TRANSPORTATION SCIENCE
Vol. 47, No. 4, November 2013, pp. 617–628
ISSN 0041-1655 (print) | ISSN 1526-5447 (online)
http://dx.doi.org/10.1287/trsc.1120.0430

A Flexible Reformulation of the Refueling Station Location Problem

S. A. MirHassani, R. Ebrazi
Faculty of Mathematics and Computer Science, Amirkabir University of Technology, Tehran, Iran
{a_mirhassani@aut.ac.ir, r.ebrazi@aut.ac.ir}

Serious environmental and economic problems of using fossil fuels in transportation sections force managers to think of alternative fuels such as hydrogen, ethanol, biodiesel, natural gas, or electricity. Meanwhile, lack of fuel network infrastructures is a major problem, which needs to be investigated considering the number and optimal location of alternative fuel stations. In the literature, two different flow-based demand modeling concepts (the maximum cover and set cover) have been proposed for solving this problem. Because of the huge number of combinations of fuel stations for covering the flow of each path, the models are impractical for the real size problems. In this paper, the flow refueling location model was reformulated and a flexible mixed-integer linear programming model was presented, which was able to obtain an optimal solution much faster than the previous set cover version. The model also could be solved in the maximum cover form in a reasonable time on the large-sized networks.

Key words: fuel station; location problem; maximum flow; mixed-integer programming
History: Received: August 2011; revisions received: November 2011, February 2012; accepted: March 2012. Published online in *Articles in Advance* September 5, 2012.

1. Introduction

For the last few decades, the dominance of fossil fuels in the transportation sector has led to serious environmental and economic damages. As a result, global warming and greenhouse gases, urban air pollution, reduction in natural resources and oil supplies, population growth, increase in energy demand, high oil prices, and energy security are the problems with which the world is faced in the future. Following electricity and heat generation, transportation was the second largest sector that caused 22% of global CO_2 emission in 2008 (International Energy Agency 2010). In most energy-consuming industrialized countries, 96% of transportation energy is derived from petroleum fuels (InterAcademy Council 2007). These facts show an urgent need for using alternative fuels such as biodiesel, ethanol, hydrogen, electricity, and natural gas (Shukla, Pekny, and Venkatasubramanian 2011; Melaina and Bremson 2008; Lund and Clark 2008).

There are two major barriers that obstruct the wide application of alternative fuels. The first difficulty is lack of refueling infrastructures due to high capital costs of building a station and handling equipment (Shukla, Pekny, and Venkatasubramanian 2011; Melaina and Bremson 2008; Kuby et al. 2009; Rodrigue, Comtois, and Slack 2009; Melendez 2006; Romm 2006; Kuby and Lim 2005; Melaina 2003). Moreover, consumers tend to buy alternative fuel vehicles (AFV) when there is a sufficient number of refueling stations (Bento 2008); consequently, AFV manufactures are disinclined to produce vehicles when there is no market for it. This situation makes a chicken-and-egg problem (Kuby and Lim 2005; Melaina 2003; Huétink, van der Vooren, and Alkemade 2010; Wang and Wang 2010; Melaina 2007; Melaina and Ross 2000; Leiby and Rubin 2004; Sperling 1990). The second obstacle is the limited driving range of AFV that is due to lower energy density, efficiency of alternative energy sources compared with gasoline, and size and weight limitations of AVF, which enforce it to have a smaller on-board fuel container or lighter battery (Shukla, Pekny, and Venkatasubramanian 2011; Wang and Wang 2010; Lim and Kuby 2010; Wang and Lin 2009; Kuby, Lim, and Upchurch 2005).

These difficulties and high capital cost of refueling station infrastructure demonstrate the importance of research on the number and location of refueling stations for developing a network that maximizes the coverage of demand, and considers the limited capability of AFV. This topic is studied through facility location models to basically deal with these difficulties. In the remaining subsections, the related literature is reviewed for finding the source of demand and applying an approach for the best deployment of the refueling station network.

1.1. Sources of Demand

Demand distribution plays a key role in optimizing the arrangement of fuel stations. The current gasoline consumption can be seen as a guide for future alternative fuel consumption. In this view, Melaina (2003) suggested four simple criteria for the effective location of early hydrogen stations. He recommended locating stations near high traffic volumes, high profit areas, potential first AFV buyer regions, and at the locations that provide fuel for vehicles during long-distance trips. Melaina (2003) also mentioned metropolitan areas and locations along interstates as places that can fulfill these criteria. In addition, he used three approaches for estimating a sufficient number of initial hydrogen fueling stations based on (1) the percentage of existing gasoline stations, which was suggested by a previous survey on fuel availability before the purchase of diesel vehicles (Sperling and Kitamura 1986; Sperling and Kurani 1987), (2) metropolitan land area, and (3) lengths of principal arterial, which accounted for regional driving intensity patterns and roads and had the highest vehicle kilometer traveling (VKT) volumes per kilometer of road. Melaina (2003) believed that the arterial road approaches appear more consistent with his suggested criteria because the land area approach adopts a uniform distribution of stations regardless of population or traffic pattern of a region, and the fact that the density of station depends on several factors such as market entry barriers, profit margins, local retail, and real potential demand (Melaina and Bremson 2008). The last subject was the field of study by Melaina and Bremson (2008), which suggested a sufficient level of station coverage in terms of station density that can fulfill the refueling needs of the general population in urban areas. This sufficient coverage level, which can be conceptually considered as the upper bound of alternative fuel stations, is based on a subset of urban areas served by relatively sparse station networks while compared with cities with similar size and population density.

Nicholas (2010) used regression analysis to show that, in particular, VKT was a good predictor of demand; but, it may be over-predicted in the central business district. This was due to high gasoline prices, a limited number of stations or traffic congestion that were conducted, and the overcount of VKT in these places. Nicholas (2010) also mentioned population as a relative estimator of demand, because when they divide the analysis region into large sub-regions, variations in the number of people generally correspond to variations in the demand. However, when subregion divisions were smaller, the correlation between the number of people and demand for gasoline was less strong. They believed that one possible explanation could be that people do not necessarily obtain fuel from the station near their homes. Thus they suggested considering the distance over which the population had fuel demand. In addition, Nicholas (2010) found that the amount of sold gasoline decreased as the distance from highways increased while the population was not as dense. So, he introduced a new population-derived metric, "population traffic," for estimating the traffic near highways. By regression analysis, Nicholas (2010) concluded that this new metric was in more correlation with the sold fuel than population or VKT. Therefore he suggested that highway entrances and along the highways can be primary candidates for planners to locate primary stations.

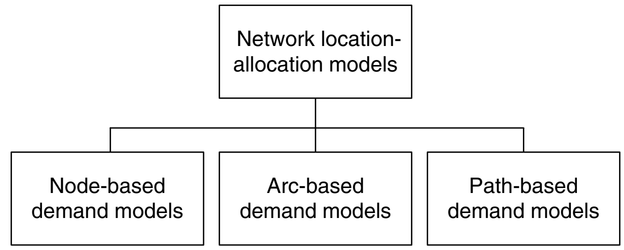

Figure 1 Division of Network Location-Allocation Models Based on Demand Considerations

1.2. Methods of Demand Coverage

In the location-allocation problem on the network and specifically fuel station location problems, there are three strategies to deal with demand (Upchurch and Kuby 2010) (Figure 1).

Traditional facility location models such as the p-median (Hakimi 1964; ReVelle and Swain 1970), fixed charge (Daskin 2008), maximal cover, set cover (Toregas et al. 1971), and p-center, which are five well-known node-based problems, all consider demand as aggregated in nodes while facilities can be located anywhere on the nodes of the network or points on arcs. The candidate facility distances from each demand node are computed as the shortest paths in a graph. The objectives of node-based demand models are simply to minimize the sum of weighted distances between the demand nodes on the graph and the nearest facility, minimize fixed facility and transportation costs (i.e., sum of distances between the demand nodes and the nearest facility multiplied by the transportation cost per distance unit per demand item), maximize the number of demand nodes that can be served within a specified distance by locating a fixed number of facilities, minimize the number of facilities to cover all demand nodes, or minimize the maximum distance between any demand and its nearest facility.

There are various facility location models that have used the concept of p-median and considered workplace or home as the origin of single-purpose refueling travel for locating refueling stations (Nicholas and Ogden 2006; Nicholas, Handy, and Sperling 2004; Goodchild and Noronha 1987).

Driven by the notation "where you drive more is where you more likely need refueling," Lin et al. (2008) supposed that any point along a road with higher VMT was more likely to be the origin of the refueling trip; actually, they quantified the aggregated fuel demand of a node by the fuel consumption of road segment and pointed to that node while hypothetically minimizing the total fuel-travel-back time (i.e., the time motorists need to reach the closest fuel stations). This model was transformed to a median problem in which the magnitude of the demand of a node was indicated by the spatial distribution of VMT.

The next strategy is to relate demand (traffic flow or population) to arcs of the road network and assess their magnitude of demand separately. The model finds paths between predetermined nodes of the network (usually, population centers) by balancing the demand coverage of the path with the cost of building an adequate number of fuel stations along the connecting arcs. The number of stations along each arc is proportional to the arc lengths and demand.

From the viewpoint of this class of model, Current, Revelle, and Cohon (1985) introduced the maximum covering/shortest path problem and Bapna, Thakur, and Nair (2002) extended it as a base model to the maximum covering/shortest spanning subgraph problem. This model used population centers as demand sources and decided about enabling an arc by locating the required number of fuel stations on it. An enabled arc covered the population of its head and tail nodes and all other nodes that lay within certain distances, the covering distance, from any point on it. This means that it related population demand to arcs. Model formulation constructed a spanning subgraph between cities through two conflicting objectives of minimizing the cost of enabling arc and maximizing population coverage. This model had two main disadvantages: (1) the spanning subgraph may lead to a roundabout and uncommon paths due to its dependence of the weights on the objectives and (2) it maximizes the population along the enabled arcs, not the traffic volume that can refuel along their shortest path (Kuby and Lim 2005), which is a main demand parameter, as discussed in Nicholas (2010).

Another strategy relates the demand of traffic flow to the path and covers them by allocating one or a combination of facilities to the nodes along this path.

Hodgson (1990) and Berman, Larson, and Fouska (1992) were the first researchers who used this schema and developed a new kind of model; namely, the flow capturing location model (FCLM). Although the concept of their model was similar to the maximum cover problem for locating the limited number of facilities (Church and Revelle 1974), unlike most classical node-based models, it assumed the traffic flow of the predetermined path from origins to destinations (OD) as a unit of demand to be covered and, if the flow passed through the location of the facility, it was considered to be "captured." To explain the reason for using path flow as the demand, they argued that many types of facilities could be the second purpose of people's trips. Since its first presentation in 1990, more than 30 different flow interceptions location models have appeared to formally characterize a wide spectrum of consumer desires and needs (Zeng, Castillo, and Hodgson 2008). It has been also extended in a number of ways; Hodgson and Rosing (1992) proposed a hybrid model (flow capturing plus p-median). Berman, Hodgson, and Krass (1995) addressed flow deviation from the shortest path. The competitive view of flow interception was investigated by Berman and Krass (1998). Zeng, Castillo, and Hodgson (2008) introduced some changes in the objective functions and/or assumptions. More details and a good review of the related literature can be found in Zeng, Castillo, and Hodgson (2008), Hodgson (1998), and Berman, Hodgson, and Krass (1995).

The limited driving range of alternative fuel makes vehicles refuel more than once on their path, which is left out in FCLM because it captures the flow if, at least, one facility is located anywhere along the path. To consider this parameter, Kuby and Lim (2005) extended FCLM and developed the flow-refueling location model (FRLM). This model used combinations of more than one station to cover a round-trip travel on longer paths. Kuby and Lim (2005) solved this problem in two phases: (1) they used an algorithm to pregenerate all the combinations of facilities that can refuel each path and (2) they located a refueling station using a mixed-integer linear programming (MILP). The main difficulty that actually limited the applicability of this approach was the enormous number of facility combinations along the long paths with several nodes. It is impractical to generate and solve the problem even for medium-sized networks. To handle larger networks, Lim and Kuby (2010) developed some heuristic algorithms (the greedy-adding, greedy-adding with substitution and genetic algorithms) and applied them to a small network with 25 nodes and, in the Florida state network, with more than 300 candidate nodes. Although the comparison of the performance of the heuristic solution with the exact solution of MILP in the first small network showed that the algorithms performed well, they found it impossible to do this comparison with the Florida state network because the algorithm (on Pentium 4, 3.2 GHz, 1 GB RAM computer) took 13 hours to generate the valid combinations for the first 39 paths of the total 2,701 paths.

This problem motivated Capar and Kuby (2011) to present a radically different mixed-binary-integer programming formulation that did not require pregeneration of presented feasible station combinations, which reduced the time required for solving FRLM problems. Although this new formulation created a more complex model, it solved the FRLM to optimality as fast as the published heuristic algorithms (Lim and Kuby 2010) and improved the solution quality.

There are other papers based on FRLM, which have studied subjects such as dispersion of candidate sites on arcs (Kuby and Lim 2007), capacitated facilities (Upchurch, Kuby, and Lim 2009), and comparison of p-median and FRLM (Upchurch and Kuby 2010).

Wang and Lin (2009) introduced a flow-based model considering the driving range assumption based on the concept of set-covering problems, which was different from FRLM in this view; so, it used only a distance matrix, which was easier to obtain as compared with the OD flow. The major innovation of this model was in using a special formulation to implement the driving range logic implicitly in the MILP model. Although using this formulation omitted the first phase of FRLM for finding a valid combination, a real-world large-scale mixed-integer problem (MIP) problem that included many constraints was not easy to solve using an exact algorithm. So, uncapacitated stations were considered for reducing the number of constraints (Wang and Lin 2009). Wang and Wang (2010) also developed a hybrid model by combing Wang and Lin's (2009) model and the classic set-covering model (Toregas et al. 1971) to simultaneously consider the nodal population demands and path demands.

Based on the problems of the refueling location model in implementing the driving range logic, this paper proposed a new reformulation of flow-based models with completely the same assumptions of FRLM and nearly similar Wang and Lin's (2009) model assumptions. The main contributions of this model were reduction of solution time for set-covering problems and the ability for solving real size maximum cover problems (i.e., FRLM) for the first time. Another important consequence of this new formulation was the abilities for considering additional assumptions about fuel station type and driver behavior, which was missing in previous models for reducing the number of constraints.

In the remaining sections, the formulation of this model is first introduced for single path and then, using this model, in §2, multipath formulation is presented for complete network. The test networks are introduced in §3 and the solution time of the new model is compared with other models. Section 4 provides conclusions and, finally, §5 suggests topics for future research.

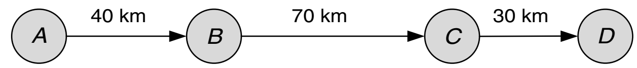

Figure 2 An Example of Simple Path

2. Model Formulation

2.1. Common Sense in Refueling

First, in this section, refueling logic is described; then, these rules are used in §2.2 to construct an expanded network for allocating stations on one path.

Figure 2 illustrates a simple example path with four nodes A, B, C, D. It is assumed that a vehicle with the driving range of $R = 100$ km wants to have a round trip starting from A without running out of fuel. The goal is to indicate all valid combinations of stations on its path.

If there is a fuel station at A and the vehicle can fill up its tank, then it could reach at most a point between nodes B and C, which means that there must be a station on node B. It demonstrates the first simple rule that the distance between two stations A and C has to be less than the driving range.

The second rule is about the amount of fuel in the tank at OD. If there is a station at the origin, it can be assumed that the vehicle has a full tank at the beginning. If not, it is not reasonable that the vehicle could have a full or empty tank because, on the previous trip, the vehicle has certainly traveled over some distance from the last station to reach the origin; therefore it would not have a full tank. On the other hand, drivers usually consider a fair amount of fuel to be able to reach the first station on their next trip; thus an empty tank assumption in the origin does not appear to be reasonable. Kuby and Lim (2005) assumed that the tank was half-full in the origin. This seems more reasonable and creates a sufficient condition for a combination to be valid for a round trip if it is only realistic in one direction. For example, it is enough to evaluate a combination for the path from A to D and, if the vehicle has a certain condition at D, it can travel back to A. Also, there is no more need to test the D–A direction. In fact, if a vehicle can reach the first station with a half-full tank, then it can fill up at the same station in the opposite direction when it is returning to the origin. In the example presented here, this rule does not prove the combination $\{C, D\}$ because the vehicle cannot reach C with just a half-full tank. The same argument can be made for the amount of fuel in the vehicle tank in the destination; meaning that, if there is no fuel station at destination D, the vehicle must reach the last station with at least a half-full tank. This fuel is enough for reaching C on the return path. Using these rules, the combinations $\{A, B, C, D\}$, $\{A, B, C\}$, $\{A, B, D\}$, $\{B, C, D\}$, $\{B, D\}$, and $\{B, C\}$ can be approved as valid combinations for refueling the instance path.

2.2. Station Allocation Problem—One Path Only

As an intermediate step toward developing the main model, the simpler problem of finding the best valid combination along a single path was presented. The approach was to construct an expanded network, which transformed the problem to the framework of finding a path that was covered with fuel stations at the minimum cost. In this section, first, the set cover version of the FRLM, which made more sense for one path was presented; actually the intention was to describe how the valid combination can be found for one path while using this approach either in set cover or maximum cover concepts made no differences. As will be seen later, this basic model had straight applications in constructing the main model of the station allocation problem. To illustrate the present modeling approach, the previous example was used and, simultaneously, its corresponding expanded network was constructed.

Here, some basic notations are presented before defining the expanded network $G = (\hat{N}^q, \hat{A}^q)$, N^q is the set of nodes existing on path q, $\hat{N}^q$ is the set of nodes existing in the expanded network and, similarly, $\hat{A}^q$ is the set of arcs belonging to the expanded network, for any two nodes i and j on the shortest path (which does not need any successive nodes on the path) $d_q(i, j)$ denotes the length of the subpath between them that lies on the referenced path q and, finally, $\text{ord}_q(i)$ defines an ordering index in the path sequence. For example, node C on the path q: A–B–C–D gets the ordering index $\text{ord}_q(C) = 3$. The four steps for constructing this expanded network are as follows:

Step 1. Add two new nodes to path nodes, a source node s (before the origin) and a sink node t (after the destination). Connect s and t to the origin node and destination node, respectively. In the example, $\hat{N}^q$ contains the nodes of path q and $\{s, t\}$, i.e., $\{s, A, B, C, D, t\}$. Also, arcs (s, A) and (D, t) are added to the expanded network arc $\hat{A}^q$.

Step 2. Connect the source node s to any other node; say, i of path q, if it is possible to begin at the origin node and get to node i with half of a tank or less. In the example, this means that

$$\forall\, i \in N^q\colon\ d_q(A, i) \le \frac{R}{2} \ \rightarrow\ (s, i) \in \hat{A}^q.$$

Accordingly, its expanded network includes arc (s, B).

Step 3. Connect any nodes i of path q to node t whenever the vehicle can reach the destination node of path with a half-full tank or less. In this step, a similar notation can be used in the example path:

$$\forall\, i \in N^q\colon\quad d_q(i, D) \le \frac{R}{2} \ \rightarrow\ (i, t) \in \hat{A}^q.$$

As a result, the expanded network contains arc (C, t).

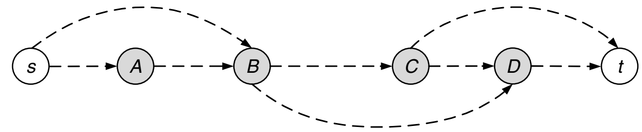

Figure 3 **Expanded Network for the Example**

Step 4. Finally, connect each node i of path q to any other node j if the ordering index of i is less than the ordering index of j and the vehicle is able to start from i with a full tank and reach j. So, for the current example:

$$\forall\, i, j \in N^q\colon\quad (\text{ord}_q(i) < \text{ord}_q(j)) \wedge (d_q(i, j) \le R) \rightarrow\ (i, j) \in \hat{A}^q.$$

This implies that the expanded network contains arcs (A, B), (B, C), (B, D), and (C, D). In this network, each arc corresponds to the consecutive feasible refueling and each directed path from the source s to the sink t corresponds to a feasible combination of fuel station that can refuel the path. Actually, Steps 1–3 represent the second vehicle's refueling rule and Step 4 satisfies the first vehicle's refueling rule. The validity of this formulation is easy to establish by showing a one-to-one correspondence between all these s–t paths and valid combinations of fuel stations. Figure 3 shows the resulting expanded network in this case. For example, the path s–B–D–t corresponds to the valid combination $\{B, D\}$.

To find the best s–t path, i.e., a path through an optimization model, node s is assigned number 1 representing its virtual supply, and node t is assigned number -1 representing its virtual demand. All other nodes are assigned number 0 considering the transshipment node. In addition to what has been mentioned, a one-path refueling station location model uses the following parameter and decision variables:

Parameter
c_i: The cost of building a station at node i.

Decision variables
y_i: 1 if a station is located at node i; otherwise 0.
x_{ij}^q: The flow on an arc (i, j), which belongs to expanded network arcs of path q (i.e., $\hat{A}^q$).

The model is formulated as follows:

$$\min\ Z = \sum_{i \in N^q} c_i y_i \tag{1}$$

s.t.

$$\sum_{\{j \mid (i,j) \in A^q\}} x_{ij}^q - \sum_{\{j \mid (j,i) \in A^q\}} x_{ji}^q = \begin{cases} 1 & i = s, \\ -1 & i = t, \\ 0 & i \ne s, t, \end{cases} \qquad \forall\, i \in \hat{N}^q, \tag{2}$$

$$\sum_{\{j \mid (j,i) \in A^q\}} x_{ji}^q \le y_i \quad \forall\, i \in N^q, \tag{3}$$

$$x_{ij}^q \ge 0 \quad \forall\, (i, j) \in \hat{A}^q, \tag{4}$$

$$y_i \in \{0, 1\} \quad \forall\, i \in N^q. \tag{5}$$

The objective function (1) minimizes the cost of building refueling stations. Constraints (2) are the mass balance constraints, which state that the outflow minus inflow must equal the virtual supply and demand of the node. Constraints (3) let the flow pass through a node only when a fuel station is located at that node. Constraints (4) ensure that the flow on each arc is greater than or equal to zero. Constraints (5) introduce the facility location variables y_i as binary variables. As a result, by solving the presented model for the designed expanded network, the resulted path demonstrated a valid and the best combination, i.e., the least cost valid combination of the refueling station. The valid paths in the presented example are q: *s-B-D-t* or *s-B-C-t*, which correspond to combinations $\{B, D\}$ and $\{B, C\}$.

It is worth mentioning that this method was based on using the same shortest path on a round trip; in other words, if the driver decides to use another path on the way back, the latter path also should be considered as the covering path.

2.3. Main Model

The main multipath model formulation was derived from the one-path model by adding summations as needed and adding constraints to integrate paths that share common nodes. Here, some additional sets are introduced:

Q: Set of all candidate paths to be covered (usually the shortest paths)

N: Set of all path-related nodes ($N = \bigcup_{q\in Q} N^q$)

Q_i: Subset of Q that contains all the paths passing node i.

The set cover form of the main model (referred to as Model 1) was formulated as

$$\text{(1)}\quad \min \quad Z = \sum_{i\in N} c_i y_i \tag{6}$$

$$\text{s.t.}\quad \sum_{\{j|(i,j)\in A^q\}} x_{ij}^q - \sum_{\{j|(j,i)\in A^q\}} x_{ji}^q = \begin{cases} 1 & i=s,\\ -1 & i=t,\\ 0 & i\neq s,t, \end{cases} \quad \forall q\in Q,\ \forall i\in \hat{N}^q, \tag{7}$$

$$\sum_{\{j|(j,i)\in A^q\}} x_{ji}^q \le y_i \quad \forall i\in N,\ \forall q\in Q_i, \tag{8}$$

$$x_{ij}^q \ge 0 \quad \forall q\in Q,\ \forall (i,j)\in \hat{A}^q, \tag{9}$$

$$y_i \in \{0,1\} \quad \forall i\in N. \tag{10}$$

The objective function (6) minimized the cost of building refueling stations on all paths. The mass balance constraints (7) were extended to include all paths to be covered in Q. The domain of constraints (8) was changed in a way in which all the paths with common nodes can make a site active; say, y_i. This prevented the excessive cost of building fuel stations and integrating the best refueling combination. Constraints (9) and (10) define flow variables x_{ij}^q nonnegative and candidate location y_i as binary variables.

Models in Wang and Lin (2009) and Wang et al. (2010) included a large number of binary variables for implementing the vehicle's refueling logic, which made it very difficult to be solved, even for a medium-sized network. Definitely, the main advantage of this new reformulation was to apply relaxed variable x_{ij}^q for implementing mass balance constraints, which satisfied the driving range assumption and implicitly indicated a valid combination of fuel stations. As will be seen in the following section, the new model was solved dramatically faster using a classic branch-and-bound solver.

Planners often want to test different strategies and select the best. To this end, a flexible model was presented that was able to simply switch between the set-covering forms with the objective of minimizing the cost of building stations and a flow-based maximum coverage model with the objective of maximizing the total flow volume.

Before presenting the FRLM new reformulation by the mentioned expanded network structure, it was essential for the designed expanded network for the path q in the previous section to include a new arc (s, t). This new arc gave the selection ability of covering a specific path (or not), so that the model could decide on selecting a path to cover (or not) considering the p facility or the budget constraint. When this flow variable of this new arc x_{st}^q equalled one mass balance constraint satisfied without any active fuel station, then the flow on this path could not be covered and vice versa. The above statements lead to the use of this variable as a decision variable.

To change the model to a flow-base maximum coverage, the following relations can be simply inserted in the objective function (6):

$$Z = \sum_{q\in Q} f_q (1 - x_{st}^q), \tag{11}$$

where f_q is traffic flow volume on each path, that is usually calculated in trips (Lim and Kuby 2010; Shukla, Pekny, and Venkatasubramanian 2011) but also can be measured as VKT by multiplying path distance by trips (Upchurch, Kuby, and Lim 2009) and adding the p facility constraint, 12, to the main model constraints.

$$\sum_{i\in N} y_i = p \tag{12}$$

It ensures that p facilities are exactly built. If there is different types of facilities with different costs, this constraint can be generalized to the budget constraint,

$$\sum_{i\in N} c_i y_i \le B. \tag{13}$$

It requires that the cost of facilities does not exceed the budget B that planners consider for developing infrastructures.

3. Performance Analysis

In this section, the solution time of the presented model is compared with that of Wang and Lin's (2009) model using two different networks. A 25-node sample network, used by Lim and Kuby (2010), Kuby and Lim (2007, 2005), Kuby, Lim, and Upchurch (2005), Hodgson (1990), Simchi-Levi and Berman (1988), and several random networks were generated by an algorithm and described in §3.2.

Kuby and Lim (2005) described situations in which some paths may not be covered by facilities just located at its nodes; however, candidate sites can be added on the arcs and these paths can be refueled by facilities located at these midarc nodes. Kuby and Lim (2007) proposed three methods for adding candidate sites to the arcs. For the comparison purpose in this paper, $t(i, j)$ midarc nodes were uniformly added (i.e., with identical space between them) along the arc (i, j) in each network using the following formula:

$$t(i,j) = \begin{cases} \left\lceil \dfrac{\text{length}(i,j)}{r} \right\rceil - 1 & \text{length}(i,j) > r, \\ 0 & \text{length}(i,j) \le r, \end{cases} \tag{14}$$

where *length*(i, j) is the length of arc (i, j) and r is an input constant. This method is motivated by the idea of preventing a network from having any longer arc than r without any candidate sites.

To show the efficiency of the new MIP formulation for the exact solution of flow-based maximum coverage problem, a moderate-sized network was generated and solved. The MIP model was generated using AIMMS 3.9 and solved using the CPLEX 12.1 solver on a system with Core(TM) 2 Duo processor (2.53 GHz) and 2.95 GB RAM.

3.1. Test Network with 25 Nodes

Figure 4 shows a 25-node test network; all 25 nodes were considered origins, destinations, and candidate sites, which made 300 ($25 \times (25-1)/2$) OD pairs. The shortest path between each OD pair was found and the input constant $r = 1$ was used for adding the mid-arc node along them, which added 155 more candidate sites to the network. Nineteen different vehicle ranges (from 2 to 20) were used for comparing this model in the set-covering form (i.e., Model 1) with Wang and Lin's (2009) model. The model running time and the optimal number of fuel stations are given in Table 1. In this table "solution times" is the time taken by the solver to solve the model and "total time" is the sum of the time taken to generate the model, including preparing the sets and constraints and solving time. Solution time ratio is defined as the proportion of the solution time of Wang and Lin's (2009) model to Model 1 and total time ratio is the proportion of total solution time in Wang and Lin's (2009) model to Model 1. At the bottom of Table 1 in the total result row, these parameters are calculated according to the sum of all tests. Average row reports the average of solution time ratio and similarly the average of total times.

As shown in the "total time ratio" column of Table 1, observe that Model 1 covers the network at least 3.74 and at most 47.79 times faster with respect to Wang and Lin's (2009) model (in average 17.46 times). The fluctuation in solution time in Wang and Lin's (2009) model is much more while varying from 20 to 191 with the variance of 2,385.16, while Model 1 varies from 1 to 6 with the variance of 2.81. Figure 5 demonstrates the solution for a vehicle range of 12, which sets 16 facilities on 15 nodes and 1 facility on the midarc.

For this problem, Wang and Lin's (2009) model had 9,141 continuous and 4,750 integer variables. It also had 18,581 constraints, while Model 1 had 37,009 continuous and only 180 integer variables with 14,311 constraints. The enormous number of integer variables applied in Wang and Lin's (2009) model can justify why the solver took 104,634 iterations for solving while taking 5,467 iterations for solving Model 1 and, subsequently, took more solution time.

In the problem network, there were two shortest paths between nodes 14 and 19 with length 7 (i.e., 14–19 and 14–20–19). To solve this problem, the path 14–19 was considered to be covered but if the path 14–20–19 was regarded as the path that the driver selected to traverse, then the model would not locate facilities at node 19 for refueling. It indicated that the concentration on selecting the shortest path or adding new constraints to the model, to be able to select one path from among some paths for covering, can reduce the total cost of building a fuel station.

3.2. Randomly Generated Networks

To generate instant networks, first, n OD nodes were randomly generated in a square $[0, 100] \times [0, 100]$. Then, the Euclidian distance matrix was calculated between these nodes. Finally, each node was connected to the m nearest adjacent nodes. The length of each arc was the Euclidian distance from head to tail. Then, the resulted undirected network was used for finding the shortest path between each OD pair. The network instance was built on the network composed of all these shortest paths.

Six random networks were generated by this method and midarc nodes were added by the input constant $r = 4$.

Table 2 compares the solution time of both models on six randomly generated networks. The number of

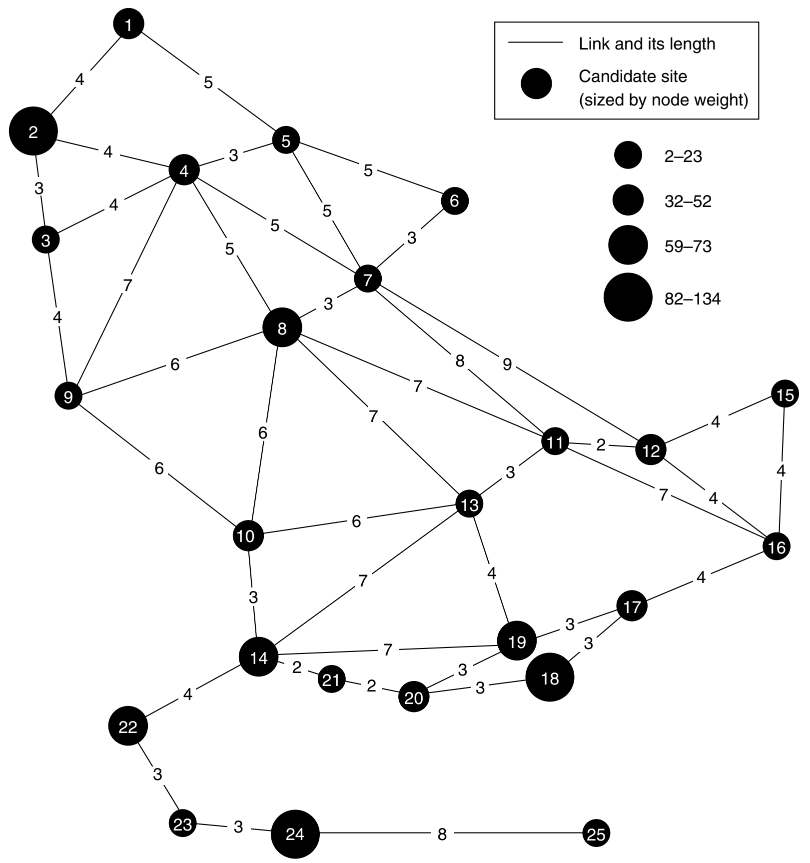

Figure 4 The 25-Node Network According to Lim and Kuby (2010)

OD nodes of the original network was set to 40, 60, and 80. Solution time, total time, solution time ratio, and total time ratio in Table 2 are defined similar to the ones in Table 1.

Results in Table 2 show that, in most cases, the longer the vehicle range, the fewer the facilities are required for covering all the paths. Also, the longer the vehicle range, the more solution time is required.

Table 1 Results of the 25-Node Network for the Driving Range of 2–20

		Solution time(s)			Total time(s)		
Driving range	No. of fuel stations	Model 1	Wang and Lin's (2009) formulation	Solution time ratio	Model 1	Wang and Lin's (2009) formulation	Total time ratio
2	87	0.51	38.83	76.14	3.06	39.35	12.86
3	59	0.61	20.08	32.92	3.06	20.67	6.75
4	42	1.10	53.24	48.40	3.85	53.68	13.94
5	36	1.84	36.43	19.80	4.57	36.93	8.08
6	28	0.98	70.89	72.34	3.68	71.42	19.41
7	24	0.64	70.42	110.03	3.31	70.84	21.40
8	20	1.18	191.12	161.97	4.01	191.64	47.79
9	19	0.90	132.23	146.92	4.10	132.93	32.42
10	17	1.73	170.67	98.65	4.65	171.17	36.81
11	17	0.95	128.31	135.06	3.90	128.82	33.03
12	16	1.34	93.18	69.54	4.22	93.71	22.21
13	16	1.48	103.41	69.87	4.54	103.94	22.89
14	15	3.78	66.38	17.56	6.77	66.97	9.89
15	15	5.85	74.01	12.65	8.92	74.52	8.35
16	14	2.54	26.52	10.44	5.40	27.20	5.04
17	14	2.09	80.65	38.59	5.09	81.13	15.94
18	14	4.45	47.80	10.74	7.38	48.44	6.56
19	14	5.24	30.36	5.79	8.24	30.81	3.74
20	14	4.20	34.62	8.24	7.61	35.04	4.60
Total result		41.41	1,469.15	35.48	96.36	1,479.21	15.35
Average				60.30			17.46

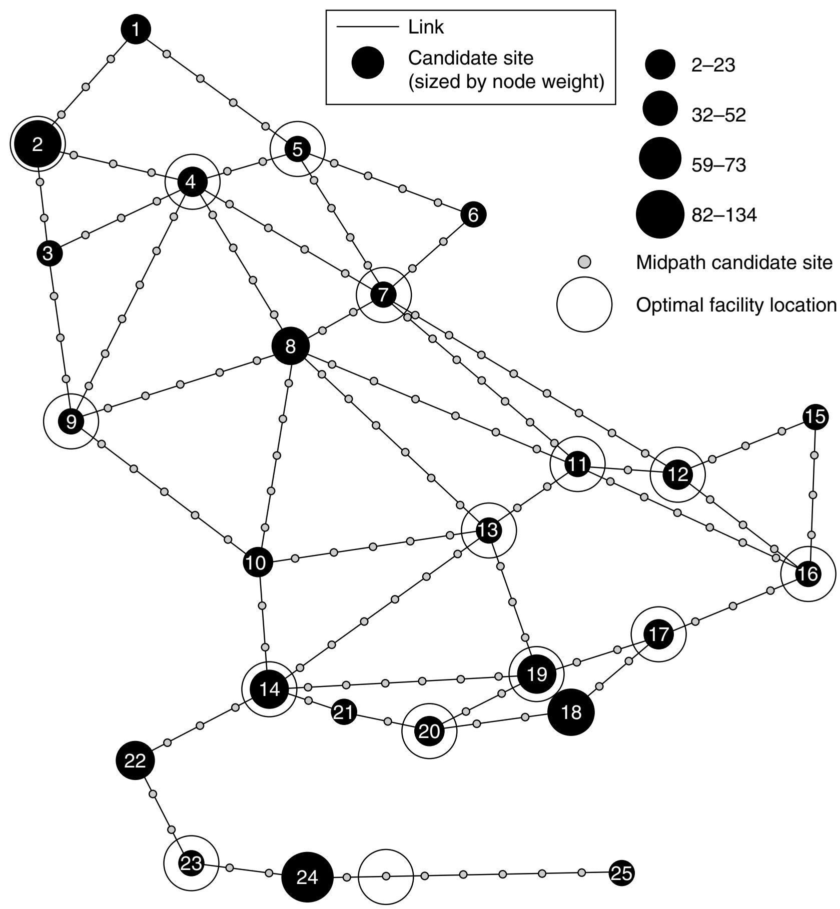

Figure 5 **Optimal Solution That Covers the Network with 16 Facilities for the Driving Range of 12**

Table 2 **Results of Six Randomly Generated Networks for the Driving Ranges of 6, 8, and 10**

Network no.	OD nodes	All nodes	Shortest paths	Driving range	Solution time(s): Model 1	Solution time(s): Wang and Lin's (2009) formulation	Solution time ratio	Total time(s): Model 1	Total time(s): Wang and Lin's (2009) formulation	Total time ratio	No. of fuel stations
1	40	159	780	6	0.94	23.76	25.28	6.06	25.95	4.28	70
				8	1.22	58.95	48.32	6.55	60.97	9.31	53
				10	6.66	210.52	31.61	12.46	212.33	17.04	41
2	40	172	780	6	0.87	23.60	27.13	6.03	25.74	4.27	89
				8	1.25	85.72	68.58	6.41	87.92	13.72	60
				10	2.64	183.58	69.54	8.09	185.64	22.95	47
3	60	181	1,770	6	2.03	53.95	26.58	16.72	64.09	3.83	86
				8	10.36	735.48	70.99	25.17	746.19	29.65	63
				10	59.90	1,517.00	25.33	75.18	1,524.66.	20.28	55
4	60	185	1,770	6	2.12	26.23	12.37	17.62	40.24	2.28	86
				8	5.10	735.03	144.12	20.43	758.62	37.13	70
				10	21.48	2,016.00	93.85	37.18	2,026.14	54.50	55
5	80	192	3,160	6	53.10	85.12	1.60	131.11	206.13	1.57	90
				8	81.57	2,429.00	29.78	157.04	2,511.74	15.99	74
				10	196.81	7,334.00	37.26	235.19	7,380.34	31.38	64
6	80	192	3,160	6	5.62	58.47	10.40	41.50	214.86	5.18	98
				8	159.73	5,331.04	33.38	197.19	5,378.37	27.28	78
				10	107.25	1,590.98	14.83	144.89	1,640.93	11.33	64
Total result					718.65	22,498.43	31.31	1,144.82	23,090.86	20.17	
Average							42.83			17.33	

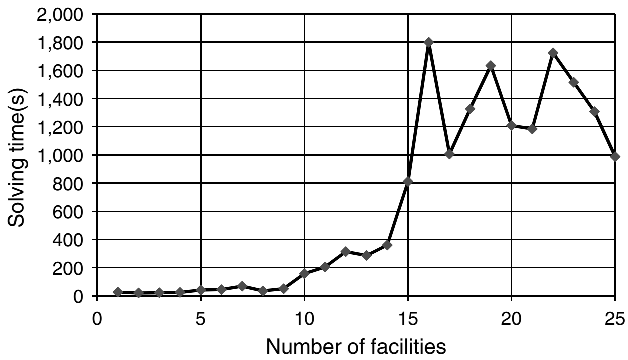

Figure 6 Solution Time for a 100-Node Network with the Driving Range of 10

Total solution time of Wang and Lin's (2009) model was 20.17 times more than that of Model 1 (95% solving time reduction with respect to Model 1).

3.3. The Maximum Coverage Model (FRLM)

To evaluate the model in flow-based maximum coverage form, a network with 100 nodes was randomly generated. These 100 OD nodes were connected by 4,950 shortest paths and 218 midpath nodes were added by the input constant $r = 2$. All these 318 nodes were served as candidate sites. Following Kuby and Lim (2007, 2005), Kuby, Lim, and Upchurch (2005), and Hodgson (1990), traffic flow volume on each shortest path f_q in the 100×100 OD matrix was calculated using a gravity model. The time taken to generate the model was 131.81 second average and Figure 6 demonstrates the results when the model was solved exactly with the driving range of 10 for the number of facilities from 1 to 25. Figure 7 shows the percentage of the refueled flows by locating 1–25 facilities.

The computational result of the maximum coverage form of Model 1 shows that it takes more time to locate more facilities but solution time is still reasonable. Also, there are some points at which the solution time decreases when the number of facilities increases. As was expected, the percentage of the refueled demand was a nondecreasing function of the number of facilities.

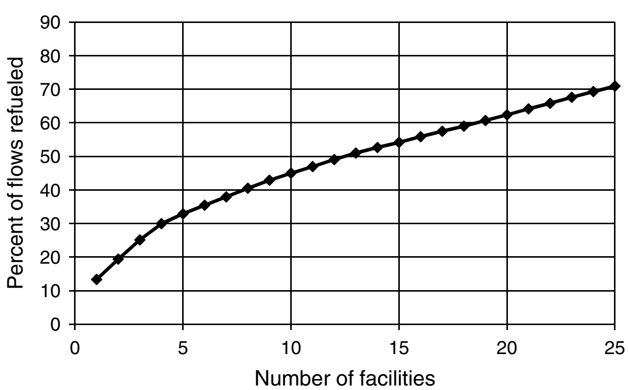

Figure 7 Trade-Off between the Number of Facilities and Percentage of Covered Flows (100 Nodes and Range of 10)

4. Conclusions

In this paper, the problem of optimal locations of refueling stations was studied and reviewed. In locating these facilities, the spatial demand, resource constraint, limited driving range, and relative location of stations to each other were the major issues. Because the combinations of facilities were required for covering the paths of networks and all previous MILP models took too much time to be solved with the commercial solver and the exact algorithms, such as the branch and bound, a new reformulation of the FRLM using an innovative expanded network was proposed here to effectively solve the large-scale set covering and the maximum coverage problems. The major advantage of the new model was the substantial reduction in the solution time of the set-covering problem and reasonable time for exactly solving the flow-based maximum coverage problem. Another advantage was its flexibility in switching between set-covering and maximum coverage concepts, which allowed the planner or decision maker to choose the optimal station deployment plan.

Tests on the 25-node benchmark and random generate problems showed that the new model had less solution time with respect to Wang and Lin's (2009) model. Other findings can be summarized as follows:

The fluctuation in solution time in the new model was much less than that in the Wang and Lin's (2009) model.

The longer the vehicle range, the more the time consumed by the solver for solving the model.

Selecting an appropriate path can reduce the total cost.

The longer the vehicle range, the fewer the facilities required for covering all the paths.

It took almost more time to locate more facilities with the maximum coverage form of the model.

5. Future Research

A considerable number of topics can be investigated in this area for future work, including Upchurch, Kuby, and Lim (2009) who handled capacitated facilities but considered fixed capacities, and Wang and Lin (2009) who considered incapacitated stations. Using this new formulation, more assumptions can be considered about capacity and the type of refueling stations proportional to the serving traffic flow.

The amount of time that drivers should wait in queues of stations indicates the level of service of fuel stations. It is necessary to study the rate of arrival flow and determine the relative location and an optimal number of dispenser in each station during peak hours.

Sustainability and robustness of fuel networks in major corridors are important factors that need to be rethought to support stations in breakdowns or failure of other stations. A great number of studies on path-based models for locating refueling stations have assumed the shortest path as a specific path that drivers choose to travel from the OD; but, in reality, drivers often select other paths that have deviations from the shortest path due to having facilities or passing through other cities. Some recent studies, such as Wang's (2011) model that can solve the location problem for any type of electric vehicle routing and Kim and Kuby's (2011) model that has the capability of considering multiple paths between an OD pair, have addressed this issue. It can be said that more studies with empirical data are still required to provide more realism for considering this driver behavior.

References

Bapna R, Thakur LS, Nair SK (2002) Infrastructure development for conversion to environmentally friendly fuel. *Euro. J. Oper. Res.* 142(3):480–496.

Bento N (2008) Building and interconnecting hydrogen networks: Insights from the electricity and gas experience in Europe. *Bento* 36(8):3019–3028.

Berman O, Krass D (1998) Flow intercepting spatial interaction model: A new approach to optimal location of competitive facilities. *Location Sci.* 6(1–4):41–65.

Berman O, Hodgson MJ, Krass D (1995) Flow-interception problems. Drezner Z, ed. *Facility Location: A Survey of Applications and Methods*, 1 ed., Vol. 1. (Springer Series in Operations Research, New York), 389–426.

Berman O, Larson RC, Fouska N (1992) Optimal location of discretionary service facilities. *Transportation Sci.* 26(3):201–211.

Capar I, Kuby M (2011) An efficient formulation of the flow refueling location model for alternative-fuel stations. *IIE Trans.* doi:10.1080/0740817X.2011.635175.

Church R, Revelle CS (1974) The maximal covering location problem. *Papers in Regional Sci.* 32(1):101–118.

Current JR, Revelle CS, Cohon JL (1985) The maximum covering/shortest path problem: A multiobjective network design and routing formulation. *Euro. J. Oper. Res.* 21(2):189–199.

Daskin MS (2008) What you should know about location modeling. *Naval Res. Logist.* 55(4):283–294.

Goodchild MF, Noronha V (1987) Location-Allocation and impulsive shopping the case of gasoline retailing. Ghosh A, Rushton G, eds. *Spatial Analysis and Location-Allocation Models* (Van Nostrand Reinhold, New York), 121–136.

Hakimi SL (1964) Optimum locations of switching centres and the absolute centres and medians of a graph. *Oper. Res.* 12(3): 450–459.

Hodgson MJ (1990) A flow capturing location-allocation model. *Geographical Anal.* 22(3):270–279.

Hodgson MJ (1998) Developments in flow-based location-allocation models. Griffith DA, Amrhein C, Huriót J-M, eds. *Economic Advances in Spatial Modelling and Methodology: Essays in Honour of Jean Paelinck*, 1 ed. (Kluwer Academic, Hingham, MA), 119–132.

Hodgson MJ, Rosing KE (1992) A network location-allocation model trading off flow capturing and *p*-median. *Ann. Oper. Res.* 40(1):247–260.

Huétink FJ, van der Vooren A, Alkemade F (2010) Initial infrastructure development strategies for the transition to sustainable mobility. *Tech. Forecasting Soc. Change* 77(8):1270–1281.

InterAcademy Council (2007) Lighting the way: Toward a sustainable energy future. Technical report, InterAcademy Council (IAC), Amsterdam.

International Energy Agency (2010) CO2 emissions from fuel combustion—Highlights. International Energy Agency (IEA), Paris.

Kim J-G, Kuby M (2011) The deviation-flow refueling location model for optimizing a network of refueling stations. *Internat. J. Hydrogen Energy* doi:10.1016/j.ijhydene.2011.08.108.

Kuby M, Lim S (2005) The flow-refueling location problem for alternative-fuel vehicles. *Socio-Econom. 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 Econom.* 7(2):129–152.

Kuby M, Lim S, Upchurch C (2005) Dispersion of nodes added to a network. *Geographical Anal.* 37(4):383–409.

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. *Internat. J. Hydrogen Energy* 34(15): 6045–6064.

Leiby P, Rubin J (2004) Understanding the transition to new fuels and vehicles: Lessons learned from experience of alternative fuel and hybrid vehicles. Sperling D, Cannon JS, eds. *The Hydrogen Energy Transition* (Elsevier Academic Press, St. Louis), 191–212.

Lim S, Kuby M (2010) Heuristic algorithms for siting alternative-fuel stations using the flow-refueling location model. *Euro. J. Oper. Res.* 204(1):51–61.

Lin Z, Ogden J, Fan Y, Chen C-W (2008) The fuel-travel-back approach to hydrogen station siting. *Internat. J. Hydrogen Energy* 33(12):3096–3101.

Lund H, Clark II WW (2008) Sustainable energy and transportation systems introduction and overview. *Utilities Policy* 16(2): 59–62.

Melaina MW (2003) Initiating hydrogen infrastructures: Preliminary analysis of a sufficient number of initial hydrogen stations in the U.S. *Internat. J. Hydrogen Energy* 28(7):743–755.

Melaina MW (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.

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

Melaina MJ, Ross MH (2000) The ultimate challenge: Developing an infrastructure for fuel cell vehicles. *Environment* 42(7):10–22.

Melendez M (2006) Transitioning to a hydrogen future: Learning from the alternative fuels experience. Technical Report NREL/TP-540-39423, National Renewable Energy Laboratory, Golden, Colorado.

Nicholas MA (2010) Driving demand: What can gasoline refueling patterns tell us about planning an alternative fuel network? *J. Transport Geography* 18(6):738–749.

Nicholas MA, Ogden J (2006) Detailed analysis of urban station siting for California hydrogen highway network. *Transportation Res. Record* 1983(1):121–128.

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

ReVelle CS, Swain RW (1970) Central facilities location. *Geographical Anal.* 2(1):30–42.

Rodrigue J-P, Comtois C, Slack B (2009) *The Geography of Transport Systems*, 2nd ed., Vol. 1. (Routledge, New York).

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

Shukla A, Pekny J, Venkatasubramanian V (2011) An optimization framework for cost effective design of refueling station infrastructure for alternative fuel vehicles. *Comput. Chemical Engrg.* doi:10.1016/j.compchemeng.2011.03.018.

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

Sperling D (1990) *New Transportation Fuels: A Strategic Approach to Technological Change*, reprint ed. (University of California Press, Berkeley, CA).

Sperling D, Kitamura R (1986) Refueling and new fuels: An exploratory analysis. *Transportation Res. Part A: General* 20(1):15–23.

Sperling D, Kurani K (1987) *Refueling and The Vehicle Purchase Decision: The Diesel Car Case* (Society of Automotive Engineers, Warrendale, PA).

Toregas C, Swain R, Revelle CS, Bergman L (1971) The location of emergency service facilities. *Oper. Res.* 19(6):1363–1373.

Upchurch C, Kuby M (2010) Comparing the *p*-median and flow-refueling models for locating alternative-fuel stations. *J. Transport Geography* 18(6):750–758.

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

Wang Y-W (2011) Locating flow-recharging stations at tourist destinations to serve recreational travelers. *Internat. J. Sustainable Transportation* 5(3):153–171.

Wang Y-W, Lin C-C (2009) Locating road-vehicle refueling stations. *Transportation Res. Part E: Logist. Transportation Rev.* 45(5): 821–829.

Wang Y-W, Wang C-R (2010) Locating passenger vehicle refueling stations. *Transportation Res. Part E: Logist. Transportation Rev.* 46(5):791–801.

Zeng W, Castillo I, Hodgson MJ (2008) A generalized model for locating facilities on a network with flow-based demand. *Networks and Spatial Economics* 10(4):579–611.