공학석사학위논문

Optimal Traffic-aware Routing of a Shared Autonomous Transportation Service

공유 자율주행 차량 서비스를 활용한 최적 교통 경로 문제

2019 년 8 월

서울대학교 대학원

산업공학과

Benjamin Lim Zhi Min

림 벤자민

Optimal Traffic-aware Routing of a Shared Autonomous Transportation Service

공유 자율주행 차량 서비스를 활용한 최적 교통 경로 문제

지도교수 문일경

이 논문을 공학석사 학위논문으로 제출함

2019 년 5 월

서울대학교 대학원

산업공학과

Benjamin Lim Zhi Min

벤자민의 산업공학석사 학위논문을 인준함

2019 년 6 월

위 원 장 ___조성준___ (인)

부위원장 ___장우진___ (인)

위 원 ___문일경___ (인)

Optimal Traffic-aware Routing of a Shared Autonomous Transportation Service

Benjamin Lim Zhi Min

Department of Industrial Engineering

The Graduate School

Seoul National University

Abstract

This thesis describes a traffic-aware routing problem with shared autonomous vehicles by incorporating jams along traffic flow due to the large population of vehicles in the network. This anticipates that autonomous vehicles will replace privately owned vehicles in the future. To provide an efficient shared common service, the dial-a-ride problem is combined with the traffic flow model to satisfy demand (origin-destination pairs), producing a system-optimal traffic assignment problem solution. Macroscopic traffic flow is modelled via the two--regime transmission model (TTM), utilizing inflow and outflow for each link. The optimal solution demonstrates that an appropriate number of vehicles is utilized regardless of the demand or fleet size due to congestion limitations.

Keywords: Two Regime Transmission Model, DARP, Shared Autonomous Vehicles, Morning Commute, Last Mile

Student Number: 2017-20583

Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1. Background and Purpose

The development of autonomous vehicle (AV) technology has brought about a new form of transportation – the shared autonomous vehicle (SAV) transportation mode. An SAV transportation service, notably autonomous taxicabs that supply an origin-destination (O-D) transportation for travellers, could possibly provide inexpensive transportability on-demand services without the requirement for an operator (Krueger et al., 2016). With the full automation of the vehicle, making driver input obsolete, together with the ability to provide a common shared service, the potential of AVs to change the transportation system landscape is undeniable. SAVs have the potential to further decrease private vehicle ownership considerably, eliminating the case of wastage wherein private cars are left idle for long durations; past studies have concluded that up to 11 personal vehicles can be replaced by a single SAV (Fagnant and Kockelman, 2015).

An SAV service presents many potential positive impacts to society, including reducing carbon emissions as well as energy consumption. However, an underlying problem, that is, congestion, has been left out in the majority of previous studies. AVs are able to travel at higher densities for all specified velocities compared to human-driven cars, therefore leading to increased capacity on the roads (Chang and Lai, 1997). However, an SAV service may in fact result in increased congestion of the roads if poorly planned. The issue of obtaining an optimal SAV route allocation is presented in an SAV routing problem, considering many other factors including fulfilment of service to all travellers. Given the possibility that SAVs will replace all personal vehicles in the future, this would relate to tens of thousands of SAVs on the road. Although extensive vehicle routing problems (VRPs) had been researched in the past, an SAV service would involve a scale many times larger. As such, this thesis addresses a large variable routing problem to determine the optimal route choice of SAVs, while considering the effect that the number of vehicles present on the road has on congestion.

An SAV routing problem bears similarities to the Dial-a-Ride problem (DARP) (Cordeau and Laporte, 2007), where passengers specify their pickup and delivery requests between origins and destinations while concurrently minimizing cost for each vehicle route. This thesis addresses the issue of morning commute/last-mile service, wherein demand is relatively fixed but high. Considering the traffic dynamics, this thesis aims to capture the congestion phenomena in the morning commute/ last-mile scenario, including the dial-a-ride behaviour of SAVs as well as the traffic flow. As such, with each SAV chained and time dependent, this problem is classified as a type of Dynamic Traffic Assignment (DTA) problem. Poor routing allocation will result in bottlenecks and gridlocks due to traffic congestion. The contributions of this thesis are as such: a linear program formulation and an analysis of the routing of SAV utilizing discrete traffic flow via the two-regime transmission model (TTM) (Balijepalli et al., 2014) to solve the SAV morning commute/ last-mile routing problem. This thesis presents the first dynamic system optimum (DSO)-DTA formulation using TTM, where previous works utilized the Link Transmission Model (LTM) or the Cell Transmission Model (CTM). This presents advantages as TTM has the ability to depict the queue spillbacks similar to LTM, but also the ability to depict peculiarly the distribution of the front shocks in the link, providing a detailed traffic state in the link. This thesis also incorporates the SAV morning commute/last-mile dial-a-ride behaviour, distributing congestion lengths and also being able to determine the optimal fleet required to fulfil the morning commute/last-mile demand. By doing so, this prevents over-supplying and underutilization of SAV and incurring additional cost.

1.2. Literature Survey

The SAV routing problem involves a large number of vehicles and every vehicle being represented in each time space. This poses the main dissimilarity between a typical DARP and this problem, including the impact of the count of vehicles on congestion conditions. TTM is able to reflect the fluctuation of traffic flow within each link depending on time and space; combined with the DARP, this results in the ability to obtain a system-optimal solution.

1.2.1. Shared Autonomous Vehicle

The potential of SAVs provides a plethora of potential advantages to our transportation system. The US National Household Travel Survey found that the number of vehicles that are idle at any time in a day amounts to more than 83% (Administration, 2009). It was also reported in Fortune that today's cars are parked 95% of the time (Morris, 2016). This suggests that there is an excess of cars to serve the current travel patterns of travellers in most locations, and that decreasing car ownership can result in “huge parking space savings,” making cities denser, more liveable and more efficient. Previous studies proposed that as AVs could reposition themselves throughout the network without fulfilling any demand from travellers, an SAV system could be implemented. With SAVs, vehicles could reposition themselves without passengers and provide service to other travellers, decreasing the number of vehicles in operation, thus optimizing land usage in reducing excess parking spaces while providing service to multiple commuters of the same family (Almeida and Arem, 2016). In addition, there is substantial environmental advantages in the form of a decrease in additional vehicle miles travelled (VMT); Shaheen et al. (2013) estimated that car-sharing members reduced their driving distance by as much as 27%, with one-fourth of them forgoing a vehicle acquisition. On the other hand, other research has proven that by staggering and planning properly the time that trips have to be fulfilled, personal vehicles could fulfil demand by several travellers in the same family unit by providing a dial-a-ride service, offering an user equilibrium formulation (Fagnant and Kockelman, 2015). SAVs have the potential to play a significant role in future transportation systems, as a cheap form of on-demand transportation service. SAVs are able to encompass car sharing through trip planning, probably implemented by the government or private taxi companies, providing a cheap taxi transportation service or an on-call transportation service that may be more efficient than current driver-reliant taxi systems. For example, SAVs could be used as a convenient transportation service for morning commute/last-mile situations (transporting people either from starting destinations to a central point or from transit drop-offs to final destinations) that can be implemented in multimodal transportation systems (Krueger et al., 2016).

Existing literature has investigated the implementation of SAVs and its impact on transportation networks. Fagnant et al. (2014) observed that each SAV may be able to replace as many as 11 personal vehicles on a grid network. Fagnant and Kockelman (2018), utilizing 10% of personal trips in Austin, Texas, investigated dynamic ride-sharing in a network and found that a replacement rate of 1:7 was observed between SAVs and personal vehicles. Spieser et al. (2014) conducted a case study based on Singapore, considering the case in which private transportation is replaced by SAVs, and found that a fleet size one third that of currently operational vehicles was sufficient. Subsequently, even though there was a reduction in the number of vehicles on the road, there had to be an increase in the number of vehicular trips to satisfy all demand. Burns et al. (2013) identified the optimal SAV fleet size to provide service to all residents within acceptable waiting times in an urban environment.

1.2.2. VRP and DARP

VRP has been studied extensively in many literatures because of its extensive usage in many transportation problems. Ropke (2005) and Kumar (2012) discussed the various classes and variants of VRPs, including the classic VRP, the pick-up and delivery problem with time windows (PDPTW), and the travelling salesman problem. The formulation of these mathematical models employs operations research methodologies and is characterized as a non-deterministic polynomial-time hard (NP-hard) type of problem.

PDPTW is closely linked to the flexible on-demand peak-hour transportation service as it features a set of goods that needs to be collected at the customer's location and then transported to the destination of the customer's choice. There has been many variants of PDPTW, including extensions where multiple types of vehicles were considered, constrained by various time windows (Bae and Moon, 2016). While PDPTW focuses on the logistics of transportation of goods, DARP is a sub-class of PDPTW that considers the transportation of passengers, where there is one or multiple passengers at a given pick-up location (Dong et al., 2009). DARP is a classic pickup and delivery problem with the objective of scheduling vehicle routing for a set of n requests by passengers. The aim is to minimize the total cost incurred by satisfying those requests. Historically, the problem was formulated for the transportation of elderly or disabled people. However, this certain type of problem can be applied in multiple forms of other transportation systems such as taxis, ambulances, and courier services (Madsen et al., 1995). There are many variants of DARP, and various algorithms to solve this mathematical problem have been proposed. Both Cordeau and Laporte (2007) and Parragh et al. (2012) have provided a comprehensive review of the different variants of DARP that have been developed. The basic case consists of a single vehicle that will service the set of requests throughout the time window. However, in the case of an SAV peak-hour transportation system, the system will rely on a multi-vehicle DARP model.

In general, the DARP can be divided into two categories: (1) static DARP and (2) dynamic DARP. In a static DARP model, the full set of user requests is made known to the operators prior to the scheduling and routing of vehicles. Various objective functions that are proposed, including the minimization of total service cost (Toth and Vigo, 1996), operational costs (Bornd et al., 1997) and, total route length (Cordeau and Laporte, 2003). The objective function also can be formulated to minimize a combination of different parameters that is defined by the user. In a dynamic DARP model, new requests by passengers are introduced into the system at different times. Given that the set of initial requests is already scheduled, and when a new request is introduced, the problem is formulated to insert and accommodate the new request by re-optimizing the objective function. The initial vehicle scheduling and routing remains unchanged in this case (Cordeau and Laporte, 2007). Various insertion algorithms have been suggested to derive a solution to solve the dynamic case including the solution algorithm REBUS developed by Madsen et al. (1995). Sayarshad and Chow (2015) also adapted a version of the travelling salesman problem with pickup and delivery to consider a non-myopic dynamic DARP model. In the case of a morning commute/ last-mile transportation system, that is, when requests for trips are set beforehand by commuters due to the daily routine to and fro from work, the set of user requests are known to the transport provider prior to the dispatch of vehicles. Hence, the dynamic case will not be considered in the current implementation. However, with an increase in the number of users in this form of transportation system in future developments, it is possible to incorporate dynamic requests from passengers in future works. The morning commute/ last-mile SAV routing problem can be further extended to ordinary impromptu trips (recreational, leisure), which will lead to dynamic demand.

Implementation of DARP requires that constraints be placed on vehicular and passenger requirements. For instance, each passenger may require a certain time window for departing and arriving. In addition, for an SAV system, it is necessary to include vehicular capacity constraints if ride-sharing is implemented (Agatz et al., 2012). However, this increases the feasible area of vehicle route choice and assignments, thus increasing the complexity of the formulation exponentially and hence the computational time. Therefore, the DARP with respect to SAVs is an NP-hard problem. Chen et al. (2016) proposed a Tabu search optimization framework to simulate SAV allocation and found that the computational effort increases exponentially as problem size increases. Ride-sharing will not be included in the initial formulation in this thesis but is expected to be added in future expansion. Also, in the typical DARP problem, researchers consider the case where each passenger has a desired departure and arrival time window (Cordeau and Laporte, 2003; Desrosiers et al., 1995; Jaw et al., 1986). However, this may result in the infeasibility if implemented in our model of fulfilling the respective time windows of passengers, coupled with long travel times affected by congestion. As such, this model considers only passengers with a desired departure times but without arrival time windows.

Given the nature of an SAV network, the model presented is in the form of a large-scale optimization problem. Compared to most DARP formulations, where heuristics or metaheuristics (Ho et al., 2018) are developed to solve the NP-hard DARP, the model proposed utilizes a continuous approximation, which is more suitable for a problem with a large number of variables. This approximation method has normally been applied to DTA models (Chiu et al., 2011) but has recently been incorporated into other VRPs such as dynamic network loading and assignment (Carey et al., 2014). The key point to note is that a constant travel time for a vehicle travelling from node to adjacent node is a common assumption in DARP models. Further research has been done to factor in the effects of congestion such as carbon emissions by vehicles between nodes, driving hours regulations with fixed travel time (Rincon-Garcia et al., 2018), extended travel times (Kok et al., 2012), and varying speed (Xiang et al., 2008). However, this assumption is not applicable to the SAV routing problem due to the large fleet size, which in turn will result in congestion depending on the number of vehicles on the route at each respective point in time. In this view, a congestion-aware routing system formulation for SAVs can be produced by integrating with system-optimal DTA.

1.2.3. Traffic-flow Model

There has been much research with respect to modelling the traffic network for DSO traffic assignment problems, where travellers cooperate in making their choices for the overall benefit of the system instead of their individual benefits. This particular routing problem deals with the time-dependent travelling pattern of commuters in a traffic grid to satisfy two objectives: (1)

O-D pairs of demands with respect to time and (2) the total system travel time (TSTT) that travellers spend in the network. To reflect the situations of the traffic networks, many traffic flow models have been developed, including microscopic models—wherein each individual vehicle is tracked to its respective route—and macroscopic models—in which the general behaviour of traffic propagation on the road is used to depict the flow of traffic and vehicle route choice (Wageningen-Kessels et al., 2015). Due to the problem statement, which assumes that there is a large number of SAVs that will be in operation at any time, and the objective, which is to identify the optimal SAV fleet size, we utilized the macroscopic model to represent the traffic conditions in our model.

The macroscopic approach is based on fundamental relationships such as that between flow and density to control traffic flow. Assuming the aggregate behaviour of groups of vehicles, it is able to reflect current traffic conditions in an easier way to validate and observe. Based on the kinematic-wave theory (KWM) developed by Lighthill and Witham (1955), and Richards (1956), traffic propagation is assumed to follow the wave motion in fluids. Daganzo (1994) formulated a discrete version of KWM, developing a traffic model known as CTM. This requires that the model be disaggregated to cell level to accurately reflect traffic conditions, which requires high computational capacity as it is directly proportional to the number of cells that the modeler specifies. LTM was then developed in both discrete form (Yperman, 2007) and continuous form (Han et al., 2016) based on Newell's shortcut solution method (Newell, 1993), where traffic flow is predicted using the difference in traffic flow at one end of the link and the other end, without having any information at any intermediate points. Utilizing the conservation of flows between incoming and outgoing traffic, the flow propagation can be found constrained by the sending and receiving flows between nodes. Nevertheless, even though LTM is able to depict the free-flow travel time delay when the link is not congested and the backward shockwave time delay when the link is fully congested, it is unable to explicitly determine the propagation of the front shocks within a link and is unable to provide a detailed traffic state within each link.

Subsequently, TTM (Balijepalli et al., 2014) was developed to address this problem. The concept of TTM will be described here briefly. Similar to the CTM and LTM, the TTM defines the traffic condition in each link via the entry and exit flows. From the flow-density relationship defined, the speed of the vehicles can be calculated. Given that the density of the vehicles at the respective parts of the link is less than or equal to the critical density, vehicles in the corresponding part of the link travel at free-flow speed. Conversely, in the opposing case, the speed of vehicles can be found by dividing the flow by the density. In addition, the TTM is able to provide the variation in queue lengths within the link through time and space, as shown in Figure 1 below.

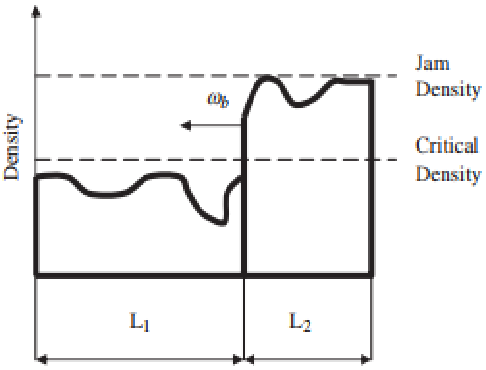

Figure 1. Link traffic state between 2 nodes

The TTM splits traffic in a link into two regimes: (1) the non-congested regime in which density is below critical density and (2) the congested regime wherein density is above critical density. TTM allows us to formulate the state of the link as non-homogenous according to its length (where LTM is unable to) and does not require a computationally intensive discretization method of cell-modelling (as in CTM for the same prescribed level of accuracy). In addition, TTM is able to deal with discontinuity in densities within congested regimes, where there may be multiple situations of varying densities within a single congested regime. Even though there exists more advanced models developed by other researchers (Aubin et al., 2008; Liu et al., 2018), TTM allows one to formulate the traffic flow propagation in a straightforward and efficient manner. This thesis builds upon the preliminary linear programming formulation of TTM to the DSO problem by Ngoduy et al. (2016), combining it with the DARP and SAV routing problem to obtain an optimal solution.

The remainder of this thesis is organized as follows. In Chapter 2, the linear programming formulation of the SAV DSO-DTA problem is presented. Chapter 3 presents numerical experimentation and results, and the conclusion and further discussions are finalized in Chapter 4. By conducting experiments, we reveal insights that would help in implementation of SAVs in real-world scenarios. A summary of related studies is shown in Table 1.

Table 1. Summary of related papers

	Considerations								Methodology	
			Congestion-awareness				Demand			
Author	SAV	DARP	TTM	LTM	CTM	Others	Static Demand	Dynamic Demand	Heuristics	LP
Balijepalli et al. (2014)			√							√
Daganzo (1994)					√					√
Yperman (2007)				√						√
Madsen et al. (1995)		√						√	√	
Sayarshad and Chow (2015)		√						√	√	
Chen et al. (2016)	√								√	
Cordeau and Laporte (2003)		√							√	
Fagnant and Kockelman (2014)	√	√							√	
Toth and Vigo (1996)		√					√		√	
Spieser et al. (2014)	√						√		√	
Carey et al. (2014)						√			√	
Dong et al. (2009)		√							√	
Levin (2017)	√	√		√			√			√
Fagnant and Kockelman (2015)	√							√	√	
Fagnant and Kockelman (2018)	√							√	√	
This Thesis (2019)	√	√	√				√			√

Chapter 2: Mathematical Model

This thesis consists of a DSO-DTA problem, dealing with a directed network with centroids defined as being both source and sink nodes at the same time. Addressing the SAV routing problem requires solving three challenges typically absent in the VRP literature. First, this model solves a large variable linear programming model, tracking the macroscopic flow of vehicles relative to each time period, unlike VRP problems which deal with the routing of a small discrete fleet of vehicles. Second, it is noted that congestion will occur due to the large fleet of vehicles and limited capacity of the roads. Therefore, at every time interval, a real-time traffic flow model must be included to account for the number of vehicles on the road. Previous literature typically uses graphs with edges that map the nodes to nodes without considering the dynamic flow of traffic. Furthermore, route choices from the embarkation of the passenger until the point where the passenger is dropped off affects the traffic congestion on the respective links, and each link may be unique in its characteristics; therefore, each link must be modelled accurately to simulate the system dynamics of traffic flow. Lastly, the utilization of TTM provides a comprehensive traffic state within each link, reflecting the propagation of front shocks within the link which enables optimally distributing the queues. Previous literature on traffic flow is able to generalise only the state within the link as a whole, providing less information for planners in finding the optimal distribution with consideration of congestion within the link.

Given that this thesis addresses the morning commute/last-mile problem, where demand is determined beforehand, trip routing behaviour can be planned in advanced. As such, computational time is not of the topmost priority. Nevertheless, fast algorithms are essential to solve the static traffic assignment problem to obtain a system-optimal solution.

2.1. Model Development

Consider a fleet of public AVs providing services similar to those provided by taxis for a large number of commuters. Each commuter possesses a specific origin, destination, and departure time, wherein each customer can be picked up only at or after his/her predefined departure time. Traffic flow is modeled using TTM to solve the DSO traffic assignment problem. This predicts the optimal time-dependent routing pattern of travelers in a network, with the objective of fulfilling all O-D demands, at the same time minimizing TSTT of travelers in the network. To identify the choice of routes that each vehicle takes in the traffic network, we created a DSO-DTA model to describe the time-varying network and demand interaction. Although Levin formulated an optimal DTA formulation using LTM, this is the first time that a formulation implementing TTM is applied to obtain a system-optimal DTA-DSO solution. As such, we termed this problem as a two-regime transmission model-dynamic system optimum-dial a ride problem (TTM-DSO-DARP).

Each vehicle on the traffic network is assumed to be solely a single class of SAVs. Past research (Fagnant and Kockelman, 2014) had applied this assumption too, given that in the future, there will be an aim to replace a large part of the vehicle fleet with SAVs ("Dubai's Autonomous Transportation Stratedgy," 2018; Singapore, 2019). This also is due to the difficulty in modelling the dynamic interaction of SAVs and human-driven vehicles sharing the same road within system-optimal DTA. In this thesis, ride-sharing was not considered due to its complexity. Also, as we are modelling the morning commuter/last-mile problem, where customers' journeys are relatively fixed daily, the demand is assumed to be known beforehand. Each vehicle is defined to carry only a single passenger, and each vehicle is able to travel to another customer immediately after the most recent drop-off. Vehicles are parked at depots (centroids) at the start and end of the time horizon. Within the time horizon, vehicles exist in only two states, travelling within links or parked.

2.2. Traffic Network

The traffic network, $G = (N, A)$, is defined as a set of links, A connected by a set of nodes, N. There exist two types of nodes—centroids and ordinary nodes. Centroids z represent depots or destinations, where O-D pairs are specified. The ordinary nodes represent junctions, where travelers are not allowed to alight nor be picked up. Links that exist between two nodes are denoted by $(i, j) \in A$ and are defined to be bidirectional. This implies that for every (i, j) link that exists, there is a parallel, reversed road link (j, i) that exists.

We define centroids to be connected to the ordinary nodes by centroids connectors. Since travel demands have to be met together with the conservation of flow of traffic, centroids are defined differently from ordinary nodes. Instead of road segments, centroid connectors represent not only the commuter embarkation and disembarkation of the SAVs, but also the parking behavior at the centroids. As such, there are no capacity constraints for centroid connectors, being used to represent just the interface between the centroids and junctions. We define A_0 as the set of links (i, j) that neither begin nor end at a centroid, while A_z is defined as the set of links that either begin or end at a centroid; that is, for each respective centroid connector (i, j) $\in A_z$, each of i or j can represent a centroid, ($i \in z$ or $j \in z$). In addition, we define

$$A_z^- = \{(i, j) \in A_z : j \in z\}$$

as the set of centroid connectors with the end points at a centroid and

$$A_z^+ = \{(i, j) \in A_z : i \in z\}$$

as the set of centroid connectors with their starting points at a centroid. The superscript "–" and "+" represents links coming into a centroid and out of a centroid, respectively.

With regards to nodes where $j \in N$, we represent the sets of incoming links going into j as Γ_j^-, and Γ_j^+ representing the set of outgoing links out of j. Centroids also possess both types of links that can represent vehicles entering and exiting the depot. Discrete time is used in this model, with one time interval representing 30 seconds, and time is indexed by $t \in (0,1, 2,..., T)$. Exogenous parameters utilized in the model are shown in Table 2 to define the network characteristics.

Table 2. Exogenous parameters for network

Notation	Parameter
Length of analysis period	T
Capacity of link (i,j)	Q_{ij}
Length of link (i,j)	L_{ij}
Maximum density of link (i,j)	K_{ij}
Free flow speed of link (i,j)	v_{ij}
Congested wave speed of link (i, j)	w_{ij}
Cost of waiting per customer per unit time	σ_W
Cost in system per vehicle per unit time	σ_V
Number of parked vehicles at centroid i when t=0	$p_i(0)$
Demand originating from r to destination s at time t	$d_{rs}(t)$
Time interval used	*30 secs*

2.3. Explanations on Constraints

The number of vehicles in each link is considered an important decision variable in this TTM-DSO-DARP problem. In addition, other decision variables in this problem include the inflow and outflow for each link, the number of passengers to keep waiting at each node and the outflow from each centroid. Decision variables in this problem are summarized, categorized, and presented in Table 3 below, and will be further elaborated in the latter parts of this thesis.

Table 3. Decision variables

Category	**Decision Variable**	**Description**
Vehicle	$p_j(t)$	*Number of parked vehicles at centroid*
Vehicle	$n_{ij}(t)$/ $n^f_{ij}(t)$ / $n^c_{ij}(t)$	*Number of vehicles in each link*
Flow	$U^s_{ij}(k)$/ $V^s_{ij}(k)$	*Inflow/ outflow of traffic*
Flow	$F^s_{ijk}(t)$	*Continuous flow from* (i,j) *to* (j,k)
Flow	$y^S_{rj}(t)$	*Outflow from centroids*
Travelers	$\omega^s_r(t)$	*Waiting demand travelling to s from r*
Travelers	$e^s_r(t)$	*Number of travelers leaving r to s*

Based on TTM (Lighthill and Whitham, 1955), we define the number of vehicles in a link going to a specific destination s at each time instant $t \in T$ as the number of vehicles within the full length of the link at time t between two nodes. This is represented by the sum of all inflow $U^s_{ij}(k)$ subtracting outflow $V^s_{ij}(k)$ for each time interval from 0 to t, going to a specific destination s. Therefore,

$$n_{ij}^{s}(t) = \sum_{k=0}^{t} \sum_{s \in Z} (U_{ij}^{s}(k) - V_{ij}^{s}(k)) \qquad \begin{array}{l} \forall (i,j) \in A_0 \cup A_z^{+} \\ \forall t \in [0, T-1] \end{array} \qquad (1)$$

Following this, the number of vehicles in each link will be the sum of vehicles in each link (*i, j*) regardless of destination s.

$$n_{ij}(t) = \sum_{s \in z} n_{ij}^{s}(t) \qquad \begin{array}{l} \forall (i,j) \in A_0 \cup A_z^{+} \\ \forall t \in [0, T-1] \end{array} \qquad (2)$$

Two conditions which provide boundaries for the number of vehicles: (1) the number of vehicles if the link is fully congested; that is, when the length of the link which is congested (L_{ij}^{c}(t)) equals the length of the link (L_{ij}(t)), and (2) the number of vehicles if the link is fully non-congested; that is, when the length of the link which is free-flowing (L_{ij}^{f}(i)) equals the length of the link (L_{ij}(t)), can be determined by TTM. As links in centroid connectors are defined to be unlimited in capacity with no constraint (that is $Q_{ij} = \infty$ for (*i, j*) ϵ A_z), only links belonging to A_0 are considered.

$$n_{ij}^{f}(t) = \sum_{k=\left(t-\frac{L_{ij}}{v_{ij}}+1\right)}^{T} U_{ij}^{s}(k) \qquad \begin{array}{l} \forall (i,j) \in A_0 \\ \forall s \in z \\ \forall t \in [\frac{L_{ij}}{v_{ij}} - 1, T] \end{array} \qquad (3)$$

$$n_{ij}^{c}(t) = KL_{ij} - \sum_{k=\left(t-\frac{L_{ij}}{w_{ij}}+1\right)}^{T} \sum_{s \in Z} V_{ij}^{s}(k) \qquad \begin{array}{l} \forall (i,j) \in A_0 \\ \forall s \in z \\ \forall t \in [\frac{L_{ij}}{w_{ij}} - 1, T] \end{array} \qquad (4)$$

The number of vehicles in a link, n_{ij} (t) is constrained by two variables: (1) the number of vehicles when fully non-congested, $n_{ij}^{f}(t)$ and (2) the number of vehicles when fully congested, $n_{ij}^{c}(t)$. Therefore,

$$n_{ij}^{f}(t) \leq n_{ij}(t) \qquad \forall (i,j) \in A_0 \quad \forall t \in [0,T] \qquad (5)$$

$$n_{ij}(t) \leq n_{ij}^{c}(t) \qquad \forall (i,j) \in A_0 \quad \forall t \in [0,T] \qquad (6)$$

At each immediate node that belongs to the ordinary nodes, the flow of vehicles has to obey the conservation of flow. Therefore, the sum of incoming traffic flow into node *i* must be equal to the sum of outgoing traffic flow out of node *i* at every time interval, *t*.

$$\sum_{(i,j)\epsilon\Gamma_j^-} \sum_{s\epsilon Z} V_{ij}^{s}(t) = \sum_{(j,k)\epsilon\Gamma_j^+} \sum_{s\epsilon Z} U_{jk}^{s}(t) \qquad \forall (i,j) \in A_0 \cup A_z^+ \quad \forall (j,k) \in A_0 \cup A_z^- \quad \forall t \in [0,T] \qquad (7)$$

For the solution of the congested regime to be unique via TTM, the following conditions must hold:

$$\sum_{s\epsilon Z} U_{ij}^{s}(t) \leq MAXFLOW - 1 \qquad \forall (i,j) \in A_0 \quad \forall t \in [0,T] \qquad (8)$$

$$\sum_{s\epsilon Z} V_{ij}^{s}(t) \leq MAXFLOW - 1 \qquad \forall (i,j) \in A_0 \quad \forall t \in [0,T] \qquad (9)$$

where $MAXFLOW = \frac{K_{ij} v_{ij} w_{ij}}{v_{ij} + w_{ij}}$.

To obtain a unique length of the congested regime, we have to prevent the total flow from reaching the critical point, equivalent to the highest flow in the triangular fundamental diagram used to formulate the KWM model. Bounding the dynamic maximum flow prevents maximum flow from occurring at any particular time at any place within the link such that there is a unique solution for TTM.

To model the network in such a way that it is tractable via continuous flow, we define $F^{s}_{ijk}(t) \in R_{+}$ as the vehicular flow from $(i,j) \in A$ to $(j,k) \in A$ moving to $s \in z$ during the time t. Therefore, for every incoming traffic flow,

$$U^{s}_{jk}(t) = \sum_{(i,j) \epsilon \Gamma^{-}_{j}} F^{s}_{ijk}(t) \qquad \begin{array}{l} \forall (i,j) \in A_0 \cup A^{-}_{z} \\ \forall (j,k) \in A_0 \cup A^{+}_{z} \\ \forall s \in z \\ \forall t \in [0,T] \end{array} \tag{10}$$

and every outgoing traffic flow,

$$V^{s}_{ij}(t) = \sum_{(j,k) \epsilon \Gamma^{+}_{j}} F^{s}_{ijk}(t) \qquad \begin{array}{l} \forall (i,j) \in A_0 \cup A^{+}_{z} \\ \forall (j,k) \in A_0 \cup A^{-}_{z} \\ \forall s \in z \\ \forall t \in [0,T] \end{array} \tag{11}$$

We now define flow and parking behaviour at centroids. We further define $y^{s}_{rj}(t)$ as the outflow from centroids, r to ordinary nodes, j where $r \in z$ and $(r, j) \in A_z$. As such, the outflow from centroid i must be equal to the incoming traffic flow into centroid link (i, j).

$$y_{ij}^{s}(t) = U_{ij}^{s}(t) \qquad \forall (i,j) \in A_z^{+} \quad \forall s \in z \quad \forall t \in [0, T-1] \qquad (12)$$

For simplicity, we assume that $\frac{L_{ij}}{v_{ij}} = 1$ for centroid connectors in A_z^{-} or A_z^{+}. Thus, incoming traffic flow into link (i,j) $\in A_z^{-}$ *or* A_z^{+} requires a time interval of only 1 to exit the centroid link:

$$V_{ij}^{s}(t+1) = U_{ij}^{s}(t) \qquad \forall (i,j) \in A_z^{-} \cup A_z^{+} \quad \forall s \in z \quad \forall t \in [0, T-1] \qquad (13)$$

Parking behaviour is modelled at the centroids. We define $p_j(t) \in R_+$ as the number of parked vehicles at $j \in z$ at time t. Then, $p_j(t)$ changes through the relationship

$$p_j(t+1) = p_j(t) + \sum_{(m,j)\in\Gamma_j^{-}} V_{mj}^{j}(t) - \sum_{(j,k)\in\Gamma_j^{+}} \sum_{s\in Z} y_{jk}^{s}(t) \qquad \forall j \in z \quad \forall t \in [0, T-1] \qquad (14)$$

Entry is not allowed to vehicles in any centroid connector $(i,j) \in A_z^{-}$ unless their final destinations corresponds to the respective centroid. As such,

$$F_{ijk}^{s}(t) = 0 \qquad \forall (j,k) \in A_z^{-} \quad \forall (i,j) \in \Gamma^{-} \quad \forall s \neq k \quad \forall t \in [0, T] \qquad (15)$$

Upon reaching a centroid, a vehicle returns to the state of being parked.

These vehicles are not allocated any demand to satisfy at that point in time and are parked at the centroid for the time being. A vehicle that leaves a centroid goes to any one of the other centroids apart from itself. Thus, the number of parked vehicles at i provides a bound on the amount of outgoing flows at i by:

$$\sum_{(i,j)\epsilon\Gamma_i^+} \sum_{s\epsilon Z} y_{ij}^s(t) \leq \mathrm{p_i}(\mathrm{t}) \qquad \begin{array}{l} \forall i \in z \\ \forall s \in z \\ \forall t \in [0, T] \end{array} \tag{16}$$

The model is assumed to have all vehicles parked at the beginning and at the end of the time horizon. Therefore, the number of vehicles in all links at $t =$ 0 and $t =$ T will be equal 0; thus

$$n_{\mathrm{ij}}(0) = 0 \qquad \forall (i, j) \in A \tag{17}$$

$$n_{\mathrm{ij}}(\mathrm{T}) = 0 \qquad \forall (i, j) \in A \tag{18}$$

The number of vehicles parked at $t = 0$ has to be the same as that at $t =$ T (end of horizon). Thus

$$\sum_{\mathrm{i \in Z}} \mathrm{p_i}\,(0) = \sum_{\mathrm{i \in Z}} \mathrm{p_i}(\mathrm{T}) \tag{19}$$

$\mathrm{p_i}(0)$ is taken as an exogenous variable that will be specified to indicate the number of SAVs initially parked at i at the beginning of the model simulation. In addition, this is an important variable for planners in identifying the ample amount of vehicles so as to satisfy all demand with minimum cost.

We further define constraints to satisfy the conditions of DARP. We define the number of commuter-trip demand originating from r to s, with r, $s \in z$ at time i as $\mathrm{d}_{\mathrm{r}}^{\mathrm{s}}(\mathrm{t})$. As this problem tackles the morning commute/ last-mile problem, perfect information about demand is known beforehand. A vehicle would not be able to pick up a customer departing at t before the designated time t. Each person-trip demand corresponds to $\mathrm{d}_{\mathrm{r}}^{\mathrm{s}}(\mathrm{t})$ vehicle trips; that is, one-person trip demand requires one vehicle to satisfy. We further define the number of serviced demands awaiting departure at r to travel to destination s as $\omega_{\mathrm{r}}^{\mathrm{s}}(t)$. Demand waiting at r to s, $\omega_{\mathrm{r}}^{\mathrm{s}}(t)$ is satisfied by vehicle trips starting at r going to s. Also, the number of travellers leaving to destination s at time t for each time t is defined as $\mathrm{e}_{\mathrm{r}}^{\mathrm{s}}(\mathrm{t})$.

With these variables, constraints are formulated to represent the movement of parked vehicles to satisfy each respective demand per time interval t. First, the number of customers leaving the depot cannot exceed the number of customers waiting:

$$\mathrm{e}_{\mathrm{r}}^{\mathrm{s}}(\mathrm{t}) \leq \omega_{\mathrm{r}}^{\mathrm{s}}(t) \qquad \forall (r,s) \in z^2 \quad \forall t \in [0,T] \qquad (20)$$

The number of customers who can leave the depot cannot be more than the number of vehicles that are leaving the depot.

$$\mathrm{e}_{\mathrm{r}}^{\mathrm{s}}(\mathrm{t}) \leq \sum_{(i,j)\epsilon\Gamma_r^+} y_{rj}^{\mathrm{s}}(t) \qquad \forall (r,s) \in z^2 \quad \forall t \in [0,T] \qquad (21)$$

The number of customers waiting at *r* going to *s* at each time interval *t+1* depends on the number of customers waiting at the time interval *t*, together with the fixed demand at time *t*, less the number of customers who are leaving:

$$\omega_r^s(t+1) = \omega_r^s(t) + d_r^s(t) - e_r^s(t) \qquad \forall (r,s) \in z^2 \quad \forall t \in [0, T-1] \qquad (22)$$

In this way, the waiting demand that varies over time can be traced. However, it must be noted that even if there is no demand for trips, that is $d_r^s(t) = 0$, there may be repositioning trips from other centroids. Waiting demand, which is inclusive of all demand, has to be satisfied when the time simulated ends; thus:

$$\omega_r^s(T) = 0 \qquad \forall (r,s) \in z^2 \qquad (23)$$

It is important to note that even though demand is defined to be satisfied at the end of the simulated time, this may result in high waiting times over the time horizon. For example, a demand at time period 1, $\omega_r^s(1)$ may be satisfied only at time period T–1 while still obeying the constraint. As such, minimizing the customer waiting time must be included in the objective function to create a more realistic scenario.

As defined earlier, the number of travellers leaving the depot and the outgoing flows from the centroids are constrained to be more than or equal to 0.

$$y_{rj}^{S}(t) \geq 0 \qquad \forall (r,s) \in z^2 \quad \forall (r,j) \in A_z^+ \quad \forall t \in [0,\, T] \qquad (24)$$

$$e_r^s(t) \geq 0 \qquad \forall (r,s) \in z^2 \quad \forall t \in [0,\, T] \qquad (25)$$

In addition, given the number of vehicles in each link, the length of the congested regime $L_{ij}^{c}(t)$ at every point in time can be approximated via the following equation through the fixed-point principle below:

$$L_{ij}^{c}(t) = \frac{n_{ij}(t) - n_{ij}^{f}(t)}{n_{ij}^{c}(t) - n_{ij}^{f}(t)} L_{ij} \qquad \forall (i,j) \in A \quad \forall t \in [0, T] \qquad (26)$$

2.4. Objective Function

The objective function can be split into two parts. The first objective was to minimize the cost of the waiting time of customers such that customers will not have to wait for a long time between their required set-off time and the actual time of departure. Minimizing the waiting time of customers also corresponds to maximizing vehicle departures and thus inflow of vehicles at the centroids. For each time interval that each customer wait, a homogenous cost is assumed to be incurred, denoted by σ_w. As such, total waiting time cost per centroid r per destination s equals $\sigma_w \omega_r^s$.

Secondly, the total system cost corresponds to the cost per unit time usage σ_p, multiplied by the number of vehicles in the transport grid at any point in time. This is indicated by the total number of SAVs present in the system, less the number of vehicles parked at the centroids at any point in time. This is because the vehicles that are not parked at centroid i at time t are incurring costs as they are in transit from node to node, incurring operating costs in terms of carbon footprint and energy consumption. Therefore, any vehicle not parked is assumed to be incurring a cost of σ_v per unit time. This includes not only vehicles that are bringing passengers from origin to destination, but also repositioning trips wherein supply of vehicles in other centroids is insufficient to meet current demand. Even though this does not segregate demand-fulfilling trips with passengers and repositioning trips, travelling time of travellers still will be minimized by maximizing the number of parked vehicles at every time interval, which is an overestimation of actual travel times. Non-holding-back conditions (Shen et al., 2007), in which the system discharges as much flow as it can, is satisfied by this objective function.

Combining these two parts wherein minimizing of (1) total waiting time cost of customers and (2) total system cost over specified time T, the optimization problem produces an optimal solution of route choice for the SAV routing problem. Thus, the objective function is

Minimize

$$C = \sigma_v \sum_{t=0}^{T} \left(\sum_{j \epsilon Z} p_{\mathrm{j}}(0) - \sum_{j \epsilon Z} p_{\mathrm{j}}(\mathrm{t}) \right) + \sum_{(r,s) \epsilon z^2} \sum_{t=0}^{T} \sigma_w \omega_r^s \tag{27}$$

Remarks:

The movement of vehicles with respect to their O-D and route choices can be described wherein (1) routes begin at centroid z and (2) routes end at z. Parked vehicles located initially at z decrease in number according to the vehicles dispatched to satisfy demand, and when flow destined to some $j \in z$ arrives at centroid *z*, $p_i(t)$ is updated accordingly. Parked vehicles are allowed to relocate to other centroids to fulfill demand at those centroids when the number of vehicles at those locations is insufficient. Upon entering the network, every individual vehicle possesses a fixed destination, and flows behave according to TTM via the following constraints. In between link flows, $F_{ijk}^{s}(t)$ behaves accordingly to inflow $U_{ij}^{s}(t)$ and outflow $V_{ij}^{s}(t)$ traffic and is used as the optimization decision variable in the objective function. Route choice is decided by the inflow and outflow variables, and turning movement variables are specified similarly depending on destination, $s \in z$.

Another objective function given by the total system travel time (TSTT) can also be used to obtain an optimal solution, given by minimizing:

$$TSTT = \sum_{(i,j)\epsilon A \cap (j,k)\epsilon A} \sum_{s \epsilon Z} \sum_{t=0}^{T} n_{ij}\ (t) + \sum_{(r,s)\epsilon z^2} \sum_{t=0}^{T} \omega_r^s.$$

The number of vehicles in each link per unit time is a decision variable to determine the number of trips departing toward destination s along (i, j) at time t, which also encapsulates both passenger carrying trips and empty trips. As the number of vehicles influences the objective value based on the constraints of the maximum number of vehicles in a link due to road capacity and congestion, only trips of repositioning or passenger carrying trips are taken when necessary.

2.5. Mathematical Formulation

With the constraints and objective function presented above, we formulated the entire linear program as below. In addition, constraint 40 was added as outflow traffic from one node to another takes place only after the amount of time that vehicles can travel on roads without congestion has elapsed at the start of simulation. As such, the linear program is as follows:

Objective:

Minimize

$$C = \sigma_v \sum_{t=0}^{T} \left(\sum_{j \epsilon Z} p_j(0) - \sum_{j \epsilon Z} p_j(t) \right) + \sum_{(r,s) \epsilon z^2} \sum_{t=0}^{T} \sigma_w \omega_r^s \tag{28}$$

$$\sum_{(i,j)\epsilon\Gamma_j^-} \sum_{s \epsilon Z} V_{ij}^s(t) = \sum_{(j,k)\epsilon\Gamma_j^+} \sum_{s \epsilon Z} U_{jk}^s(t) \qquad \forall (i,j) \in A_0 \cup A_z^+ \quad \forall (j,k) \in A_0 \cup A_z^- \quad \forall t \in [0,T] \tag{29}$$

$$n_{ij}^s(t) = \sum_{k=0}^{t} \sum_{s \in Z} \left(U_{ij}^s(k) - V_{ij}^s(k) \right) \qquad \forall (i,j) \in A_0 \cup A_z^+ \quad \forall t \in [0, T-1] \tag{30}$$

$$n_{ij}(t) = \sum_{s \in z} n_{ij}^s(t) \qquad \forall (i,j) \in A_0 \cup A_z^+ \quad \forall t \in [0, T-1] \tag{31}$$

$$n_{ij}^f(t) = \sum_{k=\left(t - \frac{L_{ij}}{v_{ij}} + 1\right)}^{T} U_{ij}^s(k) \qquad \forall (i,j) \in A_0 \quad \forall s \in z \quad \forall t \in \left[\frac{L_{ij}}{v_{ij}} - 1, T\right] \tag{32}$$

$$n_{ij}^c(t) = KL_{ij} - \sum_{k=\left(t - \frac{L_{ij}}{w_{ij}} + 1\right)}^{T} \sum_{s \in Z} V_{ij}^s(k) \qquad \forall (i,j) \in A_0 \quad \forall s \in z \tag{33}$$

$$\forall t \in [\frac{\mathrm{L_{ij}}}{\mathrm{w_{ij}}} - 1, T]$$

$$n_{\mathrm{ij}}^{\mathrm{f}}(\mathrm{t}) \leq n_{\mathrm{ij}}(\mathrm{t}) \qquad \forall (i,j) \in A_0 \quad \forall t \in [0,T] \qquad (34)$$

$$n_{\mathrm{ij}}(\mathrm{t}) \leq n_{\mathrm{ij}}^{\mathrm{c}}(\mathrm{t}) \qquad \forall (i,j) \in A_0 \quad \forall t \in [0,T] \qquad (35)$$

$$U_{jk}^{s}(t) = \sum_{(i,j)\epsilon\Gamma_j^-} F_{\mathrm{ijk}}^{\mathrm{s}}(\mathrm{t}) \qquad \forall (i,j) \in A_0 \cup A_z^- \quad \forall (j,k) \in A_0 \cup A_z^+ \quad \forall s \in z \quad \forall t \in [0,T] \qquad (36)$$

$$V_{ij}^{s}(t) = \sum_{(j,k)\epsilon\Gamma_j^+} F_{\mathrm{ijk}}^{\mathrm{s}}(\mathrm{t}) \qquad \forall (i,j) \in A_0 \cup A_z^+ \quad \forall (j,k) \in A_0 \cup A_z^- \quad \forall s \in z \quad \forall t \in [0,T] \qquad (37)$$

$$F_{\mathrm{ijk}}^{\mathrm{s}}(\mathrm{t}) \geq 0 \qquad \forall (i,j) \in A \quad \forall (j,k) \in A \quad \forall s \in z \quad \forall t \in [0,T] \qquad (38)$$

$$p_{\mathrm{j}}(\mathrm{t}+1) = p_{\mathrm{j}}(\mathrm{t}) + \sum_{(m,\mathrm{j})\in\Gamma_j^-} V_{mj}^{j}(t) - \sum_{(j,k)\epsilon\Gamma_j^+} \sum_{s\epsilon Z} y_{jk}^{s}(t) \qquad \forall j \in z \quad \forall t \in [0, T-1] \qquad (39)$$

$$F_{\mathrm{ijk}}^{\mathrm{s}}(\mathrm{t}) = 0 \qquad \forall (i,j) \in A_0 \cup A_z^+ \quad \forall (j,k) \in \Gamma_j^+ \quad \forall s \in z \qquad (40)$$

$$\forall t \in [0, \frac{L_{ij}}{v_{ij}} - 1]$$

$$F^{s}_{ijk}(t) = 0 \qquad \forall (j,k) \in A^{-}_{z},\ \forall (i,j) \in \Gamma^{-}_{j},\ \forall s \neq k,\ \forall t \in [0,T] \tag{41}$$

$$\sum_{(i,j)\epsilon\Gamma^{+}_{i}} \sum_{s\epsilon Z} y^{s}_{ij}(t) \leq p_i(t) \qquad \forall i \in z,\ \forall s \in z,\ \forall t \in [0,T] \tag{42}$$

$$y^{s}_{ij}(t) = U^{s}_{ij}(t) \qquad \forall (i,j) \in A^{+}_{z},\ \forall s \in z,\ \forall t \in [0,T-1] \tag{43}$$

$$V^{s}_{ij}(t+1) = U^{s}_{ij}(t) \qquad \forall (i,j) \in A^{-}_{z} \cup A^{+}_{z},\ \forall s \in z,\ \forall t \in [0,T-1] \tag{44}$$

$$\sum_{i \in Z} p_i(0) = \sum_{i \in Z} p_i(T) \tag{45}$$

$$e^{s}_{r}(t) \leq \omega^{s}_{r}(t) \qquad \forall (r,s) \in z^2,\ \forall t \in [0,T] \tag{46}$$

$$\mathrm{e}_{\mathrm{r}}^{\mathrm{s}}(\mathrm{t}) \leq \sum_{(i,j)\epsilon\Gamma_r^+} y_{rj}^{\mathrm{s}}(t) \qquad \forall (r,s) \in z^2 \quad \forall t \in [0,T] \qquad (47)$$

$$\omega_{\mathrm{r}}^{s}(\mathrm{t}+1) = \omega_{\mathrm{r}}^{\mathrm{s}}(\mathrm{t}) + \mathrm{d}_{\mathrm{r}}^{\mathrm{s}}(\mathrm{t}) - \mathrm{e}_{\mathrm{r}}^{\mathrm{s}}(\mathrm{t}) \qquad \forall (r,s) \in z^2 \quad \forall t \in [0,T-1] \qquad (48)$$

$$\sum_{s\epsilon Z} U_{ij}^{s}(t) \leq MAXFLOW - 1 \qquad \forall (i,j) \in A_0 \quad \forall t \in [0,T] \qquad (49)$$

$$\sum_{s\epsilon Z} V_{ij}^{s}(t) \leq MAXFLOW - 1 \qquad \forall (i,j) \in A_0 \quad \forall t \in [0,T] \qquad (50)$$

$$\omega_{\mathrm{r}}^{\mathrm{s}}(T) = 0 \qquad \forall (r,s) \in z^2 \qquad (51)$$

$$n_{\mathrm{ij}}(0) = 0 \qquad \forall (i,j) \in A \qquad (52)$$

$$n_{\mathrm{ij}}(\mathrm{T}) = 0 \qquad \forall (i,j) \in A \qquad (53)$$

$$U_{ij}^{s}(0) = 0 \qquad \forall (i,j) \in A \quad \forall s \in z \qquad (54)$$

$$\mathrm{y}_{\mathrm{rj}}^{\mathrm{S}}(\mathrm{t}) \geq 0 \qquad \forall (r,s) \in z^2 \quad \forall (r,j) \in A_z^+ \quad \forall t \in [0,T] \qquad (55)$$

$$\mathrm{e}_{\mathrm{r}}^{\mathrm{s}}(\mathrm{t}) \geq 0 \qquad \forall (r,s) \in z^2 \quad \forall t \in [0,T] \qquad (56)$$

Chapter 3: Computational Experiments

3.1. Test Network

Figure 2 shows the grid network used to conduct experiments on the mathematical formulation developed in Chapter 2. This network is similar to past studies conducted by Levin et al. (Levin, 2017) and Duell et al. (Duell et al., 2016) on DTA formulations.

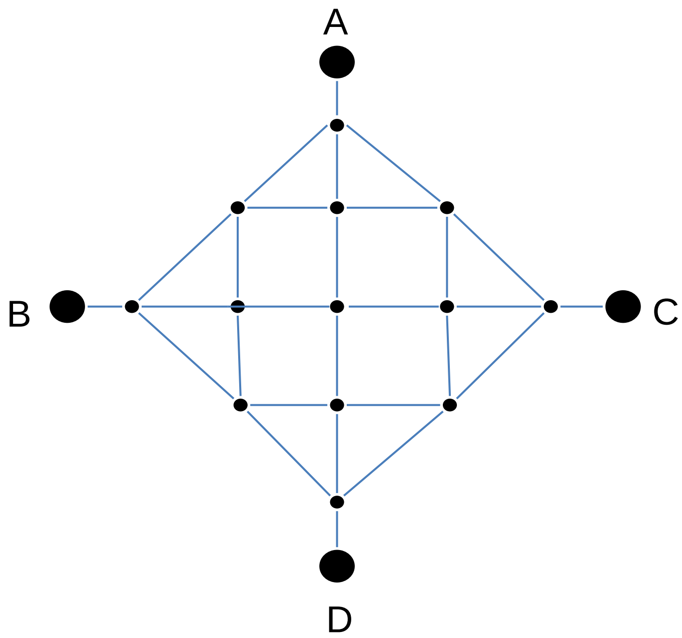

Figure 2: Grid network with 4 centroids

The network consists of four centroids, representing origins and destinations, as well as depots where vehicles are parked. Each centroid is connected to the grid network by a centroid connector, with the grid network consisting of 13 ordinary nodes. All links are bidirectional, and the exogenous parameters used in our simulation are stated in Table 4.

Table 4. Assigned data values

Parameters	Notation	Value [unit]
Capacity of link (i,j)	Q_{ij}	6
Length of link (i,j)	L_{ij}	0.5 [miles]
Free flow speed of link (i,j)	v_{ij}	30 [Mph]
Congested wave speed of link (i, j)	w_{ij}	15 [Mph]
Cost of waiting per customer per unit time	σ_W	1 [$]
Cost in system per vehicle per unit time	σ_V	1 [$]
Number of vehicles parked at depot i at beginning of simulation	$p_i(0)$	3000
Person-trip demand leaving from r towards at time t	$d_{rs}(t)$	**Peak Hour Demand** A - D: 1639 D – A: 139 B – C: 1340 C – B: 290
Time interval used	-	30 [seconds]
Length of analysis period	T	80 [time intervals]

For each scenario, demand was chosen to be over 20 time intervals, with default simulation time set to T= 80, resulting in an additional 60 time intervals to allow all O-D pairs to be satisfied. Each time interval represented 30 seconds, thus accounting for 10 minutes of demand and 30 minutes of simulation. (The objective function maximized the total number of parked vehicles at any point in time, minimizing the travel time taken for each SAV to fulfil its route to satisfy the respective demand at each point in time). Peak hour demand was used to test the model, wherein demands was set to extremely high values such that the number of trips exceeded the capacity of network at 100% demand and were not distributed proportionately such that SAVs had to make repositioning trips to other centroids to satisfy demand at that point. The linear program proposed above produces an optimal solution that balances waiting time and vehicle travel time. Even though waiting time can be reduced by discharging as many vehicles into the network as possible, this results in congestion in the network and thus increases vehicle travel time. Since the model is a linear program, it produces a solution that is not only a local optima but also a global optimum. However, recent research (Shen and Zhang, 2014) has shown that high dimensionality, system-optimal DTA may have numerous solutions in which the optimal objective function value is the same but queue distribution and route choice is different in each solution. As such, path flows are nonunique, reducing the difficulty in finding a system optima solution. Because the problem is formulated such that a system- optimal path is found, any optimal solution is sufficient. Since this formulation is the first of its kind (to this author's knowledge), comparison of results with previous works were not possible. The linear programming formulation was implemented in IBM CPLEX 12.8.0 using an Intel i7-3720QM CPU clocked at 2.6GHz with 16GB RAM.

3.2. Comparison with Static Traffic Assignment Formulation

To assess the performance of this formulation, we modify the above TTM linear programming formulation to obtain a non-system optimal DTA solution, that is the static traffic assignment (STA) solution. This assumes that demand is satisfied as soon as it appears; that is, there is no waiting time for customers at the centroids. This also means that all departing vehicles will be associated with demand present at that point in time. This provides a user optimal solution but not system optimal. Customers' choices are based on their myopic decisions rather than anticipating the traffic condition along the route so as to minimize actual experienced travel time. A similar formulation was done by Ziliaskopoulos et al. (2000), but our formulation differs by implementing TTM instead of CTM with multiple destinations. Therefore, we replace constraints 38, 43 through 46 and 55 through 56 with the following constraint:

$$\sum_{(r,j)\epsilon\Gamma_r^+} y_{rj}^s(t) = \mathrm{d}_\mathrm{r}^\mathrm{s}(\mathrm{t}) \qquad \forall(r,s) \in Z \quad \forall t \in [0,T] \qquad (56)$$

At the time when demand is present, the system chooses the most optimal route choice based on current road conditions; that is, there is no balance between delaying departure, taking into account the congestion that will be resulted from releasing an additional vehicle into the system. As a consequence, paths become congested. This produces the lower bound for

TSTT for each particular demand scenario, presenting a basis for comparison. Furthermore, this formulation also illustrates the case in which only personal vehicles are present on the road, and each traveller determines his/her departure time based on personal preference without concern that his/her decision will affect road conditions; that is, the optimal solution found is a user equilibrium optimal but not system optimal.

3.3. Experiments

To evaluate the performance of the TTM-DSO-DARP linear programming formulation, we varied different important parameters such as demand, fleet size, and number of time intervals to observe its effect on the decision variables. All departing vehicles depend on the number of person-trips, and the optimal solution depends on the trade-off between time spent waiting to depart from the depot and the time spent in travelling (with and without congestion).

The number of vehicles on the road affects the level of congestion; this in turn relates to the amount of demand of travelers going from centroid to centroid. As such, using the TTM-DSO-DARP model, we study the impact of different levels of demand on the service levels and time taken to travel. Experiments ranging from 10% of total demand to 100% of demand were used. The number of SAVs in the entire system was set to 1,720, half the number of trips for the 100% demand scenario, and the start of the simulation was done with all SAVs split equally between the four centroids.

3.3.1. Effects of Change in Demand on Utilization Rate

First, in Figure 3, the straight line indicates the total number of SAVs that are available, which was fixed at 1720 for all cases. All available SAVs were not used even though demand exceeded the total amount of SAVs available. Figure 3 shows that as demand increased, the number of SAVs used increased less than proportionately. It can be deduced that there is a limit on the number of SAVs that can be used based on the traffic network, and it was suboptimal to utilize all SAVs available due to an increase in congestion, in turn increasing travel time by travelers. Thus, this model can be used in the planner's perspective, to be able to identify the optimal number of SAVs such that there is no oversupply of vehicles leading to increased cost.

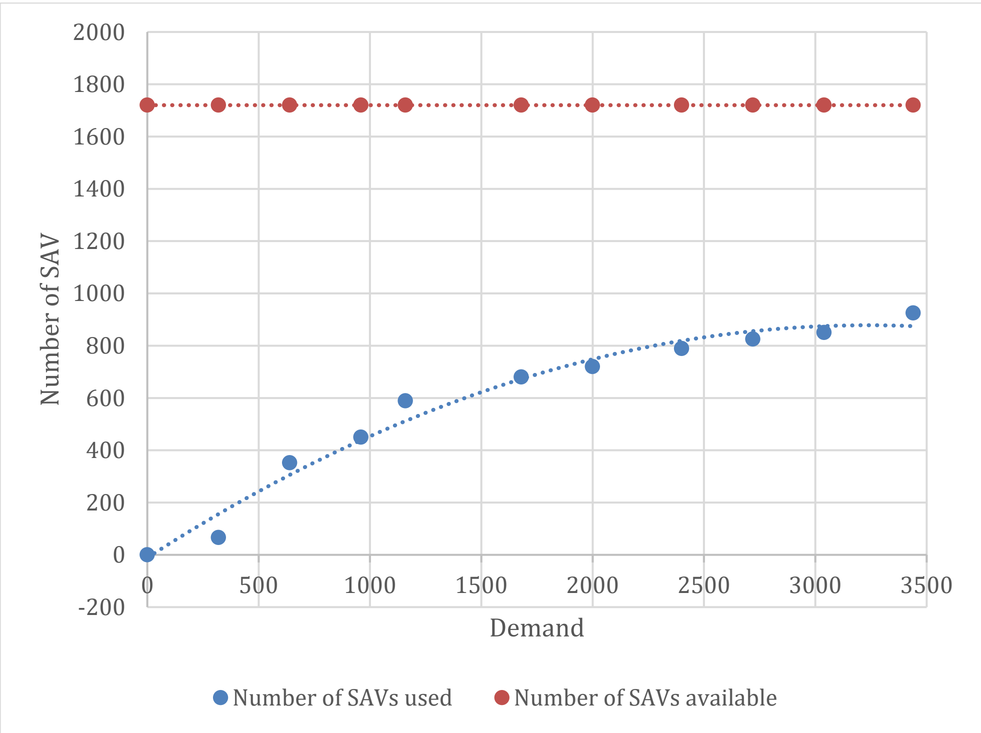

Figure 3. Effects of demand on number of SAVs used

By dividing the demand by fleet size, we can obtain the average demand fulfilled per SAV. The result suggests that results in the previous literature (Fagnant and Kockelman, 2014), wherein each SAV can replace 11 personal vehicles may not be optimal due to other considerations such as the demand and the road network. Increasing the number of SAVs in the network may do more harm than good, considering the negative impacts that may arise due to congestion.

3.3.2. Effects of Change in Demand on VMT

Average VMT per passenger increased with demand as seen in Figure 4 for the case of the TTM-DSO-DARP model. For the STA model, average vehicle miles per commuter were constant as each dispatch was based on the user optimal route choice, regardless of demand at any point in time. Vehicle miles traveled per passenger increased as vehicles took less direct routes where overall traveling cost was significantly less than if extended traveling time were spent in traffic congestion. However, vehicle miles traveled increased initially but decreased after a certain point of demand. This could be due to demand increasing leading to immediate trip chaining by the SAVs, resulting in a less need to carry out repositioning trips while satisfying demand on short notice.

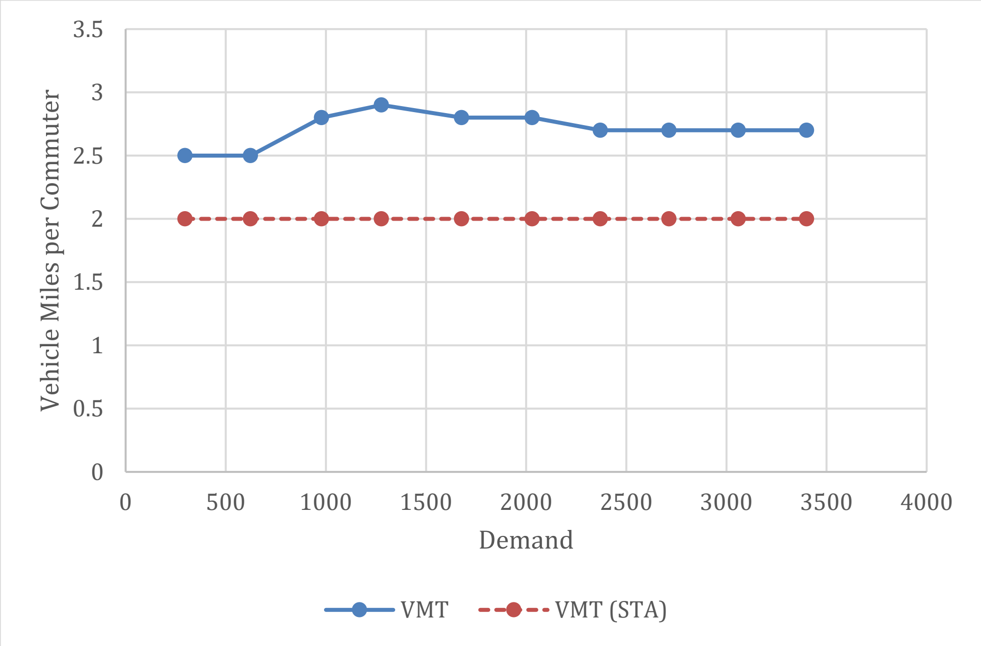

Figure 4. Effects of demand on VMT

3.3.3. Effects of Change in Demand on Total Travel Time

Figure 5 shows the change in average TSTT, average waiting time, and average vehicle travel time caused by the increase in demand. Most of the change in average TSTT was due to the change in waiting time, with average waiting time increasing by up to 3 time intervals from 30% demand to 100% demand. We can deduce that the average waiting time and average vehicle travel time increased due to empty repositioning trips to tackle high demand at the respective nodes. Increase in average vehicle travel times when demand increased could be attributed to the greater density of cars on the roads per unit time, together with an increase in less direct routes to avoid congestion. Also, as only 20 time intervals of demand were considered, the change in average vehicle times could be larger if longer periods of demand were used instead.

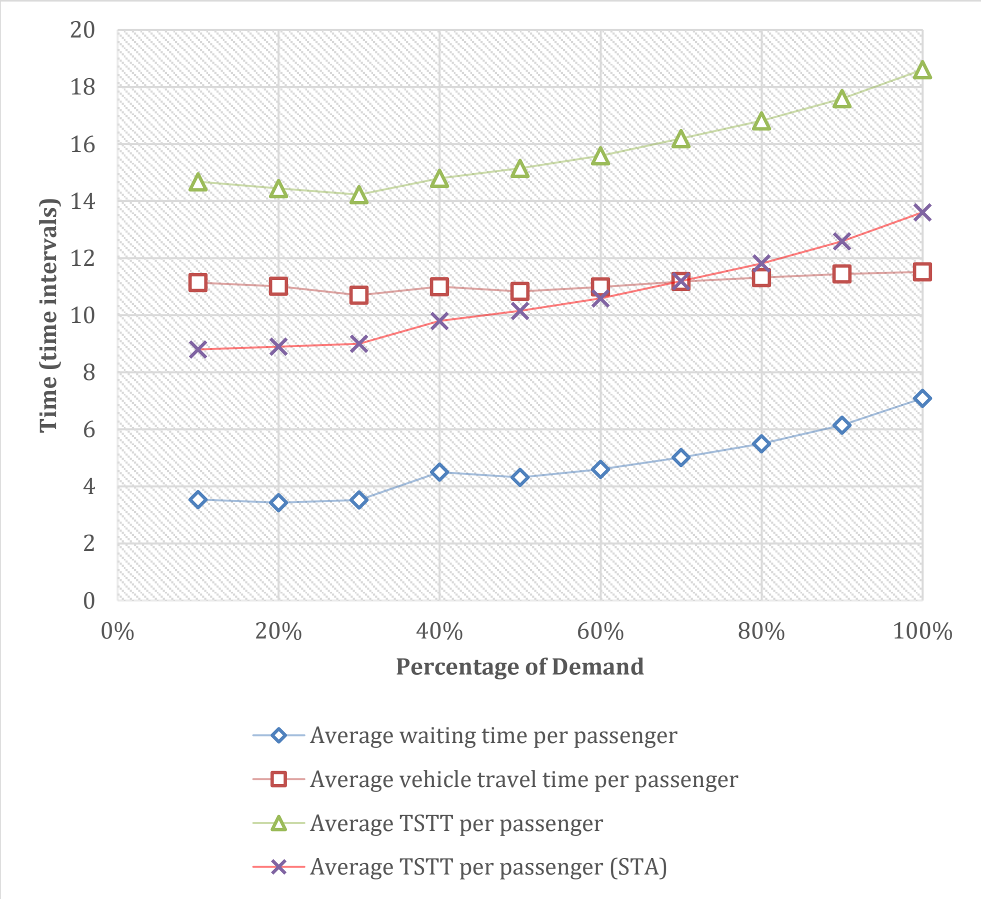

Figure 5. Effects of demand on waiting time, vehicle travel time and total travel time

Furthermore, as we are investigating the DSO solution, the model is sensitive to the exogenous parameters defined, that is, the length of each link or the capacity of the link. If the flow of traffic from a high-capacity link leading to a low-capacity link is increased, congestion will correspondingly increase at the same time, and a bottleneck will occur. Increasing demand produces the same effect as decreasing the capacity of the entire network. From the TSTT utilizing the STA formulation where no waiting was considered, each passenger required 9.5 minutes of traveling time at low demand, but traveling time increased exponentially as demand increased. At low demand, TSTT of the STA formulation was lower than the vehicle travel time per passenger of this thesis's model, but at 70%, it exceeded that of this model. This was because departing vehicles of the STA formulation only considered each of their instantaneous travel times but not foresee the traffic conditions in the future. Even though with low demand, the STA formulation performed better with lower TSTT, but in the case of a morning commute/last mile service, where demand is expected to be many times greater, the STA formulation will perform badly. Even though overall TSTT of this thesis's model is larger than that of the STA formulation, most of the time is spent waiting. This implies that customers do not have to waste time stuck in the vehicle commuting, but instead while waiting, customers have the ability to continue doing other things, maximizing their time instead of being constrained in a vehicle.

3.3.4. Effects of Change in Fleet Size on Total Travel Time

We further investigated the effect of fleet size on the average waiting time per passenger, the average vehicle travel time per passenger, and the average TSTT per passenger. Demand was kept at a constant level of 100%, utilizing the high-demand scenario. Figure 6 reflects the result. At a fleet size of 800, the average TSTT per passenger was 15 minutes. The average vehicle travel time per passenger stayed relatively constant even though fleet size decreased, but average waiting time per passenger—thus average TSTT per passenger—increased by a power function as fleet size decreased. This is due to the increase in repositioning trips that are required to satisfy demand at the respective nodes caused by a lack of supply of SAVs.

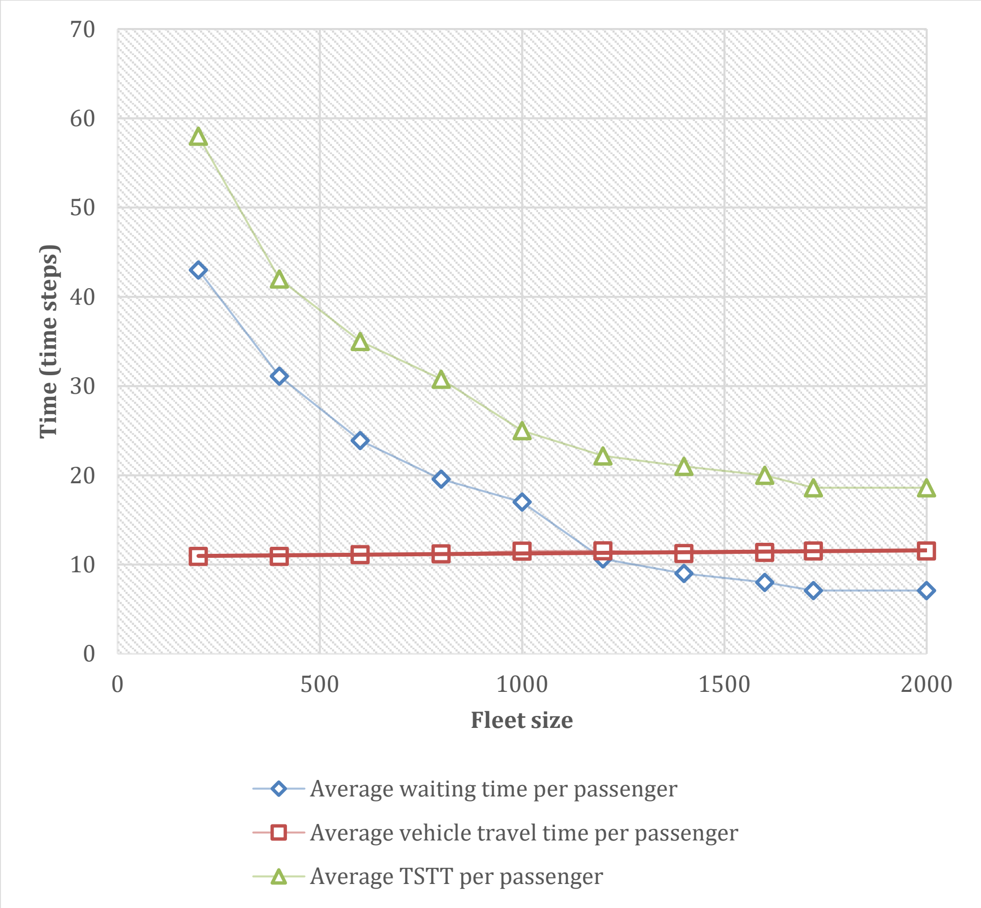

Figure 6. Effects of fleet size on average waiting time, vehicle travel time and TSTT

3.3.5. Effects of Change in Time Intervals on Computational Time and Complexity

Based on the test network, we compare three aspects of the complexity of the TTM model: (1) the number of constraints, (2) the number of variables, and (3) the computational time. We ignore the impact of space, that is, the number of nodes and centroids, by fixing it at a constant and explore the influence of the time domain on the model. High peak distribution of demand over 20 time intervals with the length of analysis period set to 80 time intervals was used.

In the experiments for the peak hour demand at default settings, 100% demand is satisfied within 832 seconds in the grid. Varying the time intervals changes the time domain for each run, leading to changes in computational time. This is shown in Table 5, noting the impact of different time intervals with respect to TTM.

Table 5. Effects of number of time intervals on computational time and complexity

Time intervals	Preparatio n Time	Run Time	Number of Constraints	Number of Variables
60	0.36 secs	3 mins 41 secs	134,390	103,213
70	0.47 secs	9 mins 14 secs	156,230	120,133
80	0.53 secs	13 mins 52 secs	178,069	137,053
90	0.62 secs	20 mins 21 secs	199,910	153,973
100	0.7 secs	25 mins 26 secs	221,750	170,893

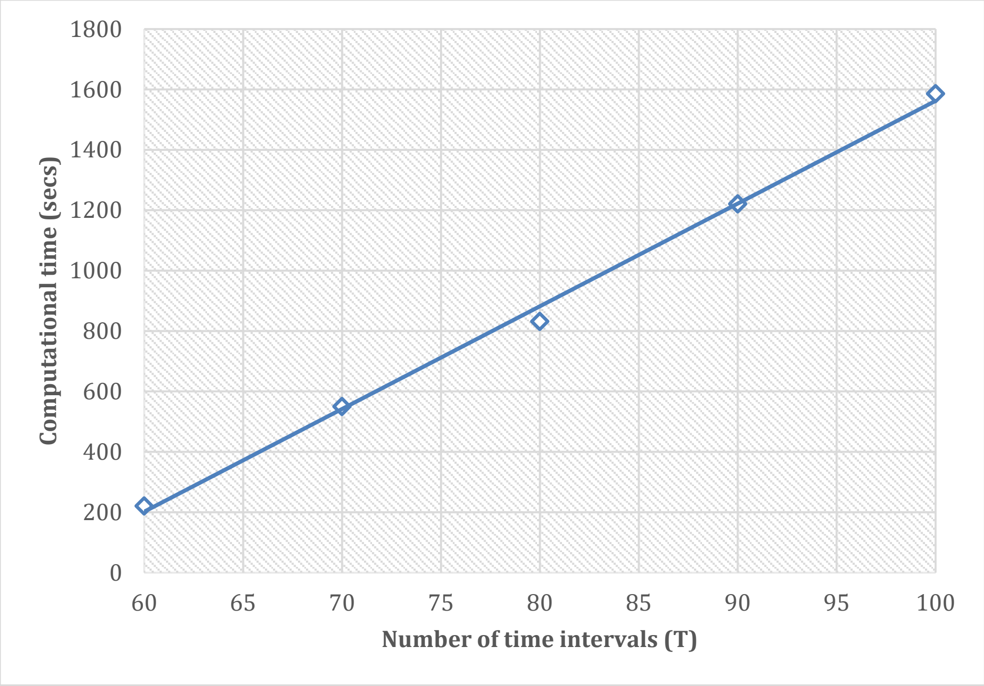

Figure 7. Effects of number of time intervals (T) on computational time

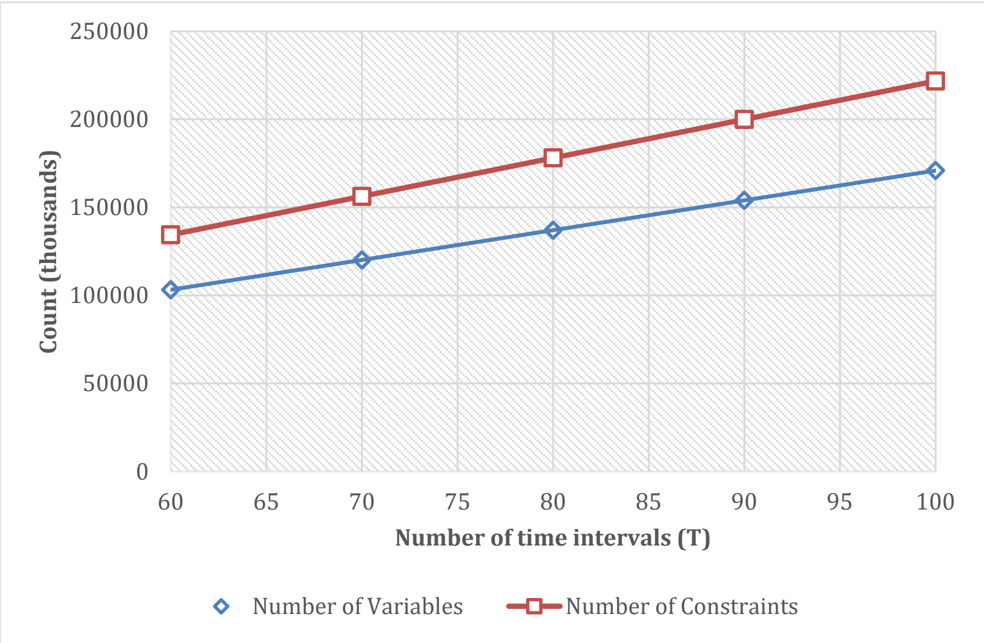

Figure 8. Effects of number of time intervals (T) on number of variables and constraints

The amount of time taken to create the grid network of TTM, including processing of data from the input file and generating the abstract models with constraints, variables, and parameters, increases linearly as number of time intervals increases. This includes processing the data file and .xlsx spreadsheet and defining a model instance with constraints, objective function, variables, and parameters. Figure 7 depicts the linear relationship between increasing the number of time intervals in the model with the computational time, while similarly Figure 8 shows that the number of variables and number of constraints also increases linearly with respect to time. This is not surprising as the model is constructed in such a way that each link (i, j) records the number of vehicles based on the inflow and outflow per unit time. Inflow and outflow of each link is further segregated into its respective final destinations, denoting each turning proportion and then defining each choice of route. Therefore, not only is each link defined by its link index; it is defined by the destinations in the model for each time interval, that is, the centroids, resulting in a linear increase as seen from the graph. As such, the user has to take these factors into consideration.

Chapter 4: Conclusions

This thesis tackled two main issues commonly found in SAV routing problems: (1) how to represent traffic flow within the system grid and (2) how to satisfy demand with vehicle supply. This novel linear programming formulation presented a way to incorporate TTM together with DARP to satisfy the demand of passengers' O-D pairs while factoring in the problem of congestion in the traffic flow if there are too many vehicles in the system. This anticipates that sooner or later, smart city proposals will aim to replace privately owned vehicles with electric SAVs to tackle the problems of environmental pollution and increasing population. This formulation is able to depict the queue lengths evolution through time and space within each link, providing a useful tool for transportation planners who will optimize autonomous routing problems in the future. Morning commute/last-mile demand is assumed in this study, while dynamic demand can be further used to expand the model in the future.

Experiments were carried out and solved to optimality based on several scenarios to test the feasibility and performance of the model. Distribution of demand, fleet size, and total time period length of simulation runs were some of the factors that were found to make a significant impact in the model. The repositioning trips that were necessary to compensate for the lack of vehicles at any centroid contributed to the increase in waiting time of passengers. In this way, it is important that transportation planners identify the ground situation and make necessary arrangements before the peak hour start, such as moving SAVs to hot zones beforehand. This will reduce each customer's travel time, improving efficiency and customer satisfaction levels.

As this linear program is a large-scale linear programming problem, we have carried out case studies to illustrate the performance of this model. Computation time increases linearly as the size of the network increases. Complexity of the model including the number of constraints and the number of variables increases linearly with time intervals. TTM is able to depict the conditions within links closer to real-life scenarios and thus provide us with a more optimal solution.

However, there are limitations to this formulation. The usage of TTM in depicting traffic flow implies that an assumption is made about the location of the congestion; that is, it occurs only at the exit of the link, propagating upstream. The link is assumed to have uniform capacity throughout, such that given the case where there is a change in the number of lanes of the road, the link has to be subdivided into two links before TTM can be used to model traffic flow. Special care has to be taken to ensure that actual scenarios are modeled accurately. Furthermore, TTM is not able to depict the case of a double shockwave, wherein a temporary bottleneck occurs between the upstream node and downstream node; that is, the regimes alternate via the order—free flow, congested, and free flow within a single link. Along the same line, the model is not able to deal with multiclass vehicles and moving bottlenecks. As such, unexpected phenomena such as random accident occurrences are unable to be captured in this model.

Nevertheless, not only does this formulation provide a parsimonious depiction of traffic dynamics; it also incorporates DARP to provide a shared autonomous transportation service by satisfying multiple O-D pairs. Although the proposed framework applies to only an O-D network without ridesharing, it plays an important role in future SAV network implementation and planning and paves the way for future work such as developing heuristics and further development involving larger-scale networks and traffic control abilities for computing a dynamic user equilibrium.

Acknowledgements

The author gratefully acknowledges the guidance and care received under Professor Ilkyeong Moon of the Industrial Engineering Department of Seoul National University. In addition, the author is thankful for the efforts and advice given by the labmates at the Supply Chain Management lab, notably Gwang Kim and Dongwook Kim.

The author would like to express his gratitude to his parents and loved ones for their unfailing support and continuous encouragement. In addition, the author is grateful to be able to accomplish this through the financial support provided by the Korean Government Scholarship Programme.

Bibliography

[1] Administration, F. H. (2009). National Household Travel Survey.

[2] Agatz, N., Erera, A., Savelsbergh, M., & Wang, X. (2012). Optimization for dynamic ride-sharing: A review. *European Journal of Operational Research, 223*(2), 295-303.

[3] Almeida, C. G. H., & Arem, B. v. (2016). Solving the User Optimum Privately Owned Automated Vehicles Assignment Problem (UO-POAVAP): A model to explore the impacts of self-driving vehicles on urban mobility. *Transportation Research Part B: Methodological, 87*, 64-88.

[4] Aubin, J.-P., Bayen, A. M., & Saint-Pierre, P. (2008). Dirichlet problems for some Hamilton–Jacobi equations with inequality constraints. *SIAM Journal on Control and Optimization, 47*(5), 2348-2380.

[5] Bae, H., & Moon, I. (2016). Multi-depot vehicle routing problem with time windows considering delivery and installation vehicles. *Applied Mathematical Modelling, 40*(13), 6536-6549. doi:https://doi.org/10.1016/j.apm.2016.01.059

[6] Balijepalli, Ngoduy, D., & Watling, D. (2014). The two-regime transmission model for network loading in dynamic traffic assignment problems. *Transportmetrica A: Transport Science, 10*(7), 563-584.

[7] Bornd, R., Gr, M., & Kuttner, F. K. C. (1997). Telebus Berlin: Vehicle Scheduling in a Dial-a-Ride System.

[8] Burns, L. D., Jordan, W. C., & Scarborough, B. A. (2013). Transforming personal mobility. *The Earth Institute, 431*, 432.

[9] Carey, M., Humphreys, P., McHugh, M., & McIvor, R. (2014). Extending travel-time based models for dynamic network loading

and assignment, to achieve adherence to first-in-first-out and link capacities. *Transportation Research Part B: Methodological, 65*, 90-104.

[10] Chang, T.-H., & Lai, I.-S. (1997). Analysis of characteristics of mixed traffic flow of autopilot vehicles and manual vehicles. *5*(6), 333-348.

[11] Chen, T. D., Kockelman, K. M., & Hanna, J. P. (2016). Operations of a shared, autonomous, electric vehicle fleet: Implications of vehicle & charging infrastructure decisions. *Transportation Research Part A: Policy and Practice, 94*, 243-254.

[12] Chiu, Y.-C., Bottom, J., Mahut, M., Paz, A., Balakrishna, R., Waller, T., & Hicks, J. (2011). Dynamic traffic assignment: A primer. *Transportation Research E-Circular*(E-C153).

[13] Cordeau, J.-F., & Laporte, G. (2003). A tabu search heuristic for the static multi-vehicle dial-a-ride problem. *Transportation Research Part B: Methodological, 37*(6), 579-594.

[14] Cordeau, J.-F., & Laporte, G. (2007). The dial-a-ride problem: models and algorithms. *Annals of operations research, 153*(1), 29-46.

[15] Daganzo, C. F. (1994). The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transportation Research Part B: Methodological, 28*(4), 269-287.

[16] Desrosiers, J., Dumas, Y., Solomon, M. M., & Soumis, F. (1995). Time constrained routing and scheduling. *Handbooks in operations research and management science, 8*, 35-139.

[17] Dong, G., Tang, J., Lai, K. K., & Kong, Y. (2009). An exact algorithm for vehicle routing and scheduling problem of free pickup and delivery service in flight ticket sales companies based on set-

partitioning model. *Journal of Intelligent Manufacturing, 22*(5), 789-799. doi:10.1007/s10845-009-0311-9

[18] Dubai's Autonomous Transportation Stratedgy. (2018).

[19] Duell, M., Levin, M. W., Boyles, S. D., & Waller, S. T. (2016). Impact of autonomous vehicles on traffic management: Case of dynamic lane reversal. *Transportation Research Record: Journal of the Transportation Research Board*(2567), 87-94.

[20] Fagnant, D. J., & Kockelman, K. M. (2014). The travel and environmental implications of shared autonomous vehicles, using agent-based model scenarios. *Transportation research part C: emerging technologies, 40*, 1-13.

[21] Fagnant, D. J., & Kockelman, K. M. (2015). *Dynamic ride-sharing and optimal fleet sizing for a system of shared autonomous vehicles.* Transportation Research Board 94th Annual Meeting, 15-1962.

[22] Fagnant, D. J., & Kockelman, K. M. (2018). Dynamic ride-sharing and fleet sizing for a system of shared autonomous vehicles in Austin, Texas. *Transportation, 45*(1), 143-158.

[23] Han, K., Piccoli, B., & Szeto, W. (2016). Continuous-time link-based kinematic wave model: formulation, solution existence, and well-posedness. *Transportmetrica B: Transport Dynamics, 4*(3), 187-222.

[24] Ho, S. C., Szeto, W., Kuo, Y.-H., Leung, J. M., Petering, M., & Tou, T. W. (2018). A survey of dial-a-ride problems: Literature review and recent developments. *Transportation Research Part B: Methodological*.

[25] Jaw, J.-J., Odoni, A. R., Psaraftis, H. N., & Wilson, N. H. (1986). A heuristic algorithm for the multi-vehicle advance request dial-a-ride problem with time windows. *Transportation Research Part B: Methodological, 20*(3), 243-257.

[26] Kok, A. L., Hans, E. W., & Schutten, J. M. (2012). Vehicle routing under time-dependent travel times: the impact of congestion avoidance. *Computers operations research, 39*(5), 910-918.

[27] Krueger, R., Rashidi, T. H., & Rose, J. M. (2016). Preferences for shared autonomous vehicles. *Transportation research part C: emerging technologies, 69*, 343-355.

[28] Kumar, S. N. (2012). A Survey on the Vehicle Routing Problem and Its Variants. *Intelligent Information Management, 04*(03), 66-74. doi:10.4236/iim.2012.43010

[29] Levin, M. W. (2017). Congestion-aware system optimal route choice for shared autonomous vehicles. *Transportation research part C: emerging technologies, 82*, 229-247.

[30] Lighthill, M. J., & Whitham, G. B. (1955). On kinematic waves II. A theory of traffic flow on long crowded roads. *Proc. R. Soc. Lond. A, 229*(1178), 317-345.

[31] Liu, H., Claudel, C., & Machemehl, R. B. (2018). A Stochastic Formulation of the Optimal Boundary Control Problem Involving the Lighthill Whitham Richards Model. *IFAC-PapersOnLine, 51*(9), 337-342.

[32] Madsen, O. B. G., Ravn, H. F., & Rygaard, J. M. (1995). A heuristic algorithm for a dial-a-ride problem with time windows, multiple capacities, and multiple objectives. *Annals of Operations Research, 60*, 193-208.

[33] Morris, D. Z. (2016). Today's Cars Are Parked 95% of the Time. Retrieved from http://fortune.com/2016/03/13/cars-parked-95-percent-of-time/

[34] Newell, G. F. (1993). A simplified theory of kinematic waves in highway traffic, part I: General theory. *Transportation Research Part B: Methodological, 27*(4), 281-287.

[35] Ngoduy, D., Hoang, N., Vu, H., & Watling, D. (2016). Optimal queue placement in dynamic system optimum solutions for single origin-destination traffic networks. *Transportation Research Part B: Methodological, 92*, 148-169.

[36] Parragh, S. N., Cordeau, J.-F., Doerner, K. F., & Hartl, R. F. (2012). Models and algorithms for the heterogeneous dial-a-ride problem with driver-related constraints. *OR Spectrum, 34*(3), 593-633. doi:10.1007/s00291-010-0229-9

[37] Richards, P. I. (1956). Shock waves on the highway. *Operations research, 4*(1), 42-51.

[38] Rincon-Garcia, N., Waterson, B., Cherrett, T. J., & Salazar-Arrieta, F. (2018). A metaheuristic for the time-dependent vehicle routing problem considering driving hours regulations–An application in city logistics. *Transportation Research Part A: Policy Practice*.

[39] Ropke, S. (2005). *Heuristic and exact algorithms for vehicle routing problems.* University of Copenhagen,

[40] Sayarshad, H. R., & Chow, J. Y. (2015). A scalable non-myopic dynamic dial-a-ride and pricing problem. *Transportation Research Part B: Methodological, 81*, 539-554.

[41] Shaheen, S., Cohen, Adam. (2013). Innovative Mobility Carsharing Outlook.

[42] Shen, W., Nie, Y., & Zhang, H. M. (2007). Dynamic Network Simplex Method for Designing Emergency Evacuation Plans. *Transportation Research Record, 2022*(1), 83-93. doi:10.3141/2022-10

[43] Shen, W., & Zhang, H. (2014). System optimal dynamic traffic assignment: Properties and solution procedures in the case of a many-to-one network. *Transportation Research Part B: Methodological, 65*, 1-17.

[44] Singapore, S. N. (2019). Transforming Singapore. Retrieved from https://www.smartnation.sg/what-is-smart-nation/initiatives/Transport

[45] Spieser, K., Treleaven, K., Zhang, R., Frazzoli, E., Morton, D., & Pavone, M. (2014). Toward a systematic approach to the design and evaluation of automated mobility-on-demand systems: A case study in Singapore. In *Road vehicle automation* (pp. 229-245): Springer.

[46] Toth, P., & Vigo, D. (1996). Fast local search algorithms for the handicapped persons transportation problem. In *Meta-Heuristics* (pp. 677-690): Springer.

[47] Wageningen-Kessels, F., van Lint, H., Vuik, K., & Hoogendoorn, S. (2015). Genealogy of traffic flow models. *EURO Journal on Transportation and Logistics, 4*(4), 445-473. doi:10.1007/s13676-014-0045-5

[48] Xiang, Z., Chu, C., & Chen, H. (2008). The study of a dynamic dial-a-ride problem under time-dependent and stochastic environments. *European Journal of Operational Research, 185*(2), 534-551.

[49] Yperman, I. (2007). The link transmission model for dynamic network loading.

[50] Ziliaskopoulos, A. K. (2000). A linear programming model for the single destination system optimum dynamic traffic assignment problem. *Transportation science, 34*(1), 37-49.

국문초록

공유 자율주행 차량 서비스를 활용한 최적 교통 경로 문제

본 연구는 네트워크 내 교통 흐름 혼잡을 고려하는 공유 자율주행 차량 경로문제(Shared Autonomous Vehicle Routing Problem)를 다루고 있다. 이 문제는 향후 자율주행차가 개인 소유의 차를 대체할 것이라는 관점에서 시작되었다. 효율적인 공유 서비스를 제공하기 위해, 기존의 다이얼 어 라이드(Dial-A-Ride) 문제에 출발지와 도착지 간의 수요를 만족하도록 하는 교통 흐름 모델을 결합해 최적의 교통 할당 문제를 제안한다. 거시적인 교통 흐름은 네트워크 각 링크에 유입 및 유출을 활용한 이중 체제 전송(Two Regime Transmission) 모델을 활용한다. 혼잡으로 인한 제약들로 인해 수요 및 차량 크기와 관계없이 최적의 해에서는 최대 차량 수가 활용되고 있음을 보여준다. 또한, 피크 교통 시간대에서는 수요에 따른 최적의 교통 할당과 차량 크기를 얻어 교통 혼잡에 활용할 수 있다.

주요어: Two Regime Transmission Model, DARP, Shared Autonomous Vehicles, Morning Commute, Last Mile

학번: 2017-20583

Appendix

i) IBM CPLEX ILOG Linear Programming Code

```
/*********************************************
 * OPL 12.8.0.0 Model
 * Author: blzm
 *********************************************/
/*********************************************************************
*****
 * DATA DECLARATIONS
*********************************************************************
****/
range NODES = 1..17;
{int} CENTROIDS = {1,2,3,4};
int T = 50;
int A[NODES][NODES] =...;// !binary code of ALL present arcs
int A0[NODES][NODES] =...;// ofal integer !binary code of arcs without
centroids
int AZP[NODES][NODES] =...;// of integer !binary code of arcs with
centroids to normal
int AZN[NODES][NODES] =...;// of integer !binary code of arcs with normal
to centroids
float L[NODES,NODES] =...;//length
float Q[NODES][NODES] =...;//capacity
float W[NODES][NODES] =...;//congested wavespeed
float K = ...;//jam density
tuple fromtotime {
       key int fromcentroid;
       key int tocentroid;
       key int time;
}
float sigmaV = 1;
float sigmaW = 1;
{fromtotime} FromToTime with fromcentroid in CENTROIDS, tocentroid in
CENTROIDS =...;
int DD[FromToTime]=...;
int TimeforWholeLinkFree = ftoi(ceil(L[i,j]/V[i,j]));
int TimeforWholeLinkCong = ftoi(ceil(L[i,j]/W[i,j]));

float MAXFLOW = (K*0.25*0.125)/(0.25+0.125);
/*********************************************************************
*****
 * DECISION VARIABLES

*********************************************************************
****/
dvar float+ NO[NODES,NODES, 0..T];
dvar float+ NC[NODES, NODES, 0..T];
dvar float+ NF[NODES, NODES,0..T];
dvar float+ NOS[NODES, NODES, CENTROIDS, 0..T];
dvar float+ y[CENTROIDS][NODES][CENTROIDS][0..T];
dvar float+ P[CENTROIDS][0..T]; //number of vehicles at each centroid
dvar float+ F[NODES,NODES,NODES, CENTROIDS, 0..T];
dvar float+ E[CENTROIDS,CENTROIDS, 0..T];//number of travelers departing r
to s
```

```
dvar float+ OM[CENTROIDS, CENTROIDS, 0..T];//number of waiting travellers
dvar float+ U[NODES, NODES, CENTROIDS, 0..T];
dvar float+ V[NODES, NODES, CENTROIDS, 0..T];

dexpr float objective = sigmaV*sum(i in CENTROIDS, t in 0..T) (P[i,0]-
P[i,t]) + sum( r, s in CENTROIDS, t in 0..T) sigmaW*OM[r,s,t];
minimize objective;

subject to {

              C1:
                     forall(i,j in NODES, s in CENTROIDS, t in
0..T:A0[i,j]==1)
                     NOS[i,j,s,t] == sum( k in 0..t) (U[i,j,s,k]-
V[i,j,s,k]);
              C2:
                     forall(i,j in NODES, t in 0..T:A0[i,j]==1)
                            NO[i,j,t] == sum(s in CENTROIDS)
NOS[i,j,s,t];
              C3:
                     forall(i,j in NODES, t in TimeforWholeLinkCong-
1..T:A0[i,j]==1)
                       NC[i,j,t] ==97 - sum(s in CENTROIDS, m in t-
TimeforWholeLinkCong+1..t) V[i,j,s,m];
              C4:
                     forall(i,j in NODES, t in TimeforWholeLinkFree-
1..T:A0[i,j]==1)
                       NF[i,j,t] == sum(s in CENTROIDS, m in t-
TimeforWholeLinkFree+1..t)  U[i,j,s,m];
              C5:
                     forall(i,j in NODES, t in 0..T:A0[i,j]==1)
                       NO[i,j,t] <= NC[i,j,t];
              C6:
                     forall(i,j in NODES, t in 0..T:A0[i,j]==1)
                       NF[i,j,t] <=NO[i,j,t];
              C7:
                     forall(s in CENTROIDS, j in NODES, t in 0..T)
                       sum(i in NODES:A0[i,j]==1||AZP[i,j]==1) V[i,j,s,t]
== sum(k in NODES:A0[j,k]==1||AZN[j,k]==1) U[j,k,s,t];
              C8:
                     forall(j,k in NODES,s in CENTROIDS, t in
0..T:A0[j,k]==1||AZN[j,k]==1)
                       U[j,k,s,t] ==  sum(i in
NODES:A0[i,j]==1||AZP[i,j]==1)F[i,j,k,s,t];
              C9:
                     forall(i,j in NODES, s in CENTROIDS, t in
0..T:A0[i,j]==1||AZP[i,j]==1)
                       V[i,j,s,t] == sum(k in
NODES:AZN[j,k]==1||A0[j,k]==1)F[i,j,k,s,t];
              C10:
                     forall(i,j,k in NODES, s in CENTROIDS, t in
0..T:A[i,j]==1 && A[j,k]==1)
                       F[i,j,k,s,t] >= 0;
              C11:
                     forall(i,j,k in NODES, s in CENTROIDS, t in 0..
TimeforWholeLinkFree-1:A0[i,j]==1 && AZP[i,j]==1 && A[j,k]==1)
                       F[i,j,k,s,t] == 0;
```

```
              C12:
              forall(i in NODES, j in CENTROIDS, t in 0..T-1:AZN[i,j]==1)

                       P[j, t+1] == P[j,t] + V[i,j,j,t] - sum(k in
NODES,s in CENTROIDS: AZP[j,k]==1)y[j,k,s,t];
              C13:
              forall(i, j, k in NODES, s in CENTROIDS, t in
0..T:A[i,j]==1 && AZN[j,k]==1 && s!=k)
                    F[i,j,k,s,t] == 0;
              C14:
                     forall(t in 0..T, i in CENTROIDS)
                sum(j in NODES: (AZP[i,j]==1))( sum(s in CENTROIDS)
y[i,j,s,t]) <= P[i,t];
              C15:
              forall(i,s in CENTROIDS,j in NODES, t in 0..T-1:
(AZP[i,j]==1))
                        y[i,j,s,t] == U[i,j,s,t];
              C16:
              forall(i, j in NODES, s in CENTROIDS, t in 0..T-1:
AZN[i,j]==1||AZP[i,j]==1)
                 V[i,j,s,t+1] == U[i,j,s,t];
              C17:
           sum(i in CENTROIDS) P[i,0] == sum(k in CENTROIDS) P[k,T];
              C18:
                     forall(r,s in CENTROIDS, t in 0..T)
                        E[r,s,t] <= OM[r,s,t];
              C19:
                     forall(r,s in CENTROIDS, t in 0..T)
                        E[r,s,t] <=   sum(j in NODES: (AZP[r,j]==1))
y[r,j,s,t];
              C20:
                     forall(r,s in CENTROIDS, t in 0..T-1: r!=s)
                        OM[r,s,t+1] == OM[r,s,t] + DD[<r,s,t>] -
E[r,s,t];
              C21:
                     forall(i, j in NODES, t in 0..T:((A0[i,j]==1)))
                        sum(s in CENTROIDS) V[i,j,s,t] <= MAXFLOW-1;
              C22:
                     forall(i, j in NODES, t in 0..T:((A0[i,j]==1)))
                 sum(s in CENTROIDS) U[i,j,s,t] <= MAXFLOW-1;
              C23:
                     forall(r,s in CENTROIDS)
                 OM[r,s, T] == 0;
              C24: //terminating condition
                     forall( i,j in NODES:A[i,j]==1)
                        NO[i,j,T]==0;
              C25: //terminating condition
                     forall( i,j in NODES, s in CENTROIDS:A[i,j]==1)
                        U[i,j,s,T]==0;
              C26:
                     forall(i,j in NODES, s in CENTROIDS:A[i,j]==1)
                        V[i,j,s,T]==0;
              C27:
                     forall(r,s in CENTROIDS ,j in NODES, t in 0..T:
(AZP[r,j]==1))
                        y[r,j,s,t]>=0;
              C28:
                     forall(r,s in CENTROIDS, t in 0..T)
```

```
                    E[r,s,t]>=0;
          C29:
                 forall(i,j in NODES, s in CENTROIDS, t in
0..TimeforWholeLinkFree-1 :(A[i,j]==1))
                    V[i,j,s,t] == 0;
          C30:
                 forall(i,j in NODES, s in CENTROIDS, t in
0..TimeforWholeLinkFree-1 :(A[i,j]==1))
                    U[i,j,s,t] == 0;
          C31:// DECISION CONSTRAINT
                 P[1,0] == 430;
                 P[2,0] == 430;
                 P[3,0] == 430;
                 P[4,0] == 430;
 }
```

ii) Two Regime Transmission Model Mathematical Proof

A brief mathematical explanation of the two-regime transmission model will be presented. A more detailed explanation can be found in supporting papers in the bibliography.

Let p(x,t) represent the traffic density and q(x,t) represent the flow of vehicles at time t and point x from the start of any link from A. Note that link index *(i, j)* has been dropped for simplicity but the following equations apply for all links in A. Speed is assumed to be solely dependent on total density along each link. Based on the LWR model, density and flow are related by the following equation:

$$\frac{\partial p}{\partial t} + \frac{\partial q}{\partial x} = 0$$

Then, based on definition, flow at length 0 corresponds to inflow, U(t):

$$q(0, t) = U(t)$$

and flow at length l(total length) corresponds to outflow, V(t):

$$q(l, t) = V(t)$$

Flow is a function of density via the equation:

$$q(x, t) = \varphi(p(x, t))$$

By assuming that φ(.) follows that of a triangular flow-density relationship,

$$q(x, t) = \varphi(p) = \begin{cases} vp, & 0 \leq p < C; \\ w(K - p), & C \leq p \leq K \end{cases}$$

where critical density, $C = \frac{Kw}{v+w}$ based on backward propagation congested speed, w and free-flow speed, v as shown by Figure A1.

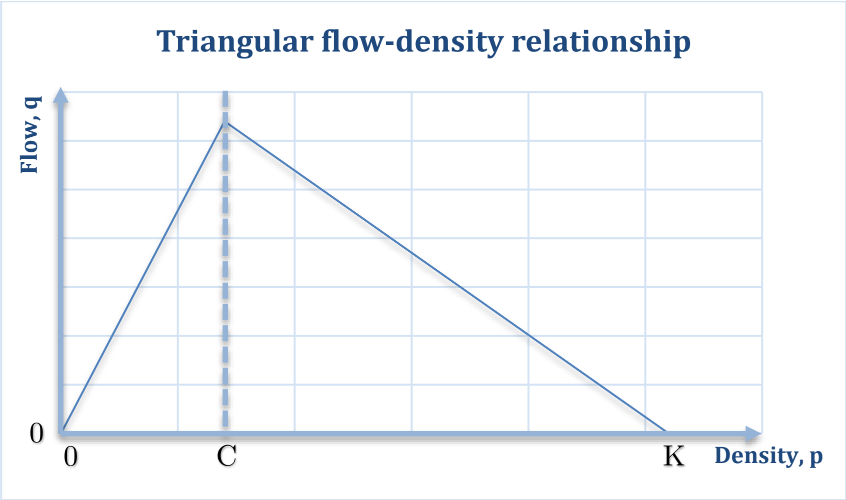

Figure A1. Triangular flow-density relationship

Based on Newell's equation, where cumulative flow along a wave varies at a fixed rate depending on condition of traffic, we can find the density at the particular time by tracing back to an earlier time of a boundary location (start or end of the link):

$$p(x,t) = \begin{cases} p(0, t - \frac{x}{v}), & 0 \le p < C; \\ p(l, t - \frac{l-x}{w}), & C \le p \le K \end{cases}$$

Combining the above equations, we can find the flow of the link depending on the densities via:

$$q(x,t) = \varphi(p) = \begin{cases} \varphi\left(p(0, t - \frac{x}{v})\right) = U(t - \frac{x}{v}), & 0 \le p < C; \\ \varphi\left(p(0, t - \frac{l-x}{w})\right) = V(t - \frac{l-x}{w}), & C \le p \le K \end{cases}$$

There may be 4 different cases that can happen at any time instant t:

1) Traffic begins at the start of the link when there are no vehicles at the downstream end.

2) There is no congestion at any part of the link, i.e. the link is free-flowing. This arises when $V(t) = U\left(t - \frac{l}{v}\right)$.
3) The whole link is congested, i.e. when $V\left(t - \frac{l}{w}\right) = U(t)$.
4) Downstream end of link is congested while upstream is free as shown in the figure below, i.e. when $U(t) > V\left(t + \frac{l}{v}\right)$, where l^f represents the length of the free-flow regime and l^c represents the length of the congested length of the link. This means that $l = l^f(t) + l^c(t)$. Refer to Figure 1.

The number of vehicles in a link is defined as the sum of all outflow $V(k)$ subtracted by the sum of inflow $U(k)$ for each time interval from 0 to t, represented by:

$$n(\mathrm{t}) = \sum_{k=0}^{t} U(k) - V(k)$$

which can also be defined based on the density p(x, t) as:

$$n(t) = \int_0^l p(x,t)dx$$

By substituting the equations of density earlier above into the above equation, we obtain:

$$n(t) = noncongested\ regime + congested\ regime$$

$$= \frac{1}{v}\int_0^{l^f(t)} U\left(t - \frac{x}{V}\right)dx + \int_0^{l^c(t)} K - \frac{1}{W}V\left(t - \frac{x}{W}\right)dx$$

We can obtain the number of vehicles at two extreme states of the link, 1) when the link is fully at free-flow speed:

$$\mathrm{n^f}\ (\mathrm{t}) = \sum_{k=\left(\mathrm{t}-\frac{\mathrm{L_{ij}}}{\mathrm{v_{ij}}}+1\right)}^{\mathrm{T}} U(k)$$

And 2) when the link is fully congested:

$$n^{\mathrm{c}}\ (\mathrm{t}) = \mathrm{K}l - \sum_{k=\left(\mathrm{t}-\frac{\mathrm{L_{ij}}}{\mathrm{w_{ij}}}+1\right)}^{T} \sum V(k)$$

As such, the number of vehicles in the link is bounded by these 2 conditions:

$$n^{\mathrm{f}}\ (\mathrm{t}) \leq n(t) \leq \mathrm{n^c}\ (\mathrm{t})$$

Inflow, U(t) and outflow, V(t) are bounded by the link capacity, given by:

$$U(t) \leq \frac{Kvw}{v+w}$$

$$V(t) \leq \frac{Kvw}{v+w}$$

Given that

$$n(i) = \lim_{t \to (i+1)-} \frac{1}{v} \int_0^{l^f(t)} U\left(t - \frac{x}{V}\right) dx + \int_0^{l^c(t)} K - \frac{1}{W} V\left(t - \frac{x}{W}\right) dx$$

and that U(t) and V(t) are steady, i.e.

$$U(t) = U(t+i) = U(i)$$

$$V(t) = V(t+i) = V(i)$$

then:

$$n(t) \approx l^c(t)\left(K - \frac{U(t)}{v} - \frac{V(t)}{w}\right) + \frac{U(t)l}{v}$$

Therefore, the length of the congested regime can be approximated as such:

$$l^c(t) \approx \frac{n(t) - \frac{U(t)l}{v}}{\left(K - \frac{U(t)}{v} - \frac{V(t)}{w}\right)}$$

A unique solution for the length of the congested regime can be found if the following conditions hold:

$$U(k) < H \qquad \forall k \in [t - \frac{l}{v}, t]$$

$$V(k) < H \qquad \forall k \in [t - \frac{l}{w}, t]$$

where $H = \frac{Kvw}{v+w}$.

Further explanations can be found in Ngoduy et al. (2016).