithm from the first part was used here. The static rerouting rates were also considered to simulate the travelers responses in the mixed traffic of connected and non-connected vehicles. To this end, the Flow-Density relationship, Fuel-Flow and Flow-Speed were presented. The three parts of analysis explained here, are summarized in Table 9. It should be noted that in all the parts of analysis except section 2 (evaluating incident scenarios) acceptance rate and MPR was considered as one factor called rerouting rate. In section 2, they were defined separately and acceptance rate follows Figure 15.

Table 9. Summary of the analysis

Part	Aspect evaluation		Rerouting strategies Examined	Update interval (sec)	Congestion Threshold	Incident Duration (sec)	Number of lanes closed	Traveler Rerouitng Rate
1	Evaluating rerouting strategies	a	Rerouting based on the real-time travel time	30, 60, 90, 120, 150	40%	0	-	20%
		b	Rerouting based on the real-time travel time	150	20%, 40%, 60%	0	-	20%
		c	1.Rerouting based on the real-time travel time 2. Rerouting based on the average travel time 3. Rerouting based on the Fuel consumption 4.Rerouting based on the modified version of travel time (upper limit of confidence interval of mean travel time)	150	40%	900	1	0,10%,20%,40%,60%,80%
		d	Best found in 1-c	best found in 1-a	40%	900	1	best found in 1-c

Table 9 (continued). Summary of the analysis

Part	Aspect evaluation		Rerouting strategies Examined	Update interval (sec)	Congestion Threshold	Incident Duration (sec)	Number of lanes closed	Traveler Rerouitng Rate
		e	Best found in 1-c	best found in 1-d	best found in 1-b	900	1	best found in 1-c
2	Evaluating incident scenarios	a	Best found in 1-c	150, 300, 600	40%	0, 600, 900, 1500, 2400, 2700	1, 2	MPR(0%,20%,40%,60%,80%, 100%) Acceptance rate of figure 15
3	Network performance at the macroscopic level	a	Best found in 1-c	150	40%	900	1	0,10%,20%,40%, 60%,80%

Rerouting Strategies

As described earlier, four rerouting algorithms were simulated in this study. Several scenarios of rerouting rates (static) were considered for each rerouting algorithm. For the base case (Table 9, 1-a and 1-b), the model was simulated without any incident to analyze the impacts of update interval (Table 9, 1-a) and congestion threshold (Table 9, 1-b) on the network performance. Then the rerouting algorithms were implemented with a simplified assumption on the update interval and congestion threshold. It should also be noted that the rerouting rate which was defined in chapter VI as the combination of MPR and acceptance rate was used in this section.

Base Case Scenario

Before proceeding to the evaluation of the rerouting strategies, a selection must be made for two common factors among all of the rerouting scenarios. Table 10 demonstrated the network performance without any connected vehicle technology. The first factor in the model is the update interval. For the update interval several values including 30, 60, 90 120, and 150 seconds were considered for the sensitivity analysis (Table 9, 1-a). Another factor which is in common for all the rerouting scenarios is the congestion threshold (Table 9, 1-b). The congestion threshold is the level at which the links are marked as congested and warning messages are sent out to the connected vehicles. The congestion threshold was studied based on comparing the current average speed to the speed limit. For the purpose of this study, average speeds that were 20%, 40% and 60% of the speed limit were analyzed for the congestion threshold (Table 9, 1-b). The first rerouting algorithm, rerouting based on the real-time travel time data, with the rerouting rate of 20%, without simulating any incident was considered for the sensitivity analysis of these factors on the performance of the model.

Table 10. Network performance for 0% connectivity, the traffic without any connected vehicle technology

Base scenario	Fuel consumption	VMT	TT
0% connectivity	21650	232683	15957

Table 11. Networkwide performance for various update intervals_ Congestion threshold of 40% (average speed ≤ 40% of speed limit) (Table 9, 1-a)

Update interval	Fuel consumption	VMT	TT
30	24352	232419	17955
60	19023	239317	14000
90	18547	238272	13650
120	19420	237690	14299
150	19721	237570	14521

Table 12. Networkwide performance for various congestion level- update interval of 150 seconds (Table 9, 1-b)

Congestion Level	Fuel consumption	VMT	TT
20%	19357	237020	14254
40%	19721	237570	14521
60%	19279	237314	14193

The results of this analysis demonstrated that, for the rerouting rate of 20%, the update interval of 90 seconds had the highest benefit in terms of fuel consumption and travel time (Table 11) . Moreover, the average speed of less that 60% of the speed limit resulted in the lowest travel time and fuel consumption (Table 12). However, because for the rest of the scenarios the simulation need to model higher rerouting rates and incidents, for computational needs, 150 seconds was assumed for the update interval and a threshold of 40% was assumed for congestion. At the end, the best approach was tested for the best values found in this section to assure that this assumption had not impacted the overall findings. In the next section, each rerouting scenarios with a simulated incident of 900 seconds on the I-10 freeway (Figure 17), update interval of 150 seconds and congestion threshold of 40% were evaluated (unless otherwise specified) .

Four Rerouting Scenarios

As mentioned earlier, in this section, the roadside communication devices disseminated congestion warning messages when meeting the congestion threshold of 40% and the information were distributed every intervals of 150 seconds. The vehicles equipped with communication devices (connected vehicles) receive these messages and can decide to change their route to alternative routes to bypass congestion or not. In order to model the congestion during the peak hour period, for all of the rerouting scenarios, one incident on the I-10 freeway (westbound lanes-peak direction) was simulated. The location of the incident is shown in Figure 17. The duration of the incident varied based on the scenarios but was 900 seconds in this scenario (Table 9, 1-c). After this period of time, the crashed vehicle was removed and the lane opens to traffic.

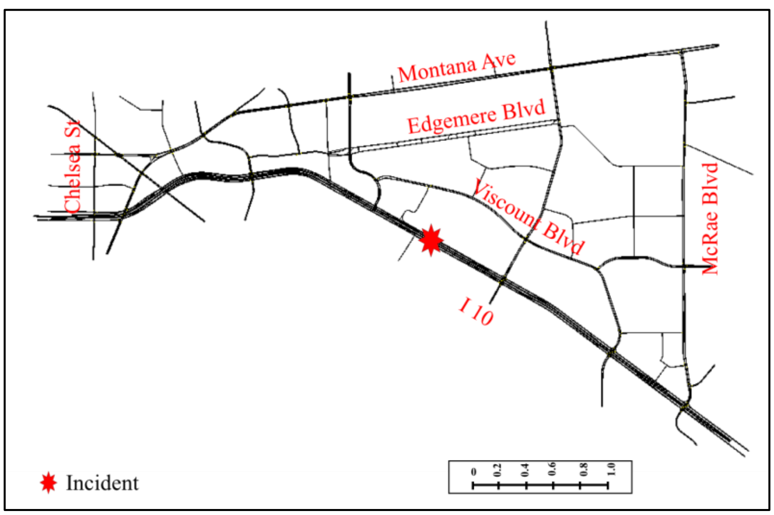


Figure 17. The location of the incident in the simulation model (West direction)

Six rerouting rate scenarios including 0%, 10%, %20, %40, %60 and %80 were assumed (Table 9, 1-c). It should also be noted that one rerouting strategy was used for each of the simulation and the algorithms were not used concurrently for each simulation in the current study.

Rerouting based on the real-time travel time data

As described earlier, the first approach for rerouting was to use the real-time travel time data collected by the roadside devices from the equipped vehicles on the road. To this end, the algorithm was developed in Python for rerouting the connected vehicles and TraCI was used for applying the implemented algorithm in the simulation model. Here, this rerouting algorithm was tested with the assumption of 40% congestion threshold, update interval of 150 seconds, a 900 seconds incident and rerouting rates varying from 0% to 80% (Table 9, 1-c).

Rerouting based on the average travel time data

For this rerouting algorithm, the average travel time was used to reroute the vehicles when their assigned route was congested. Before final analysis of this approach, a sensitivity analysis was performed to find an appropriate value for the interval over which the average should be estimated. For this analysis, the rerouting rates of 20% and 40% with 5 values of interval including 30, 60, 90, 120, 150 seconds were evaluated.

Rerouting based on fuel consumption

The third approach involved calculating the real-time fuel consumption over the links using the data collected by connected vehicles. To this end, VT-CPFM model was used for estimating the fuel consumption. This estimated fuel consumption was considered as the costs on the links for finding the least cost path from the last vehicle position to the destination. Similar assumptions including the congestion threshold of 40%, the update interval of 150 seconds and the rerouting rates between 0% and 80% were simulated.

Rerouting based on the upper-bound confidence interval of travel time

The last approach developed in this study also considered the travel time for rerouting. The only modification that is applied to estimating the cost involved using the standard deviation of the travel time on the network links (equation 32). This will result to having higher costs when the individual vehicles' speed varies significantly on a link.

Similar parameters as the other approaches were then used to explore the impacts of these parameters on the overall performance of the network. The assumed parameters are the congestion threshold of 40%, the update interval of 150 seconds, a 900-second incident and rerouting rate which varies from 0% to 80% (Table 9, 1-c).

First, a sensitivity analysis was performed on the interval for the rerouting based on the average travel time. The intervals including 30, 60, 90, 120, and 150 seconds were analyzed. The results of the analysis for 20% and 40% of market penetration is shown in Table 13 and Table 14. Based on the results of the sensitivity analysis, 60 seconds with lowest network-wide fuel consumption and travel time was assumed for the rest of the analysis.

Table 13. The impact of length of time over which the average travel time was calculated (20% MPR) (Table 9-1-c (2))

Interval for Average	Fuel consumption	VMT	TT
30	20587	237569	15165
60	18267	240262	13444
90	18956	237247	13957
120	19918	236883	14669
150	20107	236738	14810

Table 14. The impact of length of time over which the average travel time was calculated (40% MPR) (Table 9-1-c (2))

Interval for Average	Fuel consumption	VMT	TT
30	18174	240154	13378
60	17170	242335	12632
90	17629	239147	12974
120	18029	239039	13269
150	18402	238811	13547

Then the four rerouting strategies were developed in Python and TraCI was used to communicate between the simulation model and the algorithms. The networkwide travel time, fuel consumption and vehicle miles traveled for each scenario were estimated and presented in Table 15.

Table 15. Comparison of four scenarios of rerouting for various rerouting rates (Table 9, 1-c)

Model	Rerouting rates	Congestion Level	Fuel consumption (L)	Vehicle Miles Traveled (mile)	Total Travel Time (hours)
Travel Time	0	0.4	23636	232643	17426
Travel Time	0.1	0.4	18457	235671	13588
Travel Time	0.2	0.4	16998	244439	12503
Travel Time	0.4	0.4	17207	243329	12659
Travel Time	0.6	0.4	17239	242732	12684
Travel Time	0.8	0.4	18434	242947	13570
Average Travel Time	0	0.4	23636	232643	17426
Average Travel Time	0.1	0.4	18648	239230	13728
Average Travel Time	0.2	0.4	18267	240262	13444
Average Travel Time	0.4	0.4	17170	242335	12632
Average Travel Time	0.6	0.4	17541	241924	12908
Average Travel Time	0.8	0.4	17924	241165	13192
Fuel Consumption	0	0.4	23636	232643	17426
Fuel Consumption	0.1	0.4	19953	235118	14695
Fuel Consumption	0.2	0.4	19302	237710	14213
Fuel Consumption	0.4	0.4	18252	240651	13436
Fuel Consumption	0.6	0.4	18851	237341	13880
Fuel Consumption	0.8	0.4	19444	236157	14315

Table 15 (continued). Comparison of four scenarios of rerouting for various rerouting rates (Table 9, 1-c)

Model	Rerouting rates	Congestion Level	Fuel consumption (L)	Vehicle Miles Traveled (mile)	Total Travel Time (hours)
Modified Travel Time	0	0.4	23636	232643	17426
Modified Travel Time	0.1	0.4	18702	239972	13768
Modified Travel Time	0.2	0.4	18437	239286	13572
Modified Travel Time	0.4	0.4	17787	240738	13090
Modified Travel Time	0.6	0.4	17496	242534	12874
Modified Travel Time	0.8	0.4	18740	242481	13797

Table 15 presented the total travel time, vehicle miles traveled and fuel consumption at the network level for four scenarios of rerouting. For the average travel time and the real-time travel time methods, the highest benefits were achieved at the rerouting rate of 20% and 40% accordingly. For fuel consumption and the modified travel time, the least network-wide travel time and fuel consumption occurred at the rerouting rate of 40% and 60%, correspondingly. With an increase in the rate of rerouting, an increase in the total travel time and fuel consumption was observed. For vehicle-miles-traveled, an increasing trend with the increase in the rerouting rate was observed. This occurred because with high congestion on the roads, the vehicles with communication technology received warning to change their path and most of the time, the suggested alternative path was not the path with the shortest distance to the destination and with the congestion on the main path, these alternative becomes more efficient in terms of travel time and fuel consumption. Therefore, the increase in the rerouting increased the vehicle-miles-traveled. The comparison among these methods presented that the first two methods, real-time travel time and the average travel time performed better than the second two methods. One difference between the two set of approaches, is that the second set considered the variance in the individuals travel time; however, the first two focused on the average travel time of all vehicles over a period of time. Based on the results, for the implemented simulation model, accounting for the variance of travel times (individual travel time) is not beneficial in the rerouting scenarios.

The rerouting strategy with the real-time travel time resulted in the lowest travel time and fuel consumption at the rerouting rate of 20%. Here the model was tested again using the best values found for the update interval and congestion threshold in Table 11 and Table 12. The results

(Table 16 and Table 17) demonstrated that the update interval of 150 seconds and congestion level of 40% resulted in better efficiency of the network.

Table 16. Networkwide travel time and fuel consumption for real-time travel time rerouting algorithm at two update intervals

Rerouting rates	Update interval	Fuel consumption (L)	Vehicle Miles Traveled (mile)	Total Travel Time (hours)
0.2	90	17233	224771	12683
0.2	150	16998	244439	12503

Table 17. Networkwide travel time and fuel consumption for real-time travel time rerouting algorithm at two congestion thresholds

Rerouting rates	Congestion threshold	Fuel consumption (L)	Vehicle Miles Traveled (mile)	Total Travel Time (hours)
0.2	0.60	18997	237839	13985
0.2	0.40	16998	244439	12503

In the next section the impacts of the incidents (duration and number of lanes closed) and the update interval of travel time on the performance of the network was evaluated. However, instead of using the rerouting rates as a factor combining the MPR of connectivity and travelers' acceptance rate, these factors were used separately. For the market penetration rate, 0%, 20%, 40%, 60%, 80% and 100% were assumed. For the acceptance rate, the Normal distribution demonstrated in Figure 15 were used. The first rerouting algorithm which resulted in lowest travel time and fuel consumption was considered in this section as well.

Incident Scenarios in a Connected Environment

In the previous section, four methods of rerouting algorithms were evaluated. Here, a sensitivity analysis on some parameters associated with incidents is performed (Table 9, 2-a).

These include the incident duration, number of lanes closed during the incident, and the update interval at various incident scenarios. It should also be noted that the acceptance rate and MPR were considered separately in this section. The acceptance rate followed the distribution presented in Figure 15 and acceptance rate was varied between 0% and 100%.

Incident Scenarios

To investigate the impacts of nonrecurring congestion on the network performance several incident scenarios were defined and evaluated (Table 9, 2-a). First, for the base case scenario with no CVs, several incidents of various duration and number of lanes closed (one lane or two lanes) on the I-10 freeway were modeled. Incident durations is a critical parameter in evaluating the congestion caused by incidents. In addition, the number of lanes affected is another crucial parameter to study the performance of the network under the congestion induced by incidents. The same scenarios were then modeled with the addition of CVs at various market penetration rates (MPR), including 20%, 40%, 60%, 80%, and 100%.

As discussed in the chapter VI, to model the random behaviour of travelers in response to congestion and warning messages, a probability distribution (Normal distribution) was used (Figure 15). These values were defined such that the vehicles on the inner lane (further away from the exit ramp) have a lower probability of choosing the suggested new route. Another assumption in this equation is that as the number of vehicles rerouted to the alternative routes increases, the probability that the next vehicle accepts the alternative route decreases.

Case 1 (base case): 0% MPR of CVs

In this case, incident scenarios with various durations and number of lanes closed were developed in SUMO. The duration of incidents included 600, 900, 1500, 1800, 2400, and 2700 seconds (Table 9, 2). Two scenarios were examined for each duration: one lane closed, and two lanes closed. The incidents were on the westbound lanes of I10 where traffic is moving towards the center of El Paso. The location of the incident is shown in Figure 17. It was assumed that there is no connectivity among vehicles. Therefore, the vehicles don't receive or send any information. The objective of this set of scenarios was to assess the networkwide impacts of incidents on overall travel time and fuel consumption.

As demonstrated in Figure 18 and Figure 19, the networkwide fuel consumption and networkwide travel time follow similar behaviour in response to the changes in the duration of incidents and number of lane closures. Increase in the duration of incidents and number of lane closure result in more fuel consumption and higher travel time.

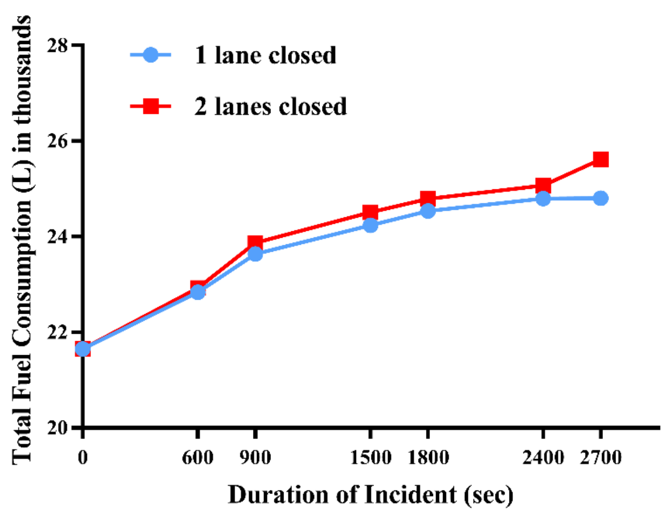


Figure 18. Total fuel consumption over the network vs the duration of incidents

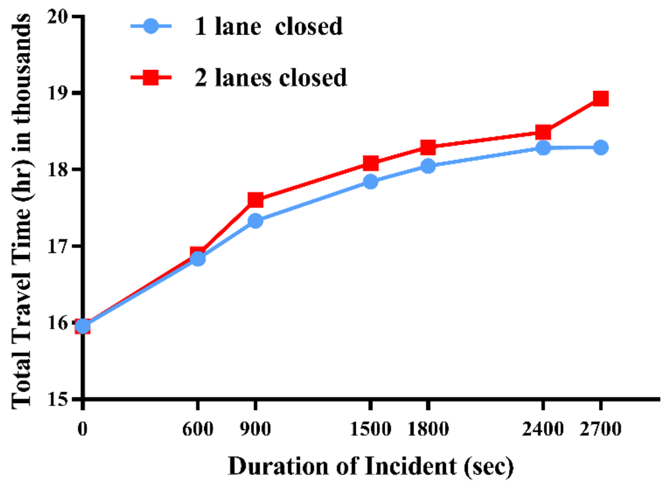


Figure 19. Total travel time over the network vs the duration of incidents

Case 2: Adding connectivity technology to the vehicles

In this section, three of incident duration scenarios (900, 1500, and 2400 seconds) with various market penetration rates of connectivity were investigated. Including connectivity, enabled the vehicles to send and receive information on congestion from other vehicles on the road and make more informed decisions on route choice. The vehicles with communication technologies follow the algorithm shown in Figure 11 for rerouting. The objective of this case is to investigate the impacts of connectivity on networkwide performance based on energy consumption, travel time and vehicle miles traveled (VMT).

As demonstrated in Figure 20 and Figure 21, the increase in the market penetration rate improves the performance of the network from both the travel time and energy consumption perspectives. However, the increase in the VMT represented in Figure 22 suggests that the rerouting scenarios do not necessarily select the routes with lower distance. However, vehicles were rerouted to the paths with lower travel time and lower fuel consumption. Figure 20 and Figure

21 also demonstrate the significant dependency of the fuel consumption and travel time on update interval of information exchange among vehicles. The results also indicate that rerouting to the shortest route (lowest travel time) reduces the fuel consumed on the network.

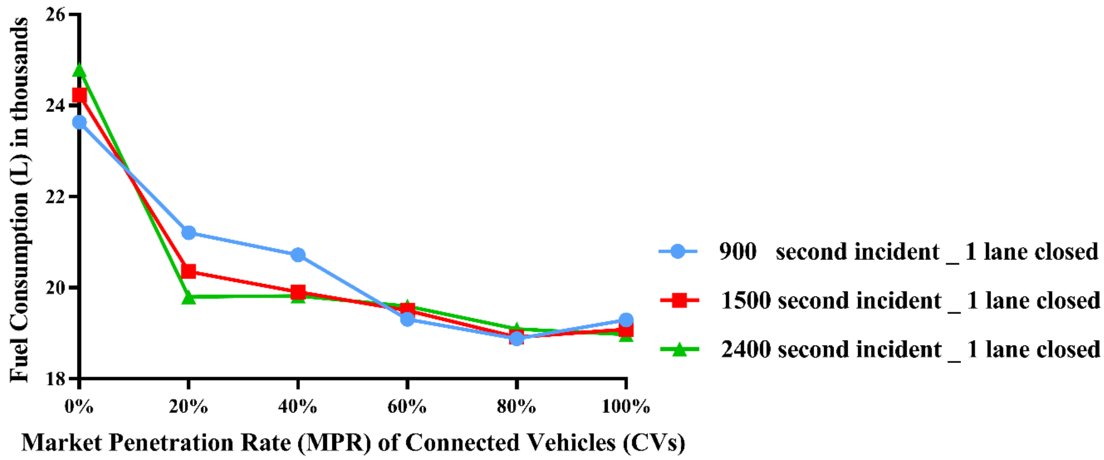


Figure 20. Total fuel consumption for different market penetration rate (MPR) of Connected Vehicles

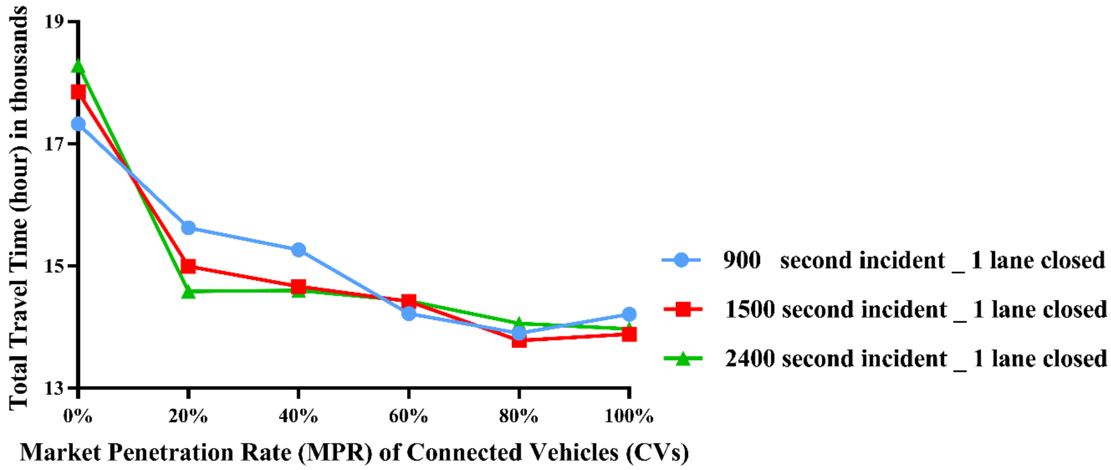


Figure 21. Total travel time for different market penetration rate (MPR) of Connected Vehicles

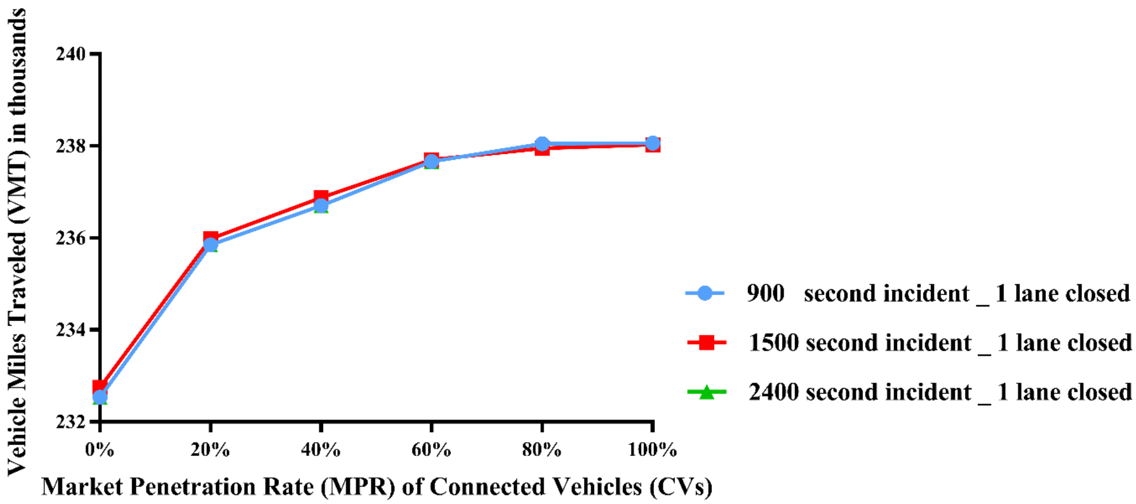


Figure 22. Total vehicle miles traveled for different market penetration rate (MPR) of Connected Vehicles

Table 18 and Table 19 include the change in the network performance measures in proportion to the base case (without any communication technology) as MPR increases. For an incident with one lane closed for 900 seconds, the highest increase in the vehicle miles traveled due to the connected vehicle technology is 2.28%, the highest reduction in fuel consumption is 20.10% and the highest reduction in travel time is 19.76% (see Table 18). For an incident with one lane closed lasting for 1500 seconds, the highest increase in VMT is 2.27%, the highest reduction in fuel consumption is 21.96% and the highest reduction in travel time is 22.76% (see Table 19). For an incident with one lane closed that lasts for 2400 seconds, the highest increase in VMT is 2.38%, the highest reduction in fuel consumption is 23.46% and the highest reduction in travel time is 23.58% (see Table 20). Although the rerouting scenario results in longer routes, it improves network performance by lowering the fuel consumption and travel time. Moreover, connected vehicle technology has greater benefits as congestion worsens. This result was examined by modelling incidents with longer durations. The results demonstrate the larger improvements in travel time and fuel consumption for the network with longer incidents and more congestion (see Table 18, Table 19 and Table 20).

Table 18. Impacts of connected vehicles technology on the network with a 900 second incident

Market penetration rate	Fuel consumption	Vehicle-miles traveled	Total Travel time
20%	-10%	1%	-10%
40%	-12%	2%	-12%
60%	-18%	2%	-18%
80%	-20%	2%	-20%
100%	-18%	2%	-18%

Table 19. Impacts of connected vehicles technology on the network with a 1500 second incident

Market penetration rate	Fuel consumption	Vehicle-miles traveled	Total Travel time
20%	-16%	1%	-16%
40%	-18%	2%	-18%
60%	-20%	2%	-19%
80%	-22%	2%	-23%
100%	-21%	2%	-22%

Table 20. Impacts of connected vehicles technology on the network with a 2400 second incident

Market penetration rate	Fuel consumption	Vehicle-miles traveled	Total Travel time
20%	-20%	1%	-20%
40%	-20%	2%	-20%
60%	-21%	2%	-21%
80%	-23%	2%	-23%

Table 20. (continued). Impacts of connected vehicles technology on the network with a 2400 second incident

Market penetration rate	Fuel consumption	Vehicle-miles traveled	Total Travel time
100%	-23%	2%	-24%

Update Interval of Rerouting

The relationship between the parameters in the rerouting algorithms and their effectiveness is unknown. Therefore, it is important to investigate the impacts of these parameters on network performance. As discussed earlier in this section of incident analysis, the CV update interval was assumed to be 600 seconds. Thus, vehicles only receive information every 10 minutes and travelers would only react to that information at ten-minute intervals. In this section, the impacts of this parameter on the operation of the network is evaluated. It should also be noted that Figure 15 was used for all of the update intervals. The same probability distribution was used for all of the update intervals for two reasons. First, it provides a more wider and general impacts of the update interval on the network performance. Second, this would keep the acceptance rate the same and a sensitivity analysis would be performed on the impacts of frequency of the communicating information on the overall performance of the network. The total vehicle-miles traveled, fuel consumption and travel time for different values of the update interval including 150, 300, and 600 seconds are shown in Figure 23, Figure 24 and Figure 25.

For an update interval of 150 seconds, the total fuel consumption and total travel time are reduced. The total vehicle-miles traveled increases as the frequency of communication increases.

Since the fuel consumption and travel time are the more crucial network performance measures, the more frequent the CVs are updated, the higher the efficiency of the network. Also, the results of the study show that an increase in VMT does not necessarily increase the travel time and fuel consumption. For the interval update of 600 seconds, there is a slight increase in the travel time and fuel consumption for MPR of 100% compared to 80% as represented in Figure 27 and Figure 28. This increase is mainly due to the rerouting of so many vehicles to alternative routes that the alternate routes become slower. Additionally, since vehicles are rerouted to signalized arterials, some of the rerouted vehicles may encounter more signal delay than expected. This issue was resolved by decreasing the update interval and providing the connected vehicles more frequent traffic data.

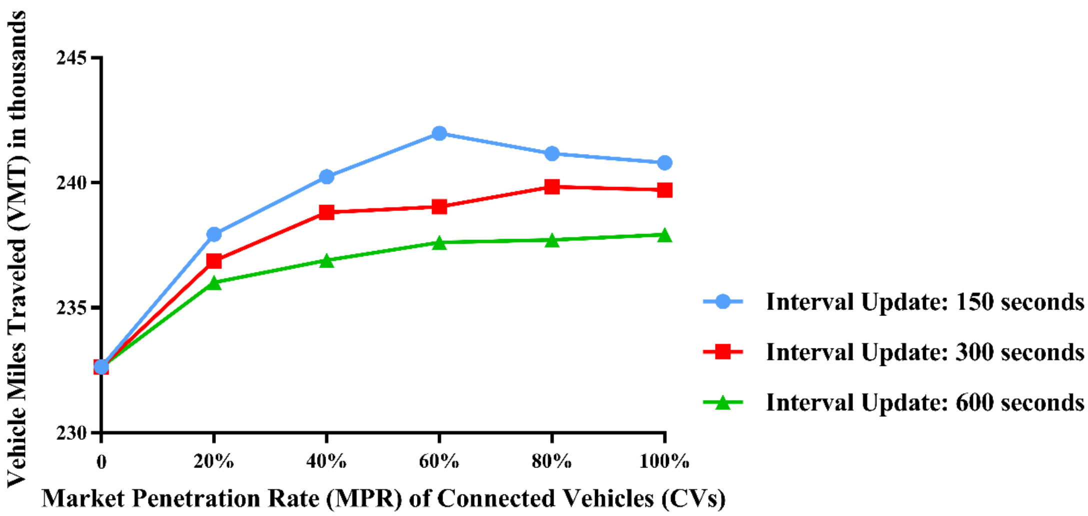


Figure 23. Total vehicle miles traveled for different market penetration rates (MPR) of connected vehicles and different update intervals

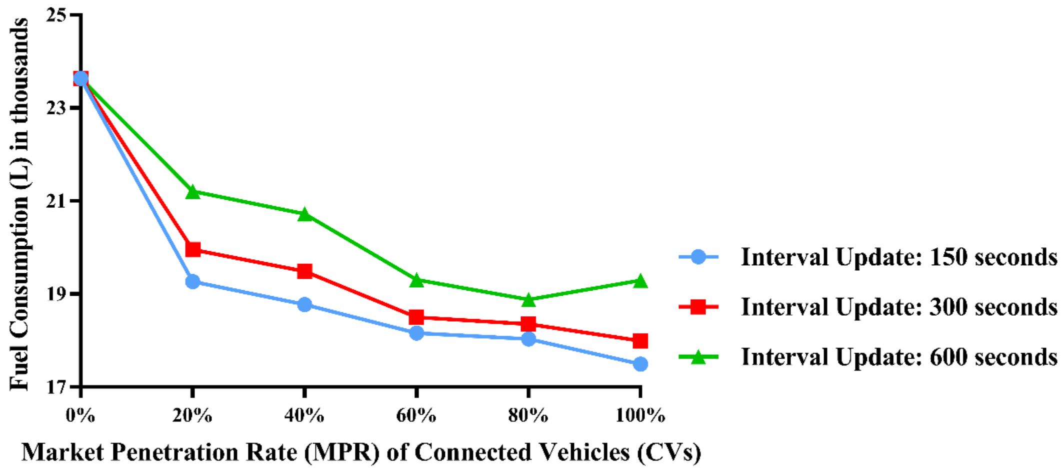


Figure 24. Total fuel consumption for different market penetration rates (MPR) of connected vehicles and different update intervals

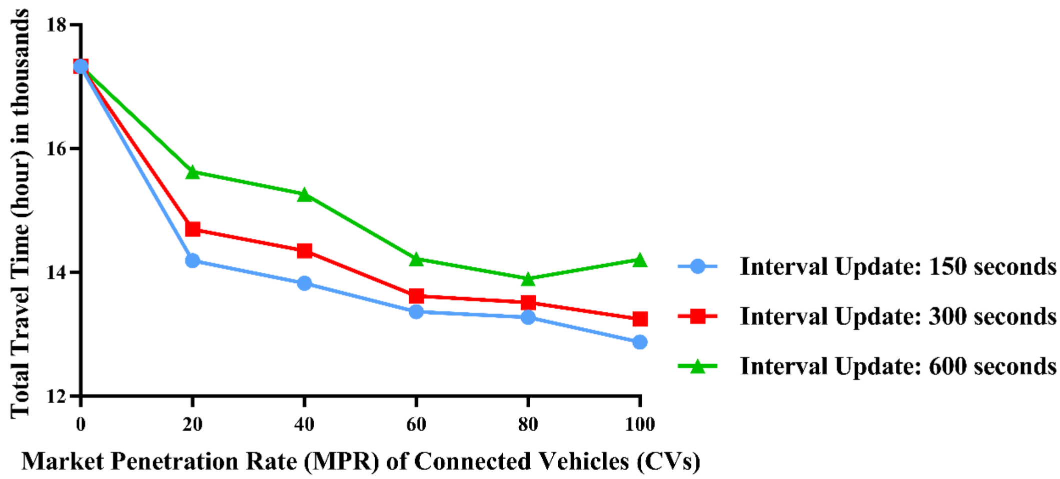


Figure 25. Total travel time for different market penetration rates (MPR) of connected vehicles and different update intervals

Table 21 and Table 22 show the network performance measures relative to the base case (without any communication technology) for update intervals of 300 seconds (Table 21) and 150 seconds (Table 22) with an incident duration of 900 seconds. Based on the results shown in Table 18, Table 21 and Table 22, the lowest assumed update interval (150 seconds) performs better with the highest benefits gained which are around 26% reduction in fuel consumption and travel time.

This would result in 3.51% increase in vehicle-miles traveled.

Table 21. Impacts of connected vehicle technology on the network with update interval of 300 seconds

Market penetration rate	Fuel consumption	Vehicle-miles traveled	Total Travel time
20%	-16%	2%	-15%
40%	-18%	3%	-17%
60%	-22%	3%	-21%
80%	-22%	3%	-22%
100%	-24%	3%	-24%

Table 22. Impacts of connected vehicle technology on the network with update interval of 150 seconds

Market penetration rate	Fuel consumption	Vehicle-miles traveled	Total Travel time
20%	-18%	2%	-18%
40%	-21%	3%	-20%
60%	-23%	4%	-23%
80%	-24%	4%	-23%
100%	-26%	4%	-26%

As presented in this section, an increase in the duration of incidents and number of lanes closed worsens the operation of the network. With the connected vehicles technology, assuming a 600 seconds interval of time for updating the network traffic data for connected vehicles, a reduction of 20% in travel time for an incident with the duration of 900 seconds, and a reduction of 23% for an incident with the duration of 2400 seconds were observed. A sensitivity analysis of the parameters demonstrated that reducing the interval of time for updating traffic information improved the efficiency of the network based on travel time. The reduction in the network-wide travel time and will reach 26% if the update interval is reduced to 150 seconds.

From an environmental perspective, the rerouting algorithm discussed here significantly reduced fuel consumption. As the amount of CO2 emitted from a vehicle is directly related to the fuel consumed, this reduction in fuel consumption will result in the reduction of pollutants and as the consequence will improve the air quality in the area. As estimated by the simulation model, the rerouting strategy will reduce the amount of fuel consumed by 26% for the case of a 900-second incident and update interval of 150 seconds. The total fuel consumed for the base case is about 23,635 liters. Therefore, the reduction is about 6,137 liters which corresponds to a reduction of 14,238,523 grams of CO2.

Table 23 provides a summary of the first and second section of the analysis part. Overall, 4 rerouting strategies were developed, and a sensitivity analysis was performed on various factors including market penetration rate of connectivity, acceptance rate, congestion threshold, update interval and incident duration on the overall performance of the network. As demonstrated in this table, the real-time travel time and the average travel time approach performed similarly. The highest benefits for these approaches occurred at the rerouting rate of 20%. The other two approaches of rerouting resulted in the lower performance. The sensitivity analysis on various factors demonstrated that all of these factors including the duration of incident, the update interval and the acceptance rate of travelers are critical factors. Deployment of connected vehicles requires to consider and evaluate all these factors to plan accordingly to achieve the highest performance. Overall, the results of the study with various scenarios confirmed the benefits of the connected vehicles deployment even at low market penetration rate.

Table 23. Impacts of connected vehicle technology on the network with update interval of 150 seconds

Part 1									
Evaluating rerouting strategies									
Rerouting strategies Examined	Update interval (sec)	Congestion Threshold	Incident Duration (sec)	Market Penetration Rate	Acceptance Rate	Traveler Rerouting Rate	Fuel consumption	VMT	TT
section a									
Rerouting based on the real-time travel time	30	40%	0	-	-	20%	24352	232419	17955
	60			-	-		19023	239317	14000
	90			-	-		18547	238272	13650
	120			-	-		19420	237690	14299
	150			-	-		19721	237570	14521
section b									
Rerouting based on the real-time travel time	150	20%	0	-	-	20%	19357	237020	14254
		40%		-	-		19721	237570	14521
		60%		-	-		19279	237314	14193
section c									
1.Rerouting based on the real-time travel time	150	40%	900	-	-	0%	23636	232643	17426
				-	-	10%	18457	235671	13588
				-	-	20%	16998	244439	12503
				-	-	40%	17207	243329	12659
				-	-	60%	17239	242732	12684
				-	-	80%	18434	242947	13570

Table 23 (continued). Impacts of connected vehicle technology on the network with update interval of 150 seconds

Part 1									
Evaluating rerouting strategies									
Rerouting strategies Examined	Update interval (sec)	Congestion Threshold	Incident Duration (sec)	Market Penetration Rate	Acceptance Rate	Traveler Rerouting Rate	Fuel consumption	VMT	TT
section c									
2. Rerouting based on the average travel time	150	40%	900	-	-	0%	23636	232643	17426
				-	-	10%	18648	239230	13728
				-	-	20%	16427	243029	12080
				-	-	40%	17170	242335	12632
				-	-	60%	17541	241924	12908
				-	-	80%	17924	241165	13192
3. Rerouting based on the Fuel consumption	150	40%	900	-	-	0%	23636	232643	17426
				-	-	10%	19953	235118	14695
				-	-	20%	19302	237710	14213
				-	-	40%	18252	240651	13436
				-	-	60%	18851	237341	13880
				-	-	80%	19444	236157	14315
4.Rerouting based on the modified version of travel time	150	40%	900	-	-	0%	23636	232643	17426
				-	-	10%	18702	239972	13768
				-	-	20%	18437	239286	13572
				-	-	40%	17787	240738	13090
				-	-	60%	17496	242534	12874
				-	-	80%	18740	242481	13797

Table 23 (continued). Impacts of connected vehicle technology on the network with update interval of 150 seconds

Part 1									
Evaluating rerouting strategies									
Rerouting strategies Examined	Update interval (sec)	Congestion Threshold	Incident Duration (sec)	Market Penetration Rate	Acceptance Rate	Traveler Rerouting Rate	Fuel consumption	VMT	TT
section d									
Rerouting based on the real-time travel time	90	40%	900			20%	17233	224771	12683
section e									
Rerouting real-time travel time	150	60%	900			20%	18997	237839	13985

Table 23 (continued). Impacts of connected vehicle technology on the network with update interval of 150 seconds

Part 2									
Evaluating incident scenarios									
Rerouting strategies Examined	Update interval (sec)	Congestion Threshold	Incident Duration (sec)	Market Penetration Rate	Acceptance Rate	Traveler Rerouting Rate	Fuel consumption	VMT	TT
section a									
Rerouting based on the real-time travel time	600	40%	900	0%	Figure 15	-	23636	232643	17330
				20%		-	21209	236025	15627
				40%		-	20723	236908	15266
				60%		-	19310	237621	14221
				80%		-	18885	237723	13905
				100%		-	19297	237936	14212
Rerouting based on the real-time travel time	600	40%	1500	0%		-	24795	232538	18286
				20%		-	19805	235849	14588
				40%		-	19821	236701	14600
				60%		-	19596	237664	14434
				80%		-	19096	238050	14062
				100%		-	18978	238065	13974
Rerouting based on the real-time travel time	150	40%	900	0%		-	23636	232643	17330
				20%		-	19273	237940	14192
				40%		-	18779	240238	13828
				60%		-	18161	241980	13368
				80%		-	18036	241169	13276
				100%		-	17498	240805	12878

Table 23 (continued). Impacts of connected vehicle technology on the network with update interval of 150 seconds

Part 2									
Evaluating incident scenarios									
Rerouting strategies Examined	Update interval (sec)	Congestion Threshold	Incident Duration (sec)	Market Penetration Rate	Acceptance Rate	Traveler Rerouting Rate	Fuel consumption	VMT	TT
section a									
Rerouting based on the real-time travel time	600	40%	2400	0%	Figure 15	-	24235	232747	17845
				20%		-	20359	235982	14998
				40%		-	19909	236880	14665
				60%		-	19500	237706	14422
				80%		-	18914	237953	13783
				100%		-	19091	238032	13889
Rerouting based on the real-time travel time	300	40%	900	0%		-	23636	232643	17330
				20%		-	19957	236886	14700
				40%		-	19492	238818	14355
				60%		-	18506	239038	13624
				80%		-	18360	239840	13516
				100%		-	17995	239722	13247

Network Fundamental Diagram

In the previous sections, the impacts of rerouting strategies on the network performance was investigated with the focus of network-wide travel time and fuel consumption. In this section, the impacts of this technology at the macroscopic level was demonstrated. To this end, a network fundamental diagram of traffic flow relationships is established using the trajectory data of vehicles taken from the microscopic model. Due to the importance of fuel consumption, which is directly related to emissions production, a fuel consumption dimension to the network fundamental diagram may prove beneficial. Aggregated fuel consumption for different levels of flow and density are estimated using an available fuel consumption model. The impacts of connectivity on the network throughput and the relationship of macroscopic traffic variables with fuel consumption for different levels of connectivity are exhibited. In this section, similar to what was done in the first section of the analysis (Rerouting Strategies), the rerouting rate (the combination of MPR and acceptance rate defined in Chapter VI) was used.

The Fuel-flow, flow-density and flow-speed diagrams for the base case scenario (0% rerouting) and for different rerouting rates (10%, 20%, 40%, 60% and 80%) of connected vehicles is presented in Figure 26,Figure 27 and Figure 28. These diagrams were presented to examine the impacts of rerouting strategies at the macroscopic level. Vehicle trajectories were employed to quantify the fuel consumption in the network using Comprehensive Power-based Fuel Consumption Model (CPFM). Toyota Camry 2012 was assumed as the typical vehicle for calibrating CPFM parameters. For rerouting of connected vehicles technology, the first approach of rerouting was used. The same assumptions on the update interval (150 seconds) and congestion level (40%) were considered. Figure 26 illustrates the relationship between the flow and fuel consumption for different rerouting rates. As it can be expected, an increase in the flow results in higher fuel consumption. However, the emergence of the connected vehicles has the potential to not only improve the mobility (higher flow for the same density presented in Figure 27), but also to reduce the fuel consumption and associated emissions.

The flow-density diagram presented to examine the throughput of the network in the presence of rerouting strategies in a connected environment. Figure 27 presenting this relationship demonstrates that with the increase of rerouting rate, higher flows are observed for the same density. Therefore, the increasing rate of connected vehicles (rerouting) increases the throughput of the network, which may prove beneficial especially in the event of blocked lanes or incidents on the roads. However, as the rerouting rate increases the rate of increase in throughput decreases. Lastly, Figure 28 demonstrates the flow-speed relationships. As presented in this figure, with the same flow over the network, higher speed was observed for higher rates of connectivity.

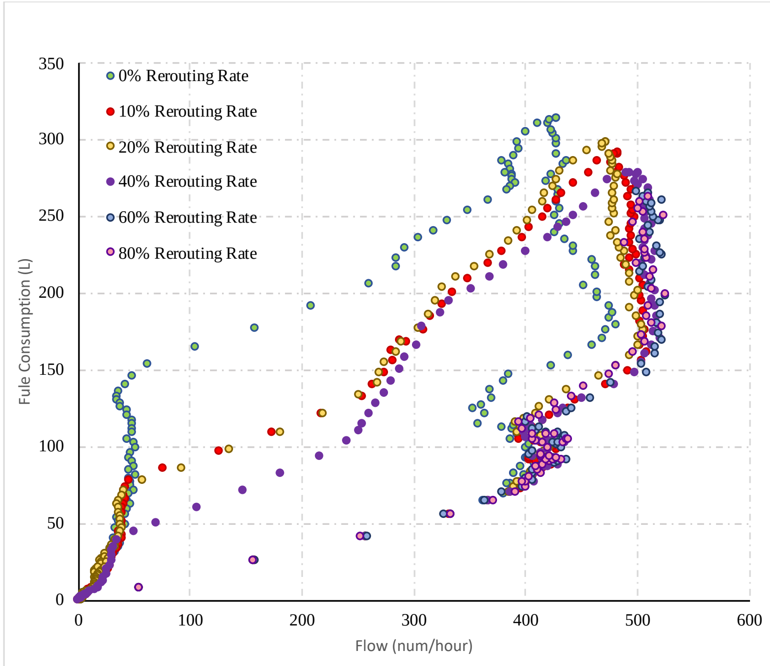


Figure 26. Fuel-flow relationship

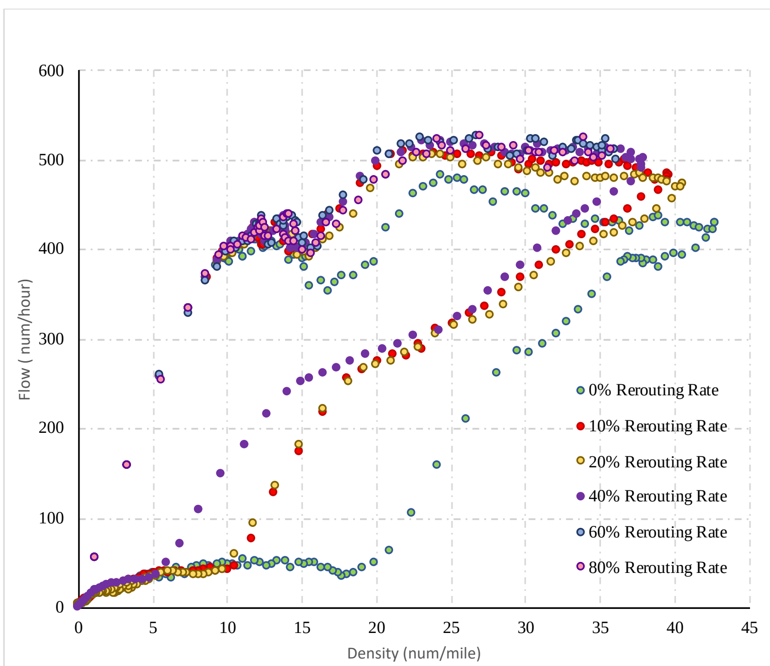


Figure 27. Flow-density diagram

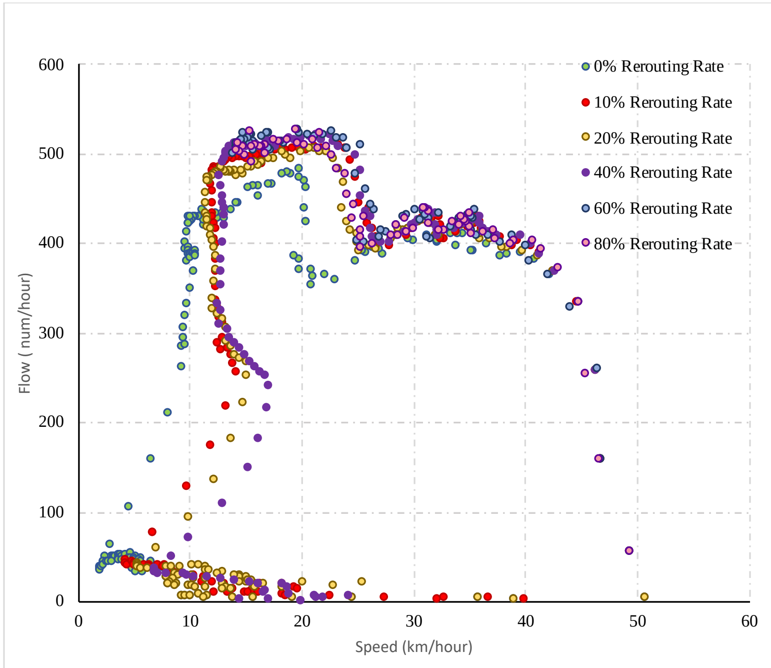


Figure 28. Flow-speed diagram

CHAPTER VIII

CONCLUSION

A route guidance system that provides information on the alternative routes during recurring and non-recurring congestions can be useful for drivers. Individual drivers may get this information from the variable message signs, radio broadcast, and some applications via smart phones. Connected vehicle technology may improve on this as it facilitates data exchange among vehicles and infrastructure. The real-time traffic information communicated between connected vehicles and infrastructure may enable a more efficient management of the transportation system which has the potential to improve mobility, increase safety, and reduce the harmful environmental impacts from the transportation system. Receiving real-time traffic data by vehicles, enables the drivers to make more informed decisions during the trip which improves the performance of the transportation system. The main objective of this study was to investigate the potential impacts of rerouting strategies in the connected environment. To this end, a microsimulation model of part of the network of El Paso, Texas was developed to assess the impacts of en-route decisions made by travelers using the information communicated by connected vehicle technology. The east part of El Paso, Texas was selected in this study. Simulation of Urban Mobility (SUMO) was utilized to develop the microsimulation model of the selected network.

First four rerouting algorithms were developed including rerouting based on the real-time travel time, average travel time, fuel consumption, and the modified travel time accounting for the variance of individuals’ travel time. The average travel time and real-time travel time approaches resulted in the highest benefits at the rerouting rate of 20%. For fuel consumption and the modified travel time approaches, the least total travel time and fuel consumption occurred at the rerouting rate of 40% and 60%, correspondingly. After this rerouting rate, an increase in the total travel time and fuel consumption was observed. For the vehicles-miles-traveled, an increasing pattern with the increase in the rerouting rate was observed.

This study also explored the impacts of connected vehicle technology on the network-wide travel time and fuel consumption for various incidents scenarios. The first implemented rerouting strategy, real-time travel time approach, which works the best among the rerouting strategies, was used in this part of study. In order to consider the travelers' responses to the rerouting information communicated by connected vehicles technology, an internet-based questionnaire was employed. A Normal distribution was developed based on the acceptance percentage estimated from the data of the survey to consider the random behaviour of travelers in response to the rerouting advice. From the traffic operation perspective, assuming a 600 seconds interval of time for updating the network traffic data for connected vehicles, a reduction of 20% in travel time for an incident with the duration of 900 seconds, and a reduction of 23% for an incident with the duration of 2400 seconds were observed. A sensitivity analysis on the model's parameters demonstrated that reducing the interval of time for updating traffic information improved the efficiency of the network based on travel time. The reduction in the network-wide travel time would reach 26% if the update interval was reduced to 150 seconds. From the environmental perspective, the rerouting algorithm discussed in this study significantly reduced the fuel consumption. As the amount of CO2 emitted from a vehicle is directly related to the fuel consumed, this reduction in fuel consumption will result in the reduction of pollutants and as the consequence improve the air quality in the area. As estimated by the simulation model, the rerouting strategy will reduce the amount of fuel consumed by 26% for the case of a 900-second incident and update interval of 150 seconds at 100% market penetration rate of connectivity.

Finally, the first rerouting approach, the real-time travel time rerouting strategy was evaluated in terms of the network-wide macroscopic variables. To this end, network-wide flow, density, speed and fuel consumption were estimated using the trajectory of all vehicles. The results show the ability of the connected vehicles technology in increasing the throughput of the network and the efficiency of the transportation system by distributing the traffic on the network.

This comprehensive study evaluated the impacts of rerouting in a connected environment on the overall performance of the network. The contribution of the current study is 1. the model which simulated a large network consisting of part of the network of El Paso including 5.6 miles of interstate I10 and 4.8 miles of arterial Montana Ave. as well as the major roads and the local streets between these two roads allowing for more comprehensive impact analysis than previous efforts. 2. Various scenarios of rerouting with the help of real-time travel time and fuel consumption were developed and compared to investigate the network-wide impacts from the traffic operation and fuel consumption perspectives 3. A travel behaviour study was performed using the stated preference data to evaluate 4. Sensitivity analysis was performed on the parameters including MPR of connectivity, Acceptance rate (which were considered together as rerouting rate in most part of the study), congestion threshold, update interval, and incident scenarios including various duration and number of lanes closed.

This study was limited to one type of vehicle and a more general study with various types of vehicles including trucks and other heavy vehicles is valuable. This study used the traffic information to generate shortest path for travelers. However, this information can be used to predict congestion and links travel time which might have more impressive impacts on the network. For example, with the use of machine learning, a framework can be built to learn from the traffic data and improve the performance as more data was received. Furthermore, in this study, the current traffic signal timing were used. However, with the communication technologies, a more efficient traffic signal can be developed. For instance, reinforcement learning can be used to penalize or reward a decision made by the signal and these rewards and penalize can help to optimize the performance of the signals as more data are received. In this study, VT-CPFM model was used for estimating fuel consumption. However, other emissions and fuel consumption can be developed and compared. Finally, collecting the socio-economic characteristic of the study area can help in finding a more accurate travel response to the communicated information.

REFERENCES

1. EPA, U. *Global, Regional, and National Fossil-Fuel CO2 Emissions*. 2017.
2. EPA, U. Inventory of U.S. Greenhouse Gas Emissions and Sinks. 2017.
3. Cummings, M. Electronic Sign Strategies and Their Benefits. *Seventh International Conference on Road Traffic Monitoring and Control*, Vol. 1994, 2003, pp. 141–144. https://doi.org/10.1049/cp:19940443.
4. Ramsay, E. D., J. Y. K. (James Y. K. . Luk, and L. ARRB Transport Research. *Route Choice under Two Australian Travel Information Systems*. ARRB Transport Research, 1997.
5. Davidsson, F., and N. Taylor. ITS Modelling in Sweden Using CONTRAM. *Transportation Research Board*, 2003.
6. Tsirimpa, A., and A. Polydoropoulou. Travelers Response to VMS in the Athens Area. *Proceedings of Research in Transport and Logistics*, 2009, pp. 179–187.
7. Chatterjee, K., N. B. Hounsell, P. E. Firmin, and P. W. Bonsall. Driver Response to Variable Message Sign Information in London. *Transportation Research Part C: Emerging Technologies*, Vol. 10, No. 2, 2002, pp. 149–169. https://doi.org/10.1016/S0968-090X(01)00008-0.
8. Erke, A., F. Sagberg, and R. Hagman. Effects of Route Guidance Variable Message Signs (VMS) on Driver Behaviour. *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 10, No. 6, 2007, pp. 447–457. https://doi.org/10.1016/J.TRF.2007.03.003.
9. Dia, H., and S. Panwai. Modelling Drivers' Compliance and Route Choice Behaviour in Response to Travel Information. *Nonlinear Dynamics*, Vol. 49, No. 4, 2007, pp. 493–509.

10. Bonsall, P. The Influence of Route Guidance Advice on Route Choice in Urban Networks. *Transportation*, Vol. 19, No. 1, 1992, pp. 1–23.

11. Bonsall, P. W., and I. A. Palmer. VMS Signs–the Importance of Phrasing the Message. *Behavioural and Network Impacts of Driver Information Systems*, 1998.

12. Lai, C. J., and K. T. Y. Sedan Drivers' Attention and Response to Variable Message Signs on Freeway in Taiwan. *The 4th International Conference on Traffic and Transportation Psychology*, 2004.

13. Li, X., Y. Cao, X. Zhao, and D. Xie. Drivers' Diversion from Expressway under Real Traffic Condition Information Shown on Variable Message Signs. *KSCE Journal of Civil Engineering*, Vol. 19, No. 7, 2015, pp. 2262–2270. https://doi.org/10.1007/s12205-014-0692-y.

14. Dia, H., and S. Panwai. Evaluation of Discrete Choice and Neural Network Approaches for Modelling Driver Compliance with Traffic Information. *Transportmetrica*, 2009, pp. 1–22. https://doi.org/10.1080/18128600903200596.

15. Dia, H., and S. Panwai. Modelling Drivers' Compliance and Route Choice Behaviour in Response to Travel Information. *Nonlinear Dynamics*, Vol. 49, No. 4, 2007, pp. 493–509. https://doi.org/10.1007/s11071-006-9111-3.

16. Gao, G., and C. Zhang. Influencing Factors of Travel Route Choice in Advanced Transportation Information Service System Using Mixed Logit Model. 2016.

17. *US Census Bureau*. 2018.

18. Khattak, A., A. Polydoropoulou, and M. Ben-Akiva. Modeling Revealed and Stated Pretrip Travel Response to Advanced Traveler Information Systems. *Transportation*

Research Record: Journal of the Transportation Research Board, Vol. 1537, No. 1, 1996, pp. 46–54. https://doi.org/10.1177/0361198196153700107.

19. Abdel-Aty, M. A., R. Kitamura, and P. P. Jovanis. Using Stated Preference Data for Studying the Effect of Advanced Traffic Information on Drivers' Route Choice. *Transportation Research Part C: Emerging Technologies*, Vol. 5, No. 1, 1997, pp. 39–50. https://doi.org/10.1016/S0968-090X(96)00023-X.

20. Bhat, C. R., and S. Castelar. A Unified Mixed Logit Framework for Modeling Revealed and Stated Preferences: Formulation and Application to Congestion Pricing Analysis in the San Francisco Bay Area. *Transportation Research Part B: Methodological*, Vol. 36, No. 7, 2002, pp. 593–616. https://doi.org/10.1016/S0191-2615(01)00020-0.

21. Paulssen, M., D. Temme, A. Vij, and J. L. Walker. Values, Attitudes and Travel Behavior: A Hierarchical Latent Variable Mixed Logit Model of Travel Mode Choice. *Transportation*, Vol. 41, No. 4, 2014, pp. 873–888. https://doi.org/10.1007/s11116-013-9504-3.

22. Ye, F., and D. Lord. Comparing Three Commonly Used Crash Severity Models on Sample Size Requirements: Multinomial Logit, Ordered Probit and Mixed Logit Models Keywords: Sample Size Crash Severity Model Multinomial Logit Model Ordered Probit Model Mixed Logit Model. 2014. https://doi.org/10.1016/j.amar.2013.03.001.

23. Savolainen, P. T., F. L. Mannering, D. Lord, and M. A. Quddus. The Statistical Analysis of Highway Crash-Injury Severities: A Review and Assessment of Methodological Alternatives. *Accident Analysis and Prevention*, Vol. 43, 2011, pp. 1666–1676. https://doi.org/10.1016/j.aap.2011.03.025.

24. Srivastava, T. Difference between Machine Learning & Statistical Modeling. https://www.analyticsvidhya.com/blog/2015/07/difference-machine-learning-statistical-modeling/.

25. UCLA: Statistical Consulting Group. ORDERED LOGISTIC REGRESSION | STATA DATA ANALYSIS EXAMPLES. *Institute for Digital Research and Education*. https://stats.idre.ucla.edu/stata/dae/ordered-logistic-regression/. Accessed Jul. 10, 2018.

26. Hensher, D. A. Stated Preference Analysis of Travel Choices: The State of Practice. *Transportation*, Vol. 21, No. 2, 1994, pp. 107–133. https://doi.org/10.1007/BF01098788.

27. Wang, F., and C. L. Ross. Machine Learning Travel Mode Choices: Comparing the Performance of an Extreme Gradient Boosting Model with a Multinomial Logit Model. *Transportation Research Record: Journal of the Transportation Research Board*, 2018, p. 036119811877355. https://doi.org/10.1177/0361198118773556.

28. Zhang, Y., and Y. Xie. Travel Mode Choice Modeling with Support Vector Machines. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2076, No. 1, 2008, pp. 141–150. https://doi.org/10.3141/2076-16.

29. Omrani, H., O. Charif, P. Gerber, A. Awasthi, and P. Trigano. Prediction of Individual Travel Mode with Evidential Neural Network Model. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2399, No. 1, 2013, pp. 1–8. https://doi.org/10.3141/2399-01.

30. Omrani, H. Predicting Travel Mode of Individuals by Machine Learning. *Transportation Research Procedia*, 2015.

31. Hagenauer, J., and M. Helbich. A Comparative Study of Machine Learning Classifiers for

Modeling Travel Mode Choice. *Expert Systems with Applications*, 2017.

32. Xie, C., J. Lu, and E. Parkany. Work Travel Mode Choice Modeling with Data Mining: Decision Trees and Neural Networks. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1854, No. 1, 2003, pp. 50–61. https://doi.org/10.3141/1854-06.

33. Golshani, N., Shabanpour, R., Mahmoudifard, S. M., Derrible, S., & Mohammadian, A. Modeling Travel Mode and Timing Decisions: Comparison of Artificial Neural Networks and Copula-Based Joint Model. *Travel Behaviour and Society 10*, 2018.

34. Zhao, X., X. Yan, A. Yu, and P. Van Hentenryck. Modeling Stated Preference for Mobility-on-Demand Transit: A Comparison of Machine Learning and Logit Models. 2018.

35. Almeida Correia, de. Automated and Connected Vehicles: Effects on Traffic, Mobility and Urban Design ETH Library. *International Journal of Transportation Science and Technology*, Vol. 6, No. 1, 2017. https://doi.org/10.1016/s2046.

36. Zhang, C. Predictive Energy Management in Connected Vehicles: Utilizing Route Information Preview for Energy Saving. *All Dissertations*, 2010.

37. Kamrani, M., B. Wali, and A. J. Khattak. Can Data Generated by Connected Vehicles Enhance Safety? *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2659, 2017, pp. 80–90. https://doi.org/10.3141/2659-09.

38. Samimi Abianeh, A., M. W. Burris, A. Talebpour, and K. C. Sinha. The Impacts of Connected Vehicles on Fuel Consumption, and Traffic Operation under Recurring and Nonrecurring Congestion. *ASCE Conference on Transportation and Development*, 2018.

39. Olia, A., H. Abdelgawad, B. Abdulhai, and S. N. Razavi. Assessing the Potential Impacts of Connected Vehicles: Mobility, Environmental, and Safety Perspectives. *Journal of Intelligent Transportation Systems*, Vol. 2450, No. September, 2015, pp. 1–15. https://doi.org/10.1080/15472450.2015.1062728.

40. Oh, J., and R. Jayakrishnan. Emergence of Private Advanced Traveler Information System Providers and Their Effect on Traffic Network Performance. *Transportation Research Record:*, 2002.

41. Abdulhai, B., and H. Look. Impact of Dynamic and Safety-Conscious Route Guidance on Accident Risk. *Journal of Transportation Engineering*, Vol. 129, No. 4, 2003, pp. 369–376. https://doi.org/10.1061/(ASCE)0733-947X(2003)129:4(369).

42. Lee, J., and B. Park. Evaluation of Route Guidance Strategies Based on Vehicle Infrastructure Integration Under Incident Conditions. *Transportation Research Record: Journal of Transportation Research Board*, Vol. 2086, No. 2086, 2008, pp. 107–114. https://doi.org/10.3141/2086-13.

43. Park, B., and J. Lee. Assessing Sustainability Impacts of Route Guidance System under Cooperative Vehicle Infrastructure Environment. *Systems and Technology, 2009. ISSST'09. …*, 2009.

44. Mei, B., H. Hu, N. Rouphail, and J.-J. Lee. Simulation Model for Studying Impact of Vehicle-to-Vehicle Wireless Communications on Traffic Network Operations. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2189, 2010, pp. 107–115. https://doi.org/10.3141/2189-12.

45. Yeo, H., S. E. Shladover, H. Krishnan, and A. Skabardonis. Microscopic Traffic

Simulation of Vehicle-to-Vehicle Hazard Alerts on Freeway. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2189, No. 1, 2010, pp. 68–77. https://doi.org/10.3141/2189-08.

46. Dion, F., J.-S. Oh, and R. Robinson. Virtual Testbed for Assessing Probe Vehicle Data in IntelliDrive Systems. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 12, No. 3, 2011, pp. 635–644. https://doi.org/10.1109/TITS.2009.2034017.

47. Kim, H. Improvement of ATIS Model Performance under Connected Vehicle Environment. *The Journal of The Korea Institute of Intelligent*, 2012.

48. Kattan, L., M. Mousavi, B. Far, C. Harschnitz, A. Radmanesh, and S. Saidi. Microsimulation Evaluation of the Potential Impacts of Vehicle-to-Vehicle Communication (V2V) in Disseminating Warning Information under High Incident Occurrence Conditions. *International Journal of Intelligent Transportation Systems Research*, Vol. 10, No. 3, 2012, pp. 137–147. https://doi.org/10.1007/s13177-012-0050-8.

49. Paikari, E., L. Kattan, S. Tahmasseby, and B. H. Far. Modeling and Simulation of Advisory Speed and Re-Routing Strategies in Connected Vehicles Systems for Crash Risk and Travel Time Reduction. *Canadian Conference on Electrical and Computer Engineering*, 2013, pp. 1–4. https://doi.org/10.1109/CCECE.2013.6567837.

50. Paikari, E., S. Tahmasseby, and B. Far. A Simulation-Based Benefit Analysis of Deploying Connected Vehicles Using Dedicated Short Range Communication. *IEEE Intelligent Vehicles Symposium, Proceedings*, 2014, pp. 980–985. https://doi.org/10.1109/IVS.2014.6856462.

51. Xiong, C., X. Chen, X. He, X. Lin, and L. Zhang. Agent-Based En-Route Diversion:

Dynamic Behavioral Responses and Network Performance Represented by Macroscopic Fundamental Diagrams. *Transportation Research Part C: Emerging Technologies*, Vol. 64, 2016, pp. 148–163. https://doi.org/10.1016/j.trc.2015.04.008.

52. Bonsall, P. The Influence of Route Guidance Advice on Route Choice in Urban Networks. *Transportation*, Vol. 19, No. 1, 1992, pp. 1–23. https://doi.org/10.1007/BF01130771.

53. Cao, Z., S. Jiang, J. Zhang, and H. Guo. A Unified Framework for Vehicle Rerouting and Traffic Light Control to Reduce Traffic Congestion. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 18, No. 7, 2017, pp. 1958–1973. https://doi.org/10.1109/TITS.2016.2613997.

54. Lee, J., and B. Park. Evaluation of Route Guidance Strategies Based on Vehicle-Infrastructure Integration Under Incident Conditions. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2086, 2008, pp. 107–114. https://doi.org/10.3141/2086-13.

55. Cazares, J. G., A. Talebpour, and R. Rajbhandari. Analyzing the Benefit of Widespread Use of V2I Communication for Improving Incident Management at a Congested Urban Corridor. *Transportation Research Board 96th*, 2017.

56. Daganzo, C. F. Urban Gridlock: Macroscopic Modeling and Mitigation Approaches. *Transportation Research Part B: Methodological*, Vol. 41, No. 1, 2007, pp. 49–62. https://doi.org/10.1016/j.trb.2006.03.001.

57. Daganzo, C., and N. Geroliminis. An Analytical Approximation for the Macroscopic Fundamental Diagram of Urban Traffic. *Transportation Research Part B:*, 2008.

58. Ji, Y., W. Daamen, S. Hoogendoorn, S. Hoogendoorn-Lanser, and X. Qian. Investigating

the Shape of the Macroscopic Fundamental Diagram Using Simulation Data. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2161, 2010, pp. 40–48. https://doi.org/10.3141/2161-05.

59. Buisson, C., and C. Ladier. Exploring the Impact of Homogeneity of Traffic Measurements on the Existence of Macroscopic Fundamental Diagrams. *Transportation Research Record:*, 2009.

60. Daganzo, C., and N. Geroliminis. An Analytical Approximation for the Macroscopic Fundamental Diagram of Urban Traffic. *Transportation Research Part B:*, 2008.

61. Edie, L. Discussion of Traffic Stream Measurements and Definitions. 1963.

62. Courbon, T., and L. Leclercq. Cross-Comparison of Macroscopic Fundamental Diagram Estimation Methods. *Procedia-Social and Behavioral Sciences*, 2011.

63. Boyacı, B., and N. Geroliminis. Exploring the Effect of Variability of Urban Systems Characteristics in the Network Capacity. *Research Board Annual Meeting, Washington, DC*, 2011.

64. Leclercq, L., and N. Geroliminis. Estimating MFDs in Simple Networks with Route Choice. *Procedia-Social and Behavioral Sciences*, 2013.

65. Geroliminis, N., and B. Boyacı. The Effect of Variability of Urban Systems Characteristics in the Network Capacity. *Transportation Research Part B: Methodological*, 2012.

66. Mahmassani, H., and J. Williams. Investigation of Network-Level Traffic Flow Relationships: Some Simulation Results. *Transportation Research*, 1984.

67. Gartner, N., and P. Wagner. Analysis of Traffic Flow Characteristics on Signalized

Arterials. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1883, 2004, pp. 94–100. https://doi.org/10.3141/1883-11.

68. Gayah, V., and V. Dixit. Using Mobile Probe Data and the Macroscopic Fundamental Diagram to Estimate Network Densities. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2390, 2013, pp. 76–86. https://doi.org/10.3141/2390-09.

69. Saberi, M., H. Mahmassani, T. Hou, and A. Zockaie. Estimating Network Fundamental Diagram Using Three-Dimensional Vehicle Trajectories. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2422, 2014, pp. 12–20. https://doi.org/10.3141/2422-02.

70. EPA. *Using MOVES for Estimating State and Local Inventories of On-Road Greenhouse Gas Emissions and Energy Consumption. Https://Www3.Epa.Gov/Otaq/Stateresources/420b12068.Pdf. Accessed Apr. 8, 2016*. 2016.

71. Barth, M., C. Malcolm, T. Younglove, and N. Hill. Recent Validation Efforts for a Comprehensive Modal Emissions Model. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1750, 2001, pp. 13–23. https://doi.org/10.3141/1750-02.

72. An, F., M. Barth, J. Norbeck, and M. Ross. Development of Comprehensive Modal Emissions Model: Operating Under Hot-Stabilized Conditions. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1587, 1997, pp. 52–62. https://doi.org/10.3141/1587-07.

73. Ahn, K., H. Rakha, A. Trani, and M. Van Aerde. Estimating Vehicle Fuel Consumption and Emissions Based on Instantaneous Speed and Acceleration Levels. *Journal of Transportation Engineering*, Vol. 128, No. 2, 2002, pp. 182–190. https://doi.org/10.1061/(ASCE)0733-947X(2002)128:2(182).

74. Rakha, H., K. Ahn, and A. Trani. Comparison of MOBILE5a, MOBILE6, VT-MICRO, and CMEM Models for Estimating Hot-Stabilized Light-Duty Gasoline Vehicle Emissions. *Canadian Journal of Civil Engineering*, Vol. 30, No. 6, 2003, pp. 1010–1021. https://doi.org/10.1139/l03-017.

75. Rakha, H., K. Ahn, and A. Trani. Development of VT-Micro Model for Estimating Hot Stabilized Light Duty Vehicle and Truck Emissions. *Transportation Research Part D: Transport and*, 2004.

76. Keller, M. Handbook of Emission Factors for Road Transport (HBEFA) 3.1. 2010.

77. Liu, B., and H. C. Frey. Development and Evaluation of a Simplified Version of MOVES for Coupling with a Traffic Simulation Model Motivation. 2012.

78. Liu, B., H. C. Frey, B. Liu, and H. C. Frey. Development of a Simplified Version of MOVES and Application to Iterative Case Studies. 2013.

79. Frey, H. C., and B. Liu. Development and Evaluation of Simplified Version of MOVES for Coupling with Traffic Simulation Model. *Transportation Research Board 92*, 2013.

80. Rakha, H., K. Ahn, K. Moran, and B. Saerens. Virginia Tech Comprehensive Power-Based Fuel Consumption Model: Model Development and Testing. *Research Part D: …*, 2011.

81. Rakha, H. A., K. Ahn, K. Moran, B. Saerens, and E. Van den Bulck. Virginia Tech

Comprehensive Power-Based Fuel Consumption Model: Model Development and Testing. *Transportation Research Part D: Transport and Environment*, Vol. 16, No. 7, 2011, pp. 492–503. https://doi.org/10.1016/j.trd.2011.05.008.

82. Washington, S., M. Karlaftis, and F. Mannering. *Statistical and Econometric Methods for Transportation Data Analysis*. 2011.

83. Saeed, T. U., Y. Qiao, S. Chen, K. Gkritza, and S. Labi. Methodology for Probabilistic Modeling of Highway Bridge Infrastructure Condition: Accounting for Improvement Effectiveness and Incorporating Random Effects. *Journal of Infrastructure Systems*, Vol. 23, No. 4, 2017, p. 04017030. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000389.

84. Ordinal Regression Using SPSS Statistics. https://statistics.laerd.com/spss-tutorials/ordinal-regression-using-spss-statistics.php.

85. Brant, R. Assessing Proportionality in the Proportional Odds Model for Ordinal Logistic Regression. *Biometrics*, Vol. 46, No. 4, 1990, p. 1171. https://doi.org/10.2307/2532457.

86. Decision Tree Tutorials & Notes | Machine Learning | HackerEarth. https://www.hackerearth.com/practice/machine-learning/machine-learning-algorithms/ml-decision-tree/tutorial/. Accessed Feb. 8, 2019.

87. Maren, A., C. Harston, and R. Pap. *Handbook of Neural Computing Applications*. 2014.

88. Rumelhart, D. E., G. E. Hinton, and R. J. Williams. Learning Representations by Back-Propagating Errors.

89. Documentation Scikit-Learn: Machine Learning in Python — Scikit-Learn 0.20.3 Documentation. https://scikit-learn.org/stable/documentation.html. Accessed Mar. 18, 2019.

90. Kingma, D. P., and J. Ba. Adam: A Method for Stochastic Optimization. 2014.

91. Xu, W. Towards Optimal One Pass Large Scale Learning with Averaged Stochastic Gradient Descent. 2011.

92. Raschka, S. When Does Deep Learning Work Better Than SVMs or Random Forests? https://www.kdnuggets.com/2016/04/deep-learning-vs-svm-random-forest.html.

93. Samimi Abianeh, A., M. W. Burris, A. Talebpour, and K. C. Sinha. Investigating the Relationship Between Network Fundamental Diagram and Fuel Consumption in a Connected Environment. *Transportation Research Board 97th*, 2018.

94. Krajzewicz, D., G. Hertkorn, and C. Rössel. SUMO (Simulation of Urban MObility)-an Open-Source Traffic Simulation. *Proceedings of the 4th Middle East Symposium on Simulation and Modelling*, 2002.

95. Krajzewicz, D., J. Erdmann, M. Behrisch, and L. Bieker. Recent Development and Applications of SUMO - Simulation of Urban Mobility. *Int. J. Advances Syst. Measure*, Vol. 5, 2012.

96. Krajzewicz, D. Traffic Simulation with SUMO – Simulation of Urban Mobility, pp. 269–293.

97. Fernandes, P., and U. Nunes. Platooning of Autonomous Vehicles with Intervehicle Communications in SUMO Traffic Simulator. 2010.

98. Krajzewicz, D., J. Erdmann, and M. Behrisch. Recent Development and Applications of SUMO-Simulation of Urban MObility. *International Journal On*, 2012.

99. Behrisch, M., L. Bieker, and J. Erdmann. SUMO–Simulation of Urban Mobility: An Overview. *Proceedings of SIMUL*, 2011.

100. Krajzewicz, D., M. Bonert, and P. Wagner. The Open Source Traffic Simulation Package SUMO. *RoboCup 2006*, 2006.

101. Cárdenas-Benítez, N., R. Aquino-Santos, P. Magaña-Espinoza, J. Aguilar-Velazco, A. Edwards-Block, and A. Medina Cass. Traffic Congestion Detection System through Connected Vehicles and Big Data. *Sensors*, Vol. 16, No. 5, 2016, p. 599. https://doi.org/10.3390/s16050599.

102. Li, W., A. Tizghadam, and A. Leon-Garcia. Robust Clustering for Connected Vehicles Using Local Network Criticality. 2012.

103. Mittal, A., H. S. Mahmassani, and A. Talebpour. Network Flow Relations and Travel Time Reliability in a Connected Environment. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2622, No. 1, 2017, pp. 24–37. https://doi.org/10.3141/2622-03.

104. Samimi Abianeh, A., M. Burris, A. Talebpour, and K. Sinha. The Impacts of Connected Vehicle Technology on Network-Wide Traffic Operation and Fuel Consumption under Various Incident Scenarios. *Transportation Planning and Technology*, Vol. 43, No. 3, 2020, pp. 293–312. https://doi.org/10.1080/03081060.2020.1735752.

105. Abianeh, Arezoo Samimi, Burris, M., A. Talebpour, and K. C. Sinha. The Impacts of Connected Vehicles on Fuel Consumption, and Traffic Operation under Recurring and Nonrecurring Congestion. *Journal of Transportation Engineering (2009).*, 2009.

106. Park, S., H. a. Rakha, K. Ahn, and K. Moran. Virginia Tech Comprehensive Power-Based Fuel Consumption Model (VT-CPFM): Model Validation and Calibration Considerations. *International Journal of Transportation Science and Technology*, Vol. 2, No. 4, 2013, pp.

317–336. https://doi.org/10.1260/2046-0430.2.4.317.

107. Experian Automotive: Midrange Cars Are Top-Selling Segment; Toyota Camry Top Vehicle - Experian Global News Blog. http://www.experian.com/blogs/news/2012/09/11/top-selling-cars/. Accessed Aug. 30, 2017.

108. Arnoldy, M., A. Talebpour, S. H. Hamdar, and J. Dong. The Effects of Safety Parameters on Vehicular Emissions: An Integrated Car Following and Fuel Consumption Modeling Approach. 2014.

109. 2012 Camry Product Info. *Toyota Motor Sales, U.S.A. Inc.* https://www.google.com/search?q=2012+Camry+Product+Info.&oq=2012+Camry+Product+Info.&aqs=chrome..69i57j69i59l2j0.1349j0j4&sourceid=chrome&ie=UTF-8. Accessed Jun. 14, 2017.

110. Rakha, H., I. Lucic, S. Demarchi, and J. Setti. Vehicle Dynamics Model for Predicting Maximum Truck Acceleration Levels. *Journal of*, 2001.

111. *Average Fuel Efficiency of U.S. Light Duty Vehicles | Bureau of Transportation Statistics*. 2016.

112. Elfar, A., C. Xavier, A. Talebpour, and H. S. Mahmassani. Traffic Shockwave Detection in a Connected Environment Using the Speed Distribution of Individual Vehicles. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2672, No. 20, 2018, pp. 203–214. https://doi.org/10.1177/0361198118794717.

APPENDIX A

SURVEY QUESTIONS

For the recent trip you made (described in the early part of survey), imagine you are driving a connected vehicle that is receiving traffic information from other vehicles on the road. What would you do if:

A) You see no congestion on the road, but your vehicle warns that there is congestion ahead in 2 miles. It estimates a 30-minute trip if you stay on your current road, or a 25 minute-trip on a different road.

How likely would you be to switch to the different road?

- ¨ I would take the different road
- ¨ I would likely take the different road
- ¨ I am unsure
- ¨ I would probably not take the different road
- ¨ I would definitely not take the different road

B) You see some congestion on the road ahead and your vehicle warns you that there is congestion ahead. It estimates a 54 minute-trip if you stay on your current road, or a 45 minute-trip on a different road.

Now that you can see some congestion, how likely would you be to switch to the different road?

- ¨ I would take the different road
- ¨ I would likely take the different road

- ¨ I am unsure
- ¨ I would probably not take the different road
- ¨ I would definitely not take the different road