A branch and price approach for routing and refueling station location model

Barış Yıldız, Okan Arslan, Oya Ekin Karaşan*

Bilkent University, Department of Industrial Engineering, Bilkent, 06800 Ankara, Turkey

ARTICLE INFO

ABSTRACT

The deviation flow refueling location problem is to locate p refueling stations in order to maximize the flow volume that can be refueled respecting the range limitations of the alternative fuel vehicles and the shortest path deviation tolerances of the drivers. We first provide an enhanced compact model based on a combination of existing models in the literature for this relatively new operations research problem. We then extend this problem and introduce the refueling station location problem which adds the routing aspect of the individual drivers. Our proposed branch and price algorithm relaxes the simple path assumption generally adopted in the existing studies and implicitly takes into account deviation tolerances without the pregeneration of the routes. Therefore, the decrease in solution times with respect to existing models is significant and our algorithm scales very efficiently to more realistic network dimensions.

© 2015 Elsevier B.V. and Association of European Operational Research Societies (EURO) within the International Federation of Operational Research Societies (IFORS). All rights reserved.

1. Introduction

Due to economic, security and environmental concerns associated with fossil fuels, the penetration of alternative fuel vehicles into the transportation network is on the rise. Alternative fuel vehicle (AFV) technologies aim at reducing the greenhouse gas emissions, the cost of transportation and the dependence on export oil. Introduction of these game-changing technologies bring about several opportunities for different players of the transportation sector. However, a widespread adoption of vehicles by the community is contingent upon the availability of refueling stations for alternative fuels. Lack of these stations is identified as one of the foremost barriers by several researchers (Bapna, Thakur, & Nair, 2002; Kuby & Lim, 2005; Melaina & Bremson, 2008; Melaina, 2003; Romm, 2006). On the other hand, establishing new refueling stations by the private sector necessitates a large number of vehicles on the road (Kuby & Lim, 2005; Melaina, 2003; 2007). This 'chicken-egg' problem (Kuby & Lim, 2005; Melaina, 2003; Wang & Wang, 2010) led to several studies flourish in the recent literature. Commonly assuming a government participation in the initial phase of refueling station establishment, the major concern has been to locate a given number of stations in a road network.

In the existing literature, different modeling approaches are used to locate the refueling stations. Early studies in this area (Goodchild & Noronha, 1987; Nicholas, Handy, & Sperling, 2004; Nicholas & Ogden, 2006) utilized the p-median model to minimize the sum of the travel times from the demand sites (i.e. homes) to the nearest refueling facilities. The motivation behind p-median models is that the vehicle owners usually prefer to refuel close to their homes (Kitamura & Sperling, 1987; Upchurch & Kuby, 2010). The p-median approach assumes that the demand is located at nodes. A different approach to the refueling station location problem considers path-based demand. This idea is initially presented in flow capturing location model (FCLM) by Hodgson (1990) and in flow intercepting location model (FILM) independently by Berman, Larson, and Fouska (1992). A path-based demand is considered to be 'captured' if the path contains a node with an open facility. In other words, a single facility is assumed to be enough to cover the whole flow on the path. The objective is to locate p facilities while capturing as much path flows as possible. Unfortunately, the single refueling stop assumption of FCLM is too restrictive to represent the real world cases in which the distance between an origin–destination (O–D) pair is larger than the range of the vehicle. This shortcoming of flow capturing approach is more severe when it comes to the AFVs which are infamous for their rather limited ranges. To handle this, Kuby and Lim (2005) introduced flow refueling location model (FRLM) that locates p refueling stations to maximize the total refueled flow volume while making sure that the vehicles never run out of fuel. Similar to FCLM, the demand is defined as a flow on the shortest path between an O–D pair. But this time, rather than a single facility, a certain set of stations enabling the round trip of the vehicle between an O–D pair is required. In other words, a 'combination of facilities' is needed to serve the demand so that the

* Corresponding author. Tel.: +90 312 290 1409; fax: +90 312 266 4126.
  E-mail addresses: baris.yildiz@bilkent.edu.tr (B. Yıldız), okan.arslan@bilkent.edu.tr (O. Arslan), karasan@bilkent.edu.tr (O.E. Karaşan).

Please cite this article as: B. Yıldız et al., A branch and price approach for routing and refueling station location model, European Journal of Operational Research (2015), http://dx.doi.org/10.1016/j.ejor.2015.05.021

vehicles do not run out of fuel while traveling. In the initial phase of the two-stage solution methodology, feasible minimal combinations that can refuel the shortest path between each O–D pair are determined by a preprocessing algorithm. These combinations are given as input to a mixed integer linear programming (MILP) formulation in the second stage. In FRLM, at least a half-full tank of fuel is required at the final destination with no refueling station (Capar, Kuby, Leon, & Tsai, 2013; Kuby & Lim, 2005, 2007; Kuby, Lines, Schultz, Xie, Kim, & Lim, 2009; MirHassani & Ebrazi, 2013). This enables the vehicle to have enough fuel to complete a round trip. If a refueling station is located at the destination node, the half-full tank requirement is relaxed. This is a very realistic assumption since no AFV driver would like to reach the destination without enough fuel to visit a refueling station on the return trip. With the same reasoning, a similar assumption is made for the origin nodes. This basic FRLM formulation is extended from different aspects and some assumptions are relaxed in further studies. The objective function is modified to maximize the total vehicle-miles traveled (Kuby et al., 2009). The feasible set of candidate sites for refueling stations is extended from the node set to include the points on the arcs as possible location points by Kuby and Lim (2007). A multi-period planning for charging station infrastructure is proposed by Chung and Kwon (2015).

The FRLM requires the generation of all combinations for all the path-based demands. Thus, building the model for even medium-sized networks requires excessive time and memory. In order to overcome this drawback, Lim and Kuby (2010) propose three heuristic algorithms: greedy-adding, greedy-adding with substitution and genetic algorithm. In a similar line of efforts, a different refueling logic is embedded into the MILP model by Capar et al. (2013). The authors propose a simple, yet powerful formulation that solves the FRLM to optimality in a reasonable amount of time.

Different approaches such as set covering are also studied in the recent literature (MirHassani & Ebrazi, 2013; Wang & Lin, 2009, 2013; Wang & Wang, 2010). Rather than locating p facilities to serve the demand, a set covering approach finds the minimum-cost combination of facilities to serve all of the O–D demand pairs. MirHassani and Ebrazi (2013) approach this problem from a different perspective to increase the size of the problems that can be solved to optimality. Initially building an expanded network in which augmented arcs correspond to path segments of the shortest paths through which vehicles can bypass nodes without refueling, the need for combinations disappears. An effective mixed-integer linear programming (MILP) formulation based on the shortest path problem is provided. They do not consider flow deviation (driver preferences) and assume a fixed simple path, namely, the shortest path, between each O–D pair. With the fixed path assumption, the resulting MILP formulation can be directly solved by a commercial solver for realistic problem instances.

All of the aforementioned studies consider only a fixed number of simple paths to connect O–D pairs. Although fixing paths and using only simple paths make problems computationally tractable, they unnecessarily restrict the solution space. It is clear that considering only a small portion of all possible paths can result in a suboptimal solution. For the simple path case, consider the example depicted in Fig. 1. In order to cover both demands between O–D pairs o_1-d_1 and o_2-d_2, two stations located at nodes A and C are required if we only consider simple paths. However, if non-simple paths are viable, a single refueling station located at node B would cover both demands. The presented example oversees capacity issues related to stations. Capacitated refueling stations are within the scope of recent studies such as Upchurch, Kuby, and Lim (2009b) and Jung, Chow, Jayakrishnan, and Park (2014). Though not within our scope, the flexibility provided by non-simple paths might prove useful in capacitated networks as well.

In the context of AFV routing, several studies flourished in the recent literature (Arslan, Yıldız, & Karaşan, 2014b; Artmeier, Haselmayr, Leucker, & Sachenbacher, 2010; Bektaş & Laporte, 2011; Erdoğan &
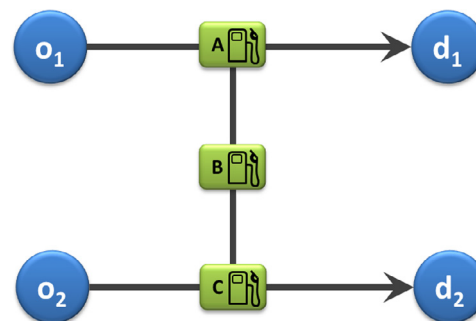

Fig. 1. Non-simple path example.

Miller-Hooks, 2012; Schneider, Stenger, & Goeke, 2014). These studies consider routing of AFVs including electric vehicles. Kuby, Araz, Palmer, and Capar (2014) also provide a decision-support tool for finding the shortest feasible path in a road network given the vehicle's driving range and station locations. However, there are very few studies in the refueling station location literature that incorporate the driver preferences into the location decisions. The effects of driver preferences such as deviating from the shortest paths is a significant factor on travel costs (Arslan, Yıldız, & Karaşan, 2014a). In this context, Kim and Kuby (2012) study simple-path deviations (i.e. cycles are excluded) from the shortest paths. The deviations are calculated by a k-shortest path algorithm before the model is solved until a predefined user tolerance deviation is reached. The deviation is defined as the percentage difference of the selected route and the shortest path. Similar to FRLM, the preprocessing time in this deviation flow refueling location model (DFRLM) is excessive when deviations are considered. Therefore Kim and Kuby (2013) propose a network transformation heuristic to solve realistic-sized problems. This transformation does allow for limited non-simple paths in the form of single cycles either at the start or end of the path. Huang, Li, and Qian (2015) also relax the commonly adopted assumption that travelers only take a shortest path between any O–D pair and study the multipath refueling location model, in which multiple deviation paths between O–D pairs can be simultaneously utilized.

In a similar context, routing is considered in a recent study by Kang and Recker (2014). In order to account for the routing decisions of the drivers, household activity pattern problem (HAPP) (Recker, 1995) is used, which is a variation of the pickup and delivery problem with time windows. The authors consider the routing decisions of the individuals in a metropolitan area and simultaneously optimize the scheduling and routing decisions of the households as well as the location of the refueling stations. The limited range of the vehicles is not considered in this study. Instead, it is presumed that each household visits a refueling station once in a day either on the way to another activity or as a single trip.

1.1. Contribution

In this paper, we study the refueling station location problem with routing considerations as a generalization of the DFRLM by Kim and Kuby (2012) and propose a branch and price algorithm as an exact solution methodology. The methodology combines existing ideas from the literature such as avoiding the explicit pregeneration of the routes and adding the flexibility of the non-simple paths in a novel manner by incorporating a path-segment based expanded network. Our unifying solution approach can also handle multiple vehicle types. We conduct extensive numerical experiments to solve this theoretically challenging and practically important problem. Our contributions to the existing literature are as follows:

Table 1
Nomenclature.

Indices	
h	Combination index
k	Candidate site index
q	O–D pair index
r	Alternative path index
Sets	
A	Set of arcs
A_{qr}	Set of arcs on alternative path rth of O–D pair q (considering a round trip)
H	Set of all combinations
H_{qr}	Set of combinations that can refuel alternative path rth of O–D pair q (considering a round trip)
K	Set of all candidate sites
K_h	Set of candidate sites in combination h
$K_{j,k}^{qr}$	Set of candidate sites that can refuel the directional arc $(j, k) \in A_{qr}$
N	Set of nodes
Q	Set of O–D pairs
R_q	Set of alternative paths between O–D pair q
Parameters	
f_{qr}	Flow on alternative path rth of O–D pair q
g_{qr}	Fraction of drivers traveling between O–D pair q who are willing to take the alternative path rth
p	Number of refueling stations to be located
Variables	
v_h	1 if all of the refueling stations in combination h is located, 0 otherwise
x_k	1 if a refueling station is located at candidate site k, 0 otherwise
y_{qr}	1 if flow on alternative path rth of O–D pair q is refueled , 0 otherwise
Note:	If only a single path between an O–D pair is considered, then the r subscript can be dropped.

- We bring different state-of-the-art models in the literature together to enhance the solution of DFRLM and show that the solution times decrease dramatically.
- We introduce the refueling location station problem with routing (RSLP-R) that generalizes DFRLM to handle the non-simple path deviations from the shortest path and present its complexity status.
- We propose a branch and price algorithm for solving the RSLP-R. The solution time decrease is significant with respect to the original DFRLM model. Moreover, because the algorithm does not require the explicit enumeration of paths, it scales very well to more realistic network dimensions.

In Section 2, we unify the state-of-the-art models to improve the solution efficiency of DFRLM. In Section 3, we present RSLP-R, provide its complexity status and detail our proposed branch and price methodology. In Section 4, an extensive computational study is conducted to attest the computational efficiency of the enhanced DFRLM as well as the proposed branch and price methodology. Section 5 concludes the study.

2. Enhancements to deviation flow refueling location model (DFRLM)

In this part, we present two enhancements to improve the solution time of the DFRLM: the first one in the modeling logic and the second one in the data generation algorithm. The parameters and variables to be used in the formulations in this section are presented in Table 1.

2.1. Model logic

The original FRLM presented by Kuby and Lim (2005) considers shortest path trips between each O–D pair. Since there is only one path for each O–D pair, the r subscript is dropped from the parameters and variables in the following FRLM formulation:

$$\text{maximize} \sum_{q \in Q} f_q y_q \tag{1}$$

$$\text{subject to} \sum_{h \in H_q} v_h \geq y_q \quad \forall q \in Q \tag{2}$$

$$x_k \geq v_h \quad \forall h \in H, k \in K_h \tag{3}$$

$$\sum_{k \in K} x_k = p \tag{4}$$

$$x_k, y_q, v_h \in \{0, 1\} \quad \forall k \in K, q \in Q, h \in H \tag{5}$$

The objective function maximizes the total flow refueled. Constraints (2) ensure that a path-based demand is satisfied only when a combination that can refuel the demand is selected. Constraints (3) ensure that whenever a combination is selected all the facilities in it are opened. Constraint (4) limits the number of facilities to be opened to p. Constraints (5) are the domain requirements. In FRLM, a shortest path for each O–D pair is considered as a demand. In the preprocessing phase, all of the facility combinations that can refuel these paths are generated. As previously mentioned, generation of these combinations require extensive amount of time, especially when the path is much longer with respect to the range of the vehicle. Capar et al. (2013) presented a different modeling logic that reduces not only the preprocessing times but also the model solution times. Without generating the feasible combinations for each path, this new logic models the 'refuelability' of the arcs. Instead of the Constraints (2) and (3) that enforce the refueling logic in the original model, the following constraints are added to the new formulation

$$\sum_{i \in K_{j,k}^q} x_i \geq y_q \quad \forall q \in Q, (j, k) \in A_q \tag{6}$$

where $K_{j,k}^q$ is the set of candidate sites that can refuel the directional arc $(j, k) \in A_q$ for the round trip between O–D pair q. This new set of constraints ensure that each arc on a given path is traversable by refueling at any of the possible candidate sites. Thus, rather than generating all feasible combinations for a given path, each arc on every path is processed once to make sure that it is traversable. Even though the new formulation also has a preprocessing part to generate the $K_{j,k}^q$ sets, generation is much faster especially for large networks.

The modeling logic extension to FRLM can also be applied to the deviation flow refueling location model (DFRLM) of Kim and Kuby (2012) which is presented below

$$\text{maximize} \sum_{q \in Q} \sum_{r \in R_q} f_q g_{qr} y_{qr} \tag{7}$$

subject to $\sum_{r \in R_q} y_{qr} \leq 1 \quad \forall q \in Q$ (8)

$\sum_{h \in H_{qr}} v_h \geq y_{qr} \quad \forall q \in Q, r \in R_q$ (9)

$x_k \geq v_h \quad \forall h \in H, k \in K_h$ (10)

$\sum_{k \in K} x_k = p$ (11)

$x_k, y_{qr}, v_h \in \{0, 1\} \quad \forall k \in K, q \in Q, h \in H, r \in R_q$ (12)

In DFRLM model, the original FRLM model by Kuby and Lim (2005) is modified to account for the deviations. A new subscript r is introduced to refer to the path alternative of the path-based demand q. The model incorporates demand decays as a function of deviation percentage from the shortest path. The parameter g_{qr} in the objective function is the fraction of drivers traveling between O–D pair q who are willing to take alternative path r. It equals to 1 for the shortest paths, and changes in a nondecreasing fashion with respect to increasing deviation distance of the alternative paths. Due to the nature of the objective function, the shorter alternative is selected among the possible set of alternative paths between an O–D pair. In other words, the flow with the highest possible fractional value contributes to the objective function. Constraints (8) ensure that at most one of the alternative paths between an O–D pair can be selected to prevent double-counting.

Observe that, similar to the study by Capar et al. (2013), Constraints (9) and (10) can be replaced by the following constraints to handle the model more efficiently

$\sum_{i \in K_{j,k}^{qr}} x_i \geq y_{qr} \quad \forall q \in Q, (j, k) \in A_q, r \in R_q$ (13)

Next, we deal with the preprocessing part of these models.

2.2. Improving data generation time

The DFRLM model considers an upper-limit on the driver tolerance as a fraction of the shortest path distance. Therefore, besides generating data for combinations, it also generates all of the paths up to a predefined distance. In order to enumerate these paths, the authors propose to solve k-shortest paths algorithm, starting at $k = 1$ and increasing it one by one until the path distance exceeds the driver's tolerance. Observe that generating these paths requires excessive amount of time and amounts to a big portion of the data preparation. However, more efficient algorithms such as 'algorithm for loopless paths near shortest path' (ANSPR0) algorithm by Carlyle and Wood (2005) exist in the literature to enumerate the paths up to a predefined distance value. Rather than solving the k-shortest paths for several times and keeping a sorted list of paths, the ANSPR0 algorithm processes arcs in a depth-first-search fashion and outputs a path if its length is less than or equal to the predefined distance. As it will be presented in the computational study section, this approach effectively reduces the preprocessing time of the model in orders of magnitude.

2.3. Decay function

Within DFRLM context, it is typically assumed that the demand decays by increasing deviation from the shortest distance. In their study, Kim and Kuby (2012) define the decay as a function of the deviation. In a recent study, Kuby, Kelley, and Schoenemann (2013) report empirical data for deviation decay in the city of Los Angeles. We assume, for each potential deviation path alternative, that

we have an associated penalty coefficient originating from an underlying demand decay model. Our proposed RSLP-R model, unlike current DFRLM studies in the literature, does not take as input a given set of alternative paths for a specific O–D pair. As such, in order to incorporate the penalty associated with a potential deviation path alternative, we transform the input data associated with the underlying demand decay model as follows: Consider a specific $q \in Q$ with m potential deviation path alternatives, and let $g_{q_1} \geq g_{q_2} \geq \cdots \geq g_{q_m}$ be the associated penalty coefficients. We can represent this particular O–D pair with m copies of it, say $q_1, \ldots, q_m$ originating from the same source and terminating in the same destination where $f_{q_i} = f_q \times (g_{q_i} - g_{q_{i+1}})$, $\forall i < m$ and $f_{q_m} = f_q \times g_{q_m}$. Observe that, with this transformation, the same percentage of flow will be refueled as the original model. In particular, if the alternative path rth is refueled in the DFRLM model, then with this transformation, demands $q_r \ldots q_m$ will all be refueled. Thus, the cumulative flow equals to $\sum_{i=r}^{m} (f_q \times (g_{q_i} - g_{q_{i+1}})) = f_q \times g_{qr}$.

2.4. Deviation flow refueling location model - enhanced (DFRLM-E)

With the above enhancements and modifications to the DFRLM model, we now propose the following DFRLM-E model that solves the same problem as DFRLM more efficiently. Note that the required path enumeration for the DFRLM-E is performed by the ANSPR0 algorithm by Carlyle and Wood (2005).

maximize $\sum_{q \in Q} \sum_{r \in R_q} f_q y_{qr}$ (14)

subject to $\sum_{r \in R_q} y_{qr} \leq 1 \quad \forall q \in Q$ (15)

$\sum_{i \in K_{j,k}^{qr}} x_i \geq y_{qr} \quad \forall q \in Q, (j, k) \in A_q, r \in R_q$ (16)

$\sum_{k \in K} x_k = p$ (17)

$x_k, y_{qr} \in \{0, 1\} \quad \forall k \in K, \forall q \in Q, r \in R_q$ (18)

In the computational study section, we present results showing that the solution times of the extended model are much faster than those of the classical one.

3. Mathematical model

In this section we formally define the refueling station location problem with routing (RSLP-R).

3.1. Problem definition and notation

An AFV trip has three components: vehicle, O–D pair and driver. For each trip, the fuel range (the maximum distance to be covered with a full fuel tank) is a function of vehicle specifications, the O–D pair indicates where the trip starts and ends and the driver preference determines how much extra driving can be tolerated by this driver. From a macroscopic view, those trips with the same vehicle, O–D pair and driver preference can be considered a single group which we call as a demand. The flow volume of a demand is given proportional to the amount of AFV trips. For each demand there is an associated traffic volume which is a function of the number of AFV trips in the considered time interval. Following the convention established in the literature, we assume that all the alternative fuel vehicle trips start with half full tank so that the driver can return on the same trip to the same station the next day with at least half full tank. A path is considered to be feasible for a given demand if it satisfies the following three conditions:

- It starts from the origin and ends in the destination node,
- There are enough refueling stations positioned on the path such that it is possible to travel without running out of fuel and arrive to the destination with at least half full fuel tank,
- Its length is not more than the threshold value that the AFV driver can tolerate.

A given demand is considered to be refueled if the designed station deployments enable a feasible path for it. In RSLP-R, the objective is to find the locations of a fixed number of refueling stations in the network such that the total volume of the refueled demand is maximized.

We now provide some basic notation. We assume the underlying physical network is represented by a weighted undirected graph with node set $N = \{1, 2, 3, \dots n\}$ and edge set E where each edge can be traversed in either direction and thus the refueling stations to be located are dual accessible. Corresponding to our physical network instance, we construct a directed weighted graph $G = (N, A)$ where $A = \{(i, j) \cup (j, i): \{i, j\} \in E\}$ and the length of each arc $a \in A$ is $l(a) \geq 0$ which is equal to the length of its corresponding edge.

Let $O, W \subseteq N$ be the sets of origin and destination nodes, respectively. We define the expanded network $\overline{G} = (\overline{N}, \overline{A})$ where:

- $\overline{N}$ contains nodes $\overline{i}$ for all $i \in O$ and $\overline{j}$ for all $j \in W$ in addition to the original set of nodes N.
- $\overline{A}$ consists of all the arcs in A plus the zero-length arcs $(\overline{i}, i)$ for all $i \in O$ and $(j, \overline{j})$ for all $j \in W$.

Between two nodes $s, t \in \overline{N}$, the shortest distance in $\overline{G}$ is denoted by $\delta^\star_{s,t}$.

We define M as the set of vehicle types. The range of a vehicle $\mu \in M$ is denoted by $r(\mu)$. A demand q is a five tuple $\langle m^q, \mathcal{S}(q), \mathcal{T}(q), \lambda^q, f_q \rangle$, where $m^q \in M$ is the vehicle type and $\mathcal{S}(q) = \overline{i}$ and $\mathcal{T}(q) = \overline{j}$ are the artificial origin and destination nodes associated with the O–D pair $i \in O, j \in W$. $\lambda^q \geq 0$ represents the maximum distance that the driver would accept to travel and f_q is the flow volume. The set of demands is denoted by Q.

A directed path is an alternating sequence of nodes and arcs $(n_0, a_1, n_1, a_2, n_2, \dots, a_\eta, n_\eta)$ with $n_i \in \overline{N}, \forall i = 0, \dots, \eta$ and $a_i = (n_{i-1}, n_i) \in \overline{A}, \forall i = 1, \dots, \eta$. A path is non-simple if it repeats nodes and is simple otherwise. Our formulation depends on the notion of *path-segments* introduced by Yıldız and Karasan (2014). Note that the idea of generating an artificial and reduced network among a fixed set of refueling locations where an edge is induced by a vehicle range dates back to a sequence of studies (including but perhaps not limited to Adler, Mirchandani, Xue, and Xia (2014); Khuller, Malekian, and Mestre (2007); Kim and Kuby (2013); Kuby et al. (2014); Lin, Gertsch, and Russell (2007); Soedarmadji and McEliece (2007); Suzuki (2008)). However, since the refueling locations are not fixed in our case, our path-segments are more flexible. In the particular case in which they correspond to shortest paths of the original network, they coincide with the MirHassani and Ebrazi (2013) definition given for fixed paths. In particular, a *path-segment* π is a directed simple path in $\overline{G}$ with an associated demand $d(\pi) \in Q$. We denote the source and destination nodes of a *path-segment* π as $s(\pi)$ and $t(\pi)$, respectively. The length of a path-segment is the sum of the lengths of the arcs on this segment and is denoted by $l(\pi)$. In our formulations, we only consider *path-segments* with total length less than the range of the vehicle type associated with it and call such path-segments *feasible*. More formally, a *path-segment* π is feasible if $l(\pi) \leq r(m^{d(\pi)})$. We define Π^q as the set of all those *feasible-path-segments* for a demand $q \in Q$ and denote the set of all the feasible path segments as Π, i.e., $\Pi = \cup_{q \in Q} \Pi^q$.

Using the same definitions and notation with Yıldız and Karasan (2014), a trip $\Phi = (\pi^1 \dots, \pi^k)$ is an ordered union of *feasible path-segments* $\pi^i, i \in 1, \dots, k$ where $t(\pi^i) = s(\pi^{i+1}), \forall i = 1, \dots, k - 1$. We call a trip feasible for a demand $q \in Q$, if $s(\pi^1) = \mathcal{S}(q), t(\pi^k) = \mathcal{T}(q)$,

$l(\Phi) = \Sigma_{i \in 1, \dots, k} l(\pi_i) \leq \lambda^q$ and a refueling station is located at $t(\pi^i)$, $\forall i = 1, \dots, k - 1$. We say an arc $a \in \pi$ if a is an arc on path-segment π. Similarly for a trip Φ, we say $\pi \in \Phi$ if π is a path-segment of Φ.

For a given node set $P \subseteq N$, let $\overline{Q}_P \subseteq Q$ be the set of demands for which there exists a feasible trip in $\overline{G}$ when a refueling station is located at every node in P. Then, RSLP-R can be formally stated as follows:

Definition 1. The refueling station location problem with routing (RSLP-R) is defined as finding a set $P^* \subseteq N$ with cardinality at most p such that the total amount of flow refueled $\sum_{q \in \overline{Q}_{P^*}} f_q$ is maximized.

Proposition 1. *RSLP-R is NP-Complete.*

Proof. Observe that for a given RSLP-R problem instance, the feasibility can be checked in polynomial time. In order to show that RSLP-R is NP-Complete, we now provide a transformation from the maximal covering location problem (MCLP) (Church & ReVelle, 1974) which is also NP-Complete (Megiddo, Zemel, & Hakimi, 1983). The MCLP is defined as selecting a combination of candidate facilities, with a cardinality less than or equal to p, such that the maximum demand is covered by the selected facilities. The parameters are the customers, $i \in I$, with a demand h_i; the facilities, $j \in J$; binary parameters a_{ij} to define the coverage of customer $i \in I$ by candidate facility $j \in J$; and a fixed number p. For this MCLP instance, we now build a graph as input to RCLP-R using the following polynomial-time transformation. For each candidate facility $j \in J$, add a node j. For each demand $i \in I$ with $a_{ij} = 1$, add two nodes i_o and i_d that represent an O–D pair q with a flow of h_i. Add the arcs (i_o, j) and (j, i_d) to the graph, both with a length of 1 unit. Consider the corresponding RSLP-R instance with a driver tolerance equal to 1, and a vehicle range of 2 units. Observe that solving this RSLP-R instance is equivalent to solving the corresponding MCLP instance. Thus, RSLP-R is NP-Complete. □

3.2. Path-segment formulation (PS)

In this subsection we present the path-segment formulation *PS* for RLSP-R and provide the details of the proposed branch and price algorithm to solve it. Recall that refueling a demand q requires to find a trip $\Phi = (\pi^1 \dots, \pi^k)$ such that a refueling station is located at the end of each path-segment $\pi \in \Phi \setminus \{\pi^k\}$ where $t(\pi^k) = \mathcal{T}(q)$. As such, our *path-segment* formulation admits a very natural representation of vehicle refueling constraints. Since there is no refueling at the interval nodes of a path-segment, it is always best to choose the shortest path among all the path segments between two nodes for the RSLP-R problem. Thus, we only need to consider the shortest path between two nodes as a path-segment. This core property is also considered by MirHassani and Ebrazi (2013) to represent refueling constraints for a vehicle traveling on a fixed path. Our methodology generalizes this approach to the whole network to relax the fixed simple path assumption.

We define the following decision variables.

$$y^q = \begin{cases} 1, & \text{if a feasible trip is built for the demand } q \in Q \\ 0, & \text{otherwise,} \end{cases}$$

$$x_i = \begin{cases} 1, & \text{if there is a refueling station located at node } i \in N \\ 0, & \text{otherwise,} \end{cases}$$

$$v^q_\pi = \begin{cases} 1, & \text{if demand } q \in Q \text{ uses path-segment } \pi \\ 0, & \text{otherwise,} \end{cases}$$

We call $y^q, q \in Q$ as the *cover* variables, $x_i, i \in N$ as the *location* variables and $v^q_\pi, q \in Q, \pi \in \Pi$ as the *path-segment* variables. With these decision variables, *PS* can be stated as follows:

$$\max \sum_{q \in Q} f_q y^q \qquad (19)$$

$$\text{s.t.} \sum_{\substack{\pi \in \Pi^q, \\ s(\pi)=i}} v_\pi^q - \sum_{\substack{\pi \in \Pi^q, \\ t(\pi)=i}} v_\pi^q = \begin{cases} y^q, & \text{if } i = \mathcal{S}(q) \\ -y^q, & \text{if } i = \mathcal{T}(q) \\ 0, & \text{otherwise} \end{cases} \quad \forall i \in \overline{N}, q \in Q,$$

(20)

$$\sum_{\pi \in \Pi^q} l(\pi) v_\pi^q \le \lambda^q \quad \forall q \in Q,$$

(21)

$$\sum_{\substack{\pi \in \Pi^q: \\ t(\pi)=i}} v_\pi^q \le x_i \quad \forall q \in Q, i \in N$$

(22)

$$\sum_{i \in N} x_i \le p$$

(23)

$$y^q \in \{0, 1\} \quad \forall q \in Q,$$

(24)

$$x_i \in \{0, 1\} \quad \forall i \in N,$$

(25)

$$v_\pi^q \in \{0, 1\} \quad \forall q \in Q, \pi \in \Pi$$

(26)

The objective function (19) is the total amount of the AFV flow volume to be captured. Constraints (20) are the flow balance equations that force a chosen demand to be carried from its source to its destination (covered) by the concatenation of *feasible-path-segments*. Constraints (21) are the maximum deviation constraints which ensure that the total length of any AFV trip is not longer than the maximum allowed. Constraints (22) enforce fuel range requirements by ensuring refueling at the end of each *feasible path-segment* that does not end in the destination node of the associated demand. Constraint (23) restricts the number of refueling stations to be at most p. Constraints (24)-(26) are the domain restrictions.

In order to strengthen the given formulation we can replace constraints (21) with the following constraints:

$$\sum_{\pi \in \Pi^q} l(\pi) v_\pi^q \le \lambda^q y^q \quad \forall q \in Q$$

(27)

This cut is very useful when solving the *PS* formulation. Indeed, as we will more formally present below, integrality of the location variables is sufficient to guarantee the integrality of the cover and path-segment variables with the inclusion of this cut in the model. A similar key result is established in FRLM context in Kuby and Lim (2005). We will call this stronger formulation as $\overline{PS}$. We now present our branch and price algorithm (B&P) to solve $\overline{PS}$. During B&P, the column generation technique is employed to solve the linear relaxation of $\overline{PS}$, say $\overline{PS}$-LP and obtain an upper bound for each node of the branch and bound tree.

3.3. LP solution (Column generation)

3.3.1. Pricing problem:

Let $\overline{RPS}$ be the restricted $\overline{PS}$ formulation with a subset of path-segment variables v_π^q. At every iteration we determine whether there exists a column with positive reduced cost such that including it to the $\overline{RPS}$ might improve the objective function. If such columns are detected, we add them to the $\overline{RPS}$ and repeat the procedure until there is no column left with a positive reduced cost.

Let ρ_i^q represent the unrestricted dual variables associated with constraints (20), and κ^q and γ_i^q be the nonnegative dual variables associated with constraints (21) and (22), respectively. For a path-segment variable v_π^q, the reduced cost $\bar{c}_\pi^q$ is given as

$$\bar{c}_\pi^q = \begin{cases} \rho_{t(\pi)}^q - \rho_{s(\pi)}^q - l(\pi)\kappa^q, & \text{if } t(\pi) = \mathcal{T}(q) \\ \rho_{t(\pi)}^q - \rho_{s(\pi)}^q - l(\pi)\kappa^q - \gamma_{t(\pi)}^q, & \text{o.w.} \end{cases}$$

(28)

Definition 2. An ordered node pair $(i, j) \in (\overline{N} \times N) \cup (N \times \overline{N})$ is called a *plausible-pair* for a demand q if it satisfies the following conditions:

- It is possible to transit from node i to node j without any refueling. More formally:

$$\delta_{i,j}^\star \le \begin{cases} r(m^q), & \text{if } i \ne \mathcal{S}(q) \text{ and } j \ne \mathcal{T}(q) \\ r(m^q)/2, & \text{o.w.} \end{cases}$$

(29)

- It is possible to visit nodes i and j without violating driver tolerance constraints. i.e.,

$$\delta_{\mathcal{S}(q),i}^\star + \delta_{i,j}^\star + \delta_{j,\mathcal{T}(q)}^\star \le \lambda^q$$

(30)

The set of all the *plausible-pairs* for a demand q is denoted by Ω^q.

In order to identify path-segment variables that price out, it is only required to check *plausible-pairs* for each demand $q \in Q$ and see if there is a pair $(i, j) \in \Omega^q$ such that, the shortest path $\pi_{i,j}^*$ from node i to j satisfies the following condition:

$$l(\pi_{i,j}^*)\kappa^q < \begin{cases} \rho_j^q - \rho_i^q, & \text{if } j = \mathcal{T}(q) \\ \rho_j^q - \rho_i^q - \gamma_j^q, & \text{o.w.} \end{cases}$$

(31)

Note that if the shortest path between a plausible pair (i, j) does not satisfy the above condition, none of the other paths connecting node i to node j can. Thus, for a plausible pair $(i, j) \in \Omega^q$, it is sufficient to check whether (31) is satisfied for the path segment $\pi_{i,j}^*$ and declare the variable $v_{\pi_{i,j}^*}^q$ as a positive reduced cost variable if this is the case.

3.3.2. Determining an initial set of columns

Defining variables as the path-segments instead of whole paths diverts from the widely used path based formulations for which the column generation technique has been applied very successfully for a wide range of problems (Lübbecke & Desrosiers, 2005). Path-segments as variables necessitate a more careful approach to determine the initial variable pool of path-segment variables (Yıldız & Karasan, 2014).

Let path segment $\pi_{i,j}^*$ be the shortest path between nodes $i, j \in \overline{N}$. Then we can define the initial variable pool as $V_0 = \{v_{\{\pi_{i,j}^*\}}^q | q \in Q, (i, j) \in \Omega^q, (i, j) \in \overline{A}\}$. Note that, a solution for the $\overline{RPS} - LP$, considering only the path-segment variables in V_0 contains enough information to derive all the needed dual variable values to properly construct the pricing problem.

3.4. IP solution

In $\overline{PS}$, all the decision variables are defined as binary. However, due to (27), requiring only the location variables as binary is sufficient to obtain a solution in which both cover and path-segment variables are also binary. Before proceeding with the formal propositions and their proofs, we need the following definition:

Definition 3. For a given solution (y, x, v) of $\overline{PS}$-LP, we call $G_v^q = (\overline{N}, A_v^q)$ as the reduced graph of demand $q \in Q$, where $A_v^q := \{a \in \overline{A} | a \in \pi, v_\pi^q > 0\}$.

Proposition 2. Let $(\hat{y}, \hat{x}, \hat{v})$ be an optimal solution for the $\overline{PS}$-LP where location variables $\hat{x}$ are all binary. Then, the cover variables $\hat{y}$ necessarily assume integral values.

Proof. Let $(\hat{y}, \hat{x}, \hat{v})$ be an optimal solution of $\overline{PS}$-LP, where $\hat{x}_i \in \{0, 1\}, \forall i \in N$ and $\hat{z}$ is the optimal solution value. Assume there exists $\hat{q} \in Q$ such that $0 < \hat{y}^{\hat{q}} < 1$. Let $U^{\hat{q}}$ be the set of trips that connect $\mathcal{S}(\hat{q})$ to $\mathcal{T}(\hat{q})$ in $G_{\hat{v}}^{\hat{q}}$. Note that $U^{\hat{q}}$ is not empty since $\hat{y}^{\hat{q}} > 0$ and the solution $(\hat{y}, \hat{x}, \hat{v})$ is feasible. Let u^* be the shortest trip in $U^{\hat{q}}$. Now consider the solution $\bar{y}, \hat{x}, \bar{v}$ where

$$\bar{y}^q = \begin{cases} 1, & \text{if } q = \hat{q} \\ \hat{y}^q, & \text{o.w.} \end{cases} \quad \bar{v}_\pi^q = \begin{cases} 1, & \text{if } q = \hat{q} \text{ and } \pi \in u^* \\ 0, & \text{if } q = \hat{q} \text{ and } \pi \notin u^* \\ \hat{v}_\pi^q, & \text{o.w.} \end{cases}$$

(32)

Observe that this new solution $(\bar{y}, \hat{x}, \bar{v})$ is feasible since location variables are all integral and $u^* \leq \lambda^q$. Let $\bar{z}$ be the objective function value for the solution $(\bar{y}, \hat{x}, \bar{v})$. Then, $\bar{z} - \hat{z} = f_q(1 - y^q) > 0$. This concludes the proof. □

Proposition 3. *Let $\hat{z}$ be the optimal solution value for $\overline{PS}$-LP obtained by the solution $(\hat{y}, \hat{x}, \hat{v})$, where location variables $\hat{x}$ and cover variables $\hat{y}$ are all binary. Then the optimal solution value for $\overline{PS}$ is equal to $\hat{z}$.*

Proof. Let $(\hat{y}, \hat{x}, \hat{v})$ be the optimal solution for $\overline{PS}$-LP where location variables $\hat{x}$ and cover variables $\hat{y}$ are all binary and assume that there exists a demand $\hat{q} \in Q$ with a positive cover variable $\hat{v}_\pi^{\hat{q}} < 1$ (if there is no such path-segment variable, then the assertion is vacuously true). $U^{\hat{q}}$ and u^* definitions are the same as their definitions in the previous proof. Now consider the solution $(\hat{y}, \hat{x}, \bar{v})$ where

$$\bar{v}_\pi^q = \begin{cases} 1, & \text{if } q = \hat{q} \text{ and } \pi \in u^* \\ 0, & \text{if } q = \hat{q} \text{ and } \pi \notin u^* \\ \hat{v}_\pi^q, & \text{o.w.} \end{cases} \quad (33)$$

Observe that the integrality of location and cover variables and u^* being the shortest trip in $U^{\hat{q}}$ ensure that the new solution $(\hat{y}, \hat{x}, \bar{v})$ is feasible with the same objective function value and strictly fewer fractional path-segment variables than the starting solution $(\hat{y}, \hat{x}, \hat{v})$. Since one can repeat this procedure as much as needed to obtain an integral solution, the proof is complete. □

Due to Propositions 2 and 3, we only need to consider the location variables in the branching phase.

Branching on location variables. Comparing the location variables x_i, $i \in N$ by the degrees of the associated node $i \in N$, we sort them in a descending order and obtain a priory list. Encountering a fractional solution, we select the fractional location variable highest in the list as the branching variable. Let x_i be the fractional location variable we chose to branch on.

- **Branching-cut-1** $x_i = 0$: In this case the set of path-segment variables $V_i = \{v_\pi^q | q \in Q, \pi \in \Pi^q \text{ and } t(\pi) = i\}$ are implicitly set to 0. Thus, we must make sure that in the pricing problem any path-segment $v_\pi^q \in V_i$ should not appear as a positive reduced cost column. This can be easily done by setting $\gamma_i^q = \infty \quad \forall q \in Q, \mathcal{T}(q) \neq i$.
- **Branching-cut-2** $x_i = 1$: In this case the path-segment variables are not affected by the branching cut and the pricing problem stays the same except for the possible change in the value of the dual variables γ_i^q.

4. Numerical experiments

Comprehensive numerical experiments are conducted to test the performance of the branch and price algorithm (B&P). Two particular network topologies are considered: 25-node road network in Fig. 2 (Simchi-Levi & Berman, 1988) and California (CA) road network in Fig. 3 (Arslan et al., 2014a). We implemented all the algorithms using Java under Linux and CPLEX 12.5 and all experiments are done on the same machine: AMD Opteron(tm) Processor 6282 SE with 2GB RAM. In the following, we first present the data and then the computational results in separate sections for each network considered.

4.1. Data

Being a commonly used network in the literature (Capar et al., 2013; Hodgson, 1990; Kim & Kuby, 2012; Kuby & Lim, 2005; Lim & Kuby, 2010; MirHassani & Ebrazi, 2013; Wang & Wang, 2010), 25-node road network constitutes a suitable test bed for us to compare the performance of B&P with the benchmark studies in the literature. The CA road network on the other hand is a close representation of
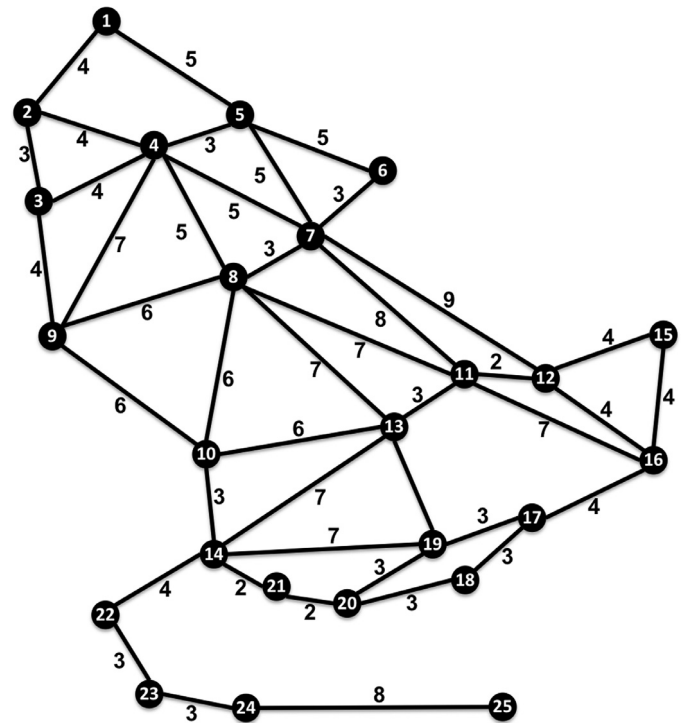
Fig. 2. 25-node road network.
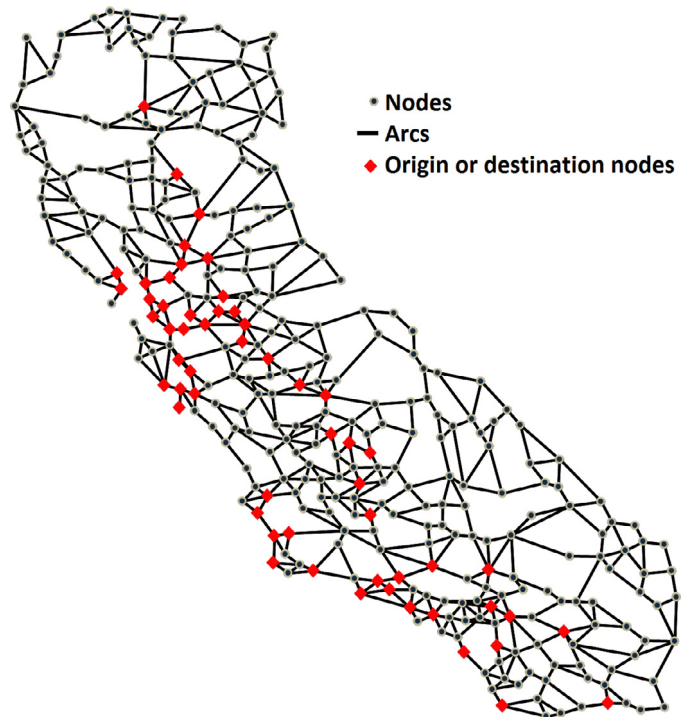
Fig. 3. California state road network.

the actual California State road network and allows us to test B&P in realistic large problem instances. The main parameters of these networks are presented in Table 2.

For the 25-node road network experiments, we generated the same test problems studied by Kim and Kuby (2012). All 25 nodes of the network are considered as O–D nodes and all the possible pairings between them are considered as O–D pairs. Note that we assume the same level of tolerances for all O–D pairs for a given setting. This

Table 2
Network and related O–D pair parameters.

Network	#nodes	#edges	Node degree			O–D pairs			
			min	max	mean	Count	min.dist	max.dist	mean.dist
25-node	25	42	1	6	3.36	300	2	38	14.23
CA	339	617	1	7	3.64	1167	30.06	463.50	153.37

Table 3
Solution time comparisons of DFRLM, DFRLM-E and B&P algorithms.

Range	Tol. (%)	Preprocessing time			Solution time in seconds (total)		
		DFRLM	DFRLM-E	B&P	DFRLM	DFRLM-E	B&P
4	0	3.85	0.15	0.17	4.14	0.42	1.51
	10	5.03	0.17	0.17	5.4	0.43	1.54
	50	54.52	0.27	0.20	54.98	1.55	1.88
8	0	3.89	0.15	0.19	4.3	0.5	2.25
	10	4.91	0.16	0.20	5.37	0.49	2.22
	50	57.68	0.27	0.23	72.22	4.3	3.72
12	0	3.97	0.15	0.21	4.46	0.37	2.51
	10	5.12	0.16	0.22	5.77	0.43	2.39
	50	82.38	0.27	0.23	130.2	11.25	4.70

can be further specified into distributions of tolerance levels by generating more demand types for the same OD pair. The flow is calculated by the gravity model proposed by Hodgson (1990). A total of 225 problem instances are obtained by considering

- 3 vehicle ranges: 4, 8 and 12
- 3 levels of driver tolerance: 0 percent, 10 percent and 50 percent
- 25 different refueling station numbers: 1, …, 25

In order to study more realistic problem instances, CA road network with 339 nodes and 617 edges test problems are used. For this set of experiments, all the urban population centers in the California are considered as O–D nodes. There are a total of 57 such centers with population more than 50,000 according to recent reports (U.S. Census Bureau, 2010). All possible pairings of these population centers that are not closer than 30 kilometers are considered as O–D pairs (1167 in total) and the volume of the flow on each pair is calculated using the gravity model (Hodgson, 1990). Our experimental design contains 64 problem instances

- 2 vehicle ranges: 100 and 150 kilometers,
- 4 levels of driver tolerance: 0 percent, 5 percent, 10 percent and 20 percent,
- 8 different refueling station limits: 1, 5, 10, …, 35.

We consider driver tolerances up to 20 percent in this realistic case study since higher tolerance values are hard to justify with economic or environmental concerns of the drivers.

4.2. 25-Node road network

Table 3 depicts the average CPU times in seconds of 25 runs ($p = 1, …, 25$) for different range and tolerance settings. For consistency with the available literature, we assumed a single vehicle type throughout our runs. The preprocessing time for DFRLM is the time it takes to generate the paths (by solving consecutive *k*-shortest path algorithms) and the minimal combination sets for these paths. The preprocessing time for DFRLM-E is for generating the paths using ANSPR0 algorithm and processing of each arc on each path, as explained in Section 2. The preprocessing time of B&P is for generating the plausible pairs for each demand. The right-most column, the solution time, shows the respective model solution time combined with the preprocessing time.

Results show that DFRLM-E runs an order of magnitude faster than its original version DFRLM in all the instances. Even though

branch and price algorithms are not as famous for their speed as their capability to handle large problem instances, it is interesting to observe that the run times of B&P are comparable to those of DFRLM-E. Apparently, problems with longer vehicle range and higher driver tolerance take longer solution times. In those cases, the number of alternative feasible paths between O–D pairs increases which makes these problems harder to tackle. Notice that the computational performances of DFRLM and DFRLM-E quickly degrade as problem gets harder whereas the solution times for B&P are more stable.

All three algorithms: DFRLM, DFRLM-E and B&P are run on all problem instances. Table 4 shows the solutions obtained by the B&P algorithm. In the table, *p* stands for the number of refueling stations, *Opt.Sol* shows the percentage of the flow that could be refueled in the optimal solution, *LP.sol* indicates the solution value for the linear relaxation of the problem, *BBN* is the number of branch and bound nodes explored by the B&P algorithm, *#Col.* indicates the total number of columns generated and *Time* is the solution time in seconds. Table 4 shows that optimal values are quite close to those of the linear relaxation solutions. This indicates the strength of the path-segment formulation which helps to make B&P a competitive alternative to the state-of-the-art models in the literature. Our results also show that the computational performance of the B&P algorithm does not vary significantly across different problem instances.

All solution values for the DFRLM and DFRLM-E are the same with those resulting from B&P except for the three cases depicted in bold in Table 4. For those instances, B&P is able to generate a better solution by utilizing non-simple paths. One example is when range is 12, the tolerance is 50 percent and $p = 6$. The refueling stations are located at nodes {4, 10, 12, 17, 20, 22} in the optimal solution. Even though there does not exist a feasible simple path between nodes 10 and 11, a non-simple path can connect these two nodes and cover the flow in between. When traveling from node 11 to node 10 on a non-simple path, the vehicle first visits node 12, refuels there, and travels to node 10 by visiting node 11 again. The travel distance in total is 13 which is just less than the tolerable maximum 13.5. This is an example of a non-simple path occurrence. It is no surprise that all these three highlighted cases share the same high range and tolerance (range = 12, tolerance = 50 percent) parameters. This is because, for a non-simple path to be feasible, the range of the vehicle should be long enough to traverse two consecutive arcs without refueling and driver tolerance should be high enough to compensate for the extra mileage of such a detour. Emergence of non-simple paths even in a quite aggregate network such as 25-node road network is an

Table 4

B&P solutions for the 25-node road network.

| Range	p	No tolerance					10 percent tolerance					50 percent tolerance				
		Opt.Sol	Lp.sol	#BBN	#Col.	Time	Opt.Sol	Lp.sol	#BBN	#Col.	Time	Opt.Sol	Lp.sol	#BBN	#Col.	Time
4	1	4.92	5.96	5	368	1.31	4.92	5.96	5	367	1.44	4.92	5.96	5	367	1.63
	2	6.31	11.91	35	379	3.08	6.31	11.91	35	377	2.98	6.31	11.91	35	377	3.84
	3	12.49	17.87	27	377	2.87	12.49	17.87	27	376	2.61	12.49	17.87	27	376	3.22
	4	20.38	23.82	19	372	2.16	20.38	23.82	19	376	2.35	20.38	23.82	19	376	2.68
	5	27.54	29.78	9	371	1.59	27.54	29.78	9	369	1.59	27.54	29.78	9	369	1.84
	6	34.01	35.73	7	370	1.58	34.01	35.73	7	373	1.52	34.01	35.73	7	373	1.75
	7	41.41	41.69	3	375	1.28	41.41	41.69	3	379	1.31	41.41	41.69	3	379	1.47
	8	45.26	47.64	9	431	2.12	45.26	47.64	9	403	2.05	45.26	47.64	9	403	2.52
	9	53.6	53.6	1	407	1.18	53.6	53.6	1	403	1.22	53.6	53.6	1	403	1.37
	10	55.97	56.71	3	473	1.39	55.97	56.71	3	441	1.32	56.08	57.98	3	441	2.47
	11	59.82	59.82	1	458	1.31	59.82	59.82	1	453	1.28	62.36	62.36	1	453	1.4
	12	61.51	61.84	5	500	1.49	61.69	61.84	5	503	1.55	64.41	64.41	3	503	1.62
	13	62.72	63.86	9	476	2.04	62.72	63.86	9	523	1.92	65.26	66.43	11	523	2.4
	14	65.12	65.88	5	488	1.62	65.12	65.88	5	502	1.7	67.66	68.45	7	502	2.28
	15	67.89	67.89	1	462	1.27	67.89	67.89	1	497	1.14	70.44	70.47	1	497	1.65
	16	69.58	69.58	1	481	1.26	69.77	69.77	1	512	1.3	72.48	72.48	1	512	1.44
	17	71.12	71.12	1	508	1.15	71.3	71.3	1	474	1.27	74.02	74.02	1	474	1.32
	18	71.81	72.23	3	470	1.4	71.99	72.42	3	474	1.44	74.84	74.84	3	474	1.34
	19	73.34	73.34	1	523	1.18	73.53	73.53	1	500	1.23	75.47	75.56	1	500	2.01
	20	73.98	73.98	1	575	1.08	74.22	74.22	1	581	1.15	76.28	76.28	1	581	1.38
	21	73.98	74.21	3	596	1.29	74.22	74.45	3	515	1.44	76.28	76.52	3	515	2.21
	22	74.45	74.45	1	598	1.12	74.68	74.68	1	517	1.42	76.75	76.75	1	517	1.41
	23	74.54	74.54	1	586	1.12	74.78	74.78	1	586	1.16	76.84	76.84	1	586	1.37
	24	74.54	74.54	1	370	1.05	74.78	74.78	1	357	1	76.84	76.84	1	357	1.23
	25	74.54	74.54	1	351	0.91	74.78	74.78	1	357	1.03	76.84	76.84	1	357	1.23
8	1	17.13	17.13	1	776	1.2	17.13	17.13	1	823	1.22	17.13	17.13	1	823	1.53
	2	32.58	32.58	1	778	1.18	32.58	32.58	1	873	1.3	32.58	32.58	1	873	2.08
	3	44.41	44.41	1	845	1.36	44.41	44.41	1	895	1.36	44.41	44.41	1	895	1.74
	4	55.97	55.97	1	892	1.38	55.97	55.97	1	958	1.43	56.08	56.08	1	958	2.06
	5	63.52	63.52	1	906	1.4	63.52	63.52	1	926	1.49	64.06	64.06	1	926	3.64
	6	68.08	68.74	5	1073	2.21	68.08	68.88	5	1093	2.33	71.61	71.61	5	1093	3.49
	7	72.32	73.95	9	1094	2.62	72.32	74.24	9	1137	2.69	75.32	79.08	7	1137	7.81
	8	75.39	79.16	17	1193	4.25	77.87	79.54	17	1144	2.63	84.56	85.64	5	1144	5.22
	9	82.35	84.25	5	1120	2.74	82.77	84.8	5	1245	3.18	92.18	92.18	9	1245	4.12
	10	87.58	89.33	9	1216	3.38	90.06	90.06	9	1189	1.88	95.99	95.99	1	1189	2.84
	11	94.41	94.41	1	1226	2.01	94.41	94.41	1	1200	1.95	98.25	98.25	1	1200	3.01
	12	96.8	96.8	1	1291	2.02	96.8	96.8	1	1318	2.04	98.76	98.76	1	1318	4.33
	13	97.78	98.07	7	1406	3.51	97.78	98.1	7	1421	3.41	99.03	99.11	7	1421	5.37
	14	98.36	98.57	9	1494	3.04	98.43	98.75	9	1494	3.41	99.45	99.45	9	1494	3.78
	15	98.48	98.97	5	1492	3.12	98.74	99.39	5	1512	3.47	99.72	99.76	7	1512	5.74
	16	99.17	99.21	3	1465	2.84	99.71	99.75	3	1496	3.05	99.81	99.85	3	1496	6.12
	17	99.24	99.29	9	1495	3.58	99.77	99.82	9	1513	3.89	99.87	99.93	15	1513	8.41
	18	99.33	99.36	3	1577	2.69	99.86	99.89	3	1463	3.17	99.97	99.98	3	1463	6.77
	19	99.39	99.39	2	1533	2.65	99.92	99.92	2	1461	2.7	100	100	2	1461	3.25
	20	99.39	99.39	2	1568	1.83	99.92	99.92	2	1424	1.81	100	100	1	1424	2.66
	21	99.39	99.39	2	1567	1.75	99.92	99.92	2	1539	1.73	100	100	1	1539	1.91
	22	99.39	99.39	1	1689	1.53	99.92	99.92	1	1250	1.4	100	100	1	1250	2.03
	23	99.39	99.39	1	1565	1.44	99.92	99.92	1	1604	1.54	100	100	1	1604	1.92
	24	99.39	99.39	1	1648	1.49	99.92	99.92	1	1339	1.35	100	100	1	1339	1.93
	25	99.39	99.39	1	481	1.01	99.92	99.92	1	619	1.08	100	100	1	619	1.24
12	1	18.23	18.23	1	1342	1.17	18.23	18.23	1	1475	1.27	18.23	18.23	1	1475	1.64
	2	34.34	34.75	3	1333	1.5	34.34	34.75	3	1562	1.91	34.34	34.75	3	1562	2.48
	3	47.9	47.9	1	1490	1.44	47.9	47.9	1	1618	1.48	49.04	49.04	1	1618	3.03
	4	57.47	57.47	1	1572	1.56	58.14	58.14	1	1833	2.16	62.64	62.64	1	1833	3.09
	5	66.18	66.18	1	1619	1.72	67.7	67.7	1	1818	1.91	72.46	72.77	1	1818	5.59
	6	72.53	74.11	9	1918	4.26	75	76	9	1992	2.83	**82.15**	82.5	3	1992	5.33
	7	80.88	81.57	9	1921	4.36	83.35	83.35	9	1941	2.27	**91.78**	91.78	1	1941	4.75
	8	87.33	87.4	7	1945	4.57	88.83	88.83	7	2009	2.48	**95.95**	95.95	1	2009	5.06
	9	92.71	92.71	1	1834	2.86	92.93	92.98	1	2027	2.85	97.59	97.75	3	2027	6.71
	10	96.83	96.83	1	1889	2.84	96.83	96.83	1	2037	2.83	98.97	99.18	1	2037	7.39
	11	97.81	98.03	5	1924	3.51	97.81	98.03	5	1994	3.45	99.54	99.55	5	1994	6.7
	12	98.66	99.16	13	1928	4.5	98.66	99.16	13	2049	4.71	99.8	99.82	13	2049	6.85
	13	99.3	99.57	5	2048	4.22	99.3	99.72	5	2187	4.7	99.89	99.94	13	2187	12.69
	14	99.85	99.85	1	1857	2.97	99.85	99.85	1	2096	3.61	99.95	100	3	2096	14.02
	15	99.93	99.93	1	1965	3.24	99.93	99.93	1	2163	3.51	100	100	1	2163	4.94
	16	100	100	1	2012	2.96	100	100	1	2062	3.07	100	100	1	2062	5.3
	17	100	100	1	1951	2.58	100	100	1	2052	2.42	100	100	2	2052	4.22
	18	100	100	2	1827	2.04	100	100	2	1901	2.05	100	100	1	1901	4.29
	19	100	100	1	1901	2.02	100	100	1	1919	1.88	100	100	1	1919	2.38
	20	100	100	4	2006	2.04	100	100	4	1863	1.66	100	100	1	1863	2.41
	21	100	100	1	1876	1.54	100	100	1	1863	1.63	100	100	1	1863	2.23
	22	100	100	1	1878	1.41	100	100	1	1725	1.37	100	100	1	1725	1.98
	23	100	100	1	1866	1.35	100	100	1	1773	1.42	100	100	1	1773	1.63
	24	100	100	1	1819	1.18	100	100	1	1669	1.33	100	100	1	1669	1.5
	25	100	100	1	640	0.99	100	100	1	984	1.05	100	100	1	984	1.18
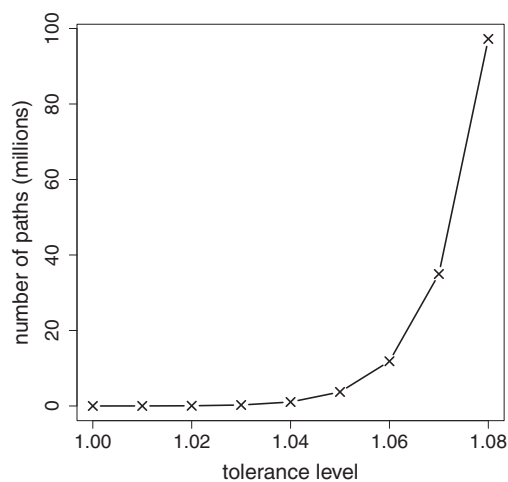
Fig. 4. Number of paths generated vs tolerance level.

interesting result which indicates that neglecting them could result in sub-optimal solutions especially in less aggregate and more realistic network instances.

4.3. CA road network

For the problem instances in CA road network, DFRLM and DFRLM-E fail to solve the problem for even minor driver tolerances. These models cannot even keep the problem in the memory in these problem instances. This is due to the exponential growth in the number of paths as the driver tolerance level increases. Illustrating this fact, Fig. 4 shows the total number of alternative paths that connect

O–D pairs for a given tolerance level. There are more than 3.5 million alternative paths for 5 percent driver tolerance and this number grows almost thirty times larger when the tolerance is increased to 8 percent. However, the B&P algorithm does not get overwhelmed by these large problem instances since it does not require the inclusion of all those paths, only a tiny fraction of which actually appear in the optimal solution.

Table 5 shows the results of the computational experiments with B&P on the CA road network. In the table, p stands for the number of refueling stations, *Sol.* is the percentage of the flow that could be refueled by the B&P algorithm solution, *%Gap* is the percentage of the optimality gap (with a time limit of 3 hours), *BBN* is the number of branch and bound nodes explored by the B&P algorithm and *#Col.* indicates the total number of columns generated and Time is the solution time in seconds. Empty rows indicate that the problem is not solved because 100 percent coverage is already established for less number of refueling stations.

As seen in the table, B&P is able to solve approximately 75 percent of the problems to optimality. For those instances, where B&P did not converge to the optimal solution in the given time limit of 3 hours, the maximum optimality gap is 0.506 percent and the average is below 0.007 percent. The results show that problems with small or large number of refueling stations are easier to solve and harder problems arise in between. Also problems with high tolerance values are naturally hard to solve since a higher number of columns are generated and considered in the solutions. The same claims and arguments are obviously true for the higher driving ranges.

Also note that higher driver tolerances make more significant differences in total flow for the medium values of p. For example consider the problem instances with $p = 5$, *range* = 100 kilometers in Table 5. For this set of problems just a 5 percent driver tolerance results in 9 percent increase in the total refuelable flow percentage.

Table 5
B&P solutions for the CA road network.

p	λ	Range = 100 kilometers					Range = 150 kilometers				
		Sol.	%Gap	BBN	#Col.	Time	Sol.	%Gap	BBN	#Col.	Time
1	1	30.545	0	1	71403	155	33.953	0	1	95141	200
	1.05	32.979	0	1	228320	146	34.439	0	1	312759	199
	1.1	33.285	0	1	372845	183	34.618	0	1	556776	221
	1.2	36.457	0	1	664187	202	36.828	0	1	1097566	306
5	1	67.084	0	1	70532	191	79.944	0	1	93654	217
	1.05	76.002	0	19	265154	811	84.136	0	3	337780	423
	1.1	79.573	0	3	409219	740	85.907	0	11	625782	1142
	1.2	82.861	0	29	877962	6556	89.078	0	109	1334565	10236
10	1	87.98	0	3	74059	239	92.984	0	15	104253	793
	1.05	91.977	0	7	246228	758	95.859	0	7	343821	878
	1.1	93.469	0	165	476869	6585	97.403	0	1	605160	989
	1.2	94.609	0.506	19	793718	10800	98.286	0	7	1270512	7262
15	1	95.008	0	1	86957	462	98.348	0	83	110644	2105
	1.05	97.793	0	27	276550	1655	99.435	0	37	354104	2204
	1.1	98.885	0	1	449125	910	99.793	0	31	633476	2497
	1.2	99.208	0.124	9	880796	10800	99.917	0.044	24	1267006	10800
20	1	98.407	0	69	91136	1980	99.89	0	619	114941	9165
	1.05	99.525	0	283	293126	10691	99.982	0.007	435	364327	10800
	1.1	99.82	0	229	499775	9885	99.969	0.028	193	666414	10800
	1.2	99.974	0.01	2	823954	10800	100	0	7	1128846	8820
25	1	99.776	0.134	331	91596	10800	100	0	365	108527	6240
	1.05	99.936	0.05	475	287709	10800	100	0	10	315231	1132
	1.1	99.996	0.004	48	452787	10800	100	0	131	595547	3620
	1.2	99.982	0.018	12	783728	10800	100	–	–	–	–
30	1	99.964	0.036	1041	91583	10800	100	–	–	–	–
	1.05	99.991	0.009	579	273906	10800	100	–	–	–	–
	1.1	99.988	0.011	211	470354	10800	100	–	–	–	–
	1.2	99.997	0.003	19	762189	10800	100	–	–	–	–
35	1	100	0	23	77066	581	100	–	–	–	–
	1.05	100	0	30	237368	1192	100	–	–	–	–
	1.1	100	0	34	413932	2146	100	–	–	–	–
	1.2	100	0	79	728364	6539	100	–	–	–	–

However, the gain for the same driver tolerance decreases when p increases or decreases. This is expected since, for small p, driver tolerance cannot compensate the scarcity of the refueling stations to generate a feasible path whereas in the case where the refueling stations are widespread, there is not so much need for a long refueling detour.

5. Conclusions

In this paper, we present the refueling station location problem with routing (RSLP-R) for locating a given number of refueling stations for alternative fuel vehicles in a road network so as to maximize the total flow covered. Driver deviations from the shortest path up to a certain tolerance value are considered as alternative paths including non-simple ones. The problem is practically important due to fact that the adoption of AFVs strongly depends on the availability of the refueling infrastructure and the high cost of this initial investment motivates the efforts for the best use of limited resources. It is theoretically challenging because the problem is NP-Complete and previous formulations of similar problems failed to handle large networks due to their modeling structures. The most important contribution of this paper to the literature is extending the size of the solvable problems. Rather than pregenerating all the path alternatives before solving the model, we apply a branch and price solution algorithm which enables us to handle problems that were not of manageable size by previous works on similar contexts. Our algorithm also decreased the solution times with respect to previous studies which is another major contribution to the literature. The efficiency of the solution technique is mainly due to the path-segment definition in our formulation. Such a formulation enables us to relax the simple path assumption and admits a very natural representation of the side constraints on the path. Observe that our path-segment formulation is general enough to accommodate a wide array of side constraints on the vehicle routes other than refueling and total distance. As such, our formulation could be quite useful to model more realistic and complex problems as well as problems in completely different contexts such as telecommunications and transportation (Yıldız & Karasan, 2014).

In its current form, our methodology does not accommodate for the capacities of the refueling stations which is in the scope of recent works such as Upchurch, Kuby, and Lim (2009a) and Jung et al. (2014). Driver's different preferences such as stopping tolerance can also be accommodated into this framework and the number of refueling stops on a given path might also be limited by a constant. In the context of this study, we also consider only the refuelability of the O–D pairs, but the costs are not taken into account. With the help of efficient solution algorithms for RSLP-R and its variants, future research in this area might be directed towards handling the node and arc capacities in the network, the stopping tolerance of the drivers, and/or different nodal pricing between infrastructure setup costs. Each of these research directions will bring more realism at the expense of additional computational complexity.

Acknowledgments

References

Adler, J. D., Mirchandani, P. B., Xue, G., & Xia, M. (2014). The electric vehicle shortest-walk problem with battery exchanges. *Networks and Spatial Economics* (pp. 1–19).

Arslan, O., Yıldız, B., & Karaşan, O. E. (2014a). Impacts of battery characteristics, driver preferences and road network features on travel costs of a plug-in hybrid electric vehicle (phev) for long-distance trips. *Energy Policy, 74*(0), 168–178.

Arslan, O., Yıldız, B., & Karaşan, O. E. (2014b). Minimum cost path problem for plug-in hybrid electric vehicles. Department of Industrial Engineering, Bilkent University, Turkey.

Artmeier, A., Haselmayr, J., Leucker, M., & Sachenbacher, M. (2010). The shortest path problem revisited: optimal routing for electric vehicles. In *Ki 2010: Advances in artificial intelligence* (pp. 309–316). Springer.

Bapna, R., Thakur, L. S., & Nair, S. K. (2002). Infrastructure development for conversion to environmentally friendly fuel. *European Journal of Operational Research, 142*(3), 480–496.

Bektaş, T., & Laporte, G. (2011). The pollution-routing problem. *Transportation Research Part B: Methodological, 45*(8), 1232–1250.

Berman, O., Larson, R. C., & Fouska, N. (1992). Optimal location of discretionary service facilities. *Transportation Science, 26*(3), 201–211.

Capar, I., Kuby, M., Leon, V. J., & Tsai, Y.-J. (2013). An arc coverpath-cover formulation and strategic analysis of alternative-fuel station locations. *European Journal of Operational Research, 227*(1), 142–151.

Carlyle, M. W., & Wood, K. R. (2005). Near-shortest and k-shortest simple paths. *Networks, 46*(2), 98–109.

Chung, S. H., & Kwon, C. (2015). Multi-period planning for electric car charging station locations: a case of korean expressways. *European Journal of Operational Research, 242*(2), 677–687.

Church, R., & ReVelle, C. (1974). The maximal covering location problem. *Papers in Regional Science, 32*(1), 101–118.

Erdoğan, S., & Miller-Hooks, E. (2012). A green vehicle routing problem. *Transportation Research Part E: Logistics and Transportation Review, 48*(1), 100–114.

Goodchild, M. F., & Noronha, V. T. (1987). Location-allocation and impulsive shopping: the case of gasoline retailing. *Spatial analysis and location-allocation models* (pp. 121–136).

Hodgson, M. J. (1990). A flow-capturing location-allocation model. *Geographical Analysis, 22*(3), 270–279.

Huang, Y., Li, S., & Qian, Z. (2015). Optimal deployment of alternative fueling stations on transportation networks considering deviation paths. *Networks and Spatial Economics* (pp. 1–22).

Jung, J., Chow, J. Y., Jayakrishnan, R., & Park, J. Y. (2014). Stochastic dynamic itinerary interception refueling location problem with queue delay for electric taxi charging stations. *Transportation Research Part C: Emerging Technologies, 40*, 123–142.

Kang, J. E., & Recker, W. (2014). Strategic hydrogen refueling station locations with scheduling and routing considerations of individual vehicles. *Transportation Science* . doi:10.1287/trsc.2014.0519.

Khuller, S., Malekian, A., & Mestre, J. (2007). To fill or not to fill: the gas station problem. In *Algorithms–esa 2007* (pp. 534–545). Springer.

Kim, J.-G., & Kuby, M. (2012). The deviation-flow refueling location model for optimizing a network of refueling stations. *International Journal of Hydrogen Energy, 37*(6), 5406–5420.

Kim, J.-G., & Kuby, M. (2013). A network transformation heuristic approach for the deviation flow refueling location model. *Computers & Operations Research, 40*(4), 1122–1131.

Kitamura, R., & Sperling, D. (1987). Refueling behavior of automobile drivers. *Transportation Research Part A: General, 21*(3), 235–245.

Kuby, M., Araz, O. M., Palmer, M., & Capar, I. (2014). An efficient online mapping tool for finding the shortest feasible path for alternative-fuel vehicles. *International Journal of Hydrogen Energy, 39*(32), 18433–18439.

Kuby, M., & Lim, S. (2005). The flow-refueling location problem for alternative-fuel vehicles. *Socio-Economic Planning Sciences, 39*(2), 125–145.

Kuby, M., & Lim, S. (2007). Location of alternative-fuel stations using the flow-refueling location model and dispersion of candidate sites on arcs. *Networks and Spatial Economics, 7*(2), 129–152.

Kuby, M., Lines, L., Schultz, R., Xie, Z., Kim, J.-G., & Lim, S. (2009). Optimization of hydrogen stations in florida using the flow-refueling location model. *International Journal of Hydrogen Energy, 34*(15), 6045–6064.

Kuby, M. J., Kelley, S. B., & Schoenemann, J. (2013). Spatial refueling patterns of alternative-fuel and gasoline vehicle drivers in los angeles. *Transportation Research Part D: Transport and Environment, 25*, 84–92.

Lim, S., & Kuby, M. (2010). Heuristic algorithms for siting alternative-fuel stations using the flow-refueling location model. *European Journal of Operational Research, 204*(1), 51–61.

Lin, S. H., Gertsch, N., & Russell, J. R. (2007). A linear-time algorithm for finding optimal vehicle refueling policies. *Operations Research Letters, 35*(3), 290–296.

Lübbecke, M. E., & Desrosiers, J. (2005). Selected topics in column generation. *Operations Research, 53*(6), 1007–1023.

Megiddo, N., Zemel, E., & Hakimi, S. L. (1983). The maximum coverage location problem. *SIAM Journal on Algebraic Discrete Methods, 4*(2), 253–261.

Melaina, M., & Bremson, J. (2008). Refueling availability for alternative fuel vehicle markets: sufficient urban station coverage. *Energy Policy, 36*(8), 3233–3241.

Melaina, M. W. (2003). Initiating hydrogen infrastructures: preliminary analysis of a sufficient number of initial hydrogen stations in the us. *International Journal of Hydrogen Energy, 28*(7), 743–755.

Melaina, M. W. (2007). Turn of the century refueling: A review of innovations in early gasoline refueling methods and analogies for hydrogen. *Energy Policy, 35*(10), 4919–4934.

MirHassani, S. A., & Ebrazi, R. (2013). A flexible reformulation of the refueling station location problem. *Transportation Science, 47*(4), 617–628.

Nicholas, M. A., Handy, S. L., & Sperling, D. (2004). Using geographic information systems to evaluate siting and networks of hydrogen stations. *Transportation Research Record: Journal of the Transportation Research Board, 1880*(1), 126–134.

Nicholas, M. A., & Ogden, J. (2006). Detailed analysis of urban station siting for california hydrogen highway network. *Transportation Research Record: Journal of the Transportation Research Board, 1983*(1), 121–128.

Recker, W. W. (1995). The household activity pattern problem: General formulation and solution. *Transportation Research Part B: Methodological, 29*(1), 61–77.

Romm, J. (2006). The car and fuel of the future. *Energy Policy, 34*(17), 2609–2614.

Schneider, M., Stenger, A., & Goeke, D. (2014). The electric vehicle-routing problem with time windows and recharging stations. *Transportation Science, 48*(4), 500–520. doi:10.1287/trsc.2013.0490.

Simchi-Levi, D., & Berman, O. (1988). A heuristic algorithm for the traveling salesman location problem on networks. *Operations Research, 36*(3), 478–484.

Soedarmadji, E., & McEliece, R. J. (2007). Optimal worst-case qos routing in constrained awgn channel network. In *Communications, 2007. icc'07. ieee international conference on* (pp. 153–157). IEEE.

Suzuki, Y. (2008). A generic model of motor-carrier fuel optimization. *Naval Research Logistics (NRL), 55*(8), 737–746.

Upchurch, C., & Kuby, M. (2010). Comparing the p-median and flow-refueling models for locating alternative-fuel stations. *Journal of Transport Geography, 18*(6), 750–758 Special Section on Alternative Fuels and Vehicles.

Upchurch, C., Kuby, M., & Lim, S. (2009a). A capacitated model for location of alternative-fuel stations. *Geographical Analysis, 41*(1), 127–148.

Upchurch, C., Kuby, M., & Lim, S. (2009b). A model for location of capacitated alternative-fuel stations. *Geographical Analysis, 41*(1), 85–106.

U.S. Census Bureau (2010). Urban, urbanized area, urban cluster, and rural population, 2010 and 2000: United states. Accessed May 18.05.14 http://www.census.gov/geo/reference/ua/urban-rural-2010.html.

Wang, Y.-W., & Lin, C.-C. (2009). Locating road-vehicle refueling stations. *Transportation Research Part E: Logistics and Transportation Review, 45*(5), 821–829.

Wang, Y.-W., & Lin, C.-C. (2013). Locating multiple types of recharging stations for battery-powered electric vehicle transport. *Transportation Research Part E: Logistics and Transportation Review, 58*(0), 76–87.

Wang, Y.-W., & Wang, C.-R. (2010). Locating passenger vehicle refueling stations. *Transportation Research Part E: Logistics and Transportation Review, 46*(5), 791–801.

Yıldız, B. & Karasan, O. E. (2014). Regenerator location problem in flexible optical networks. Under Review in Operations Research.