Title

Strategic Hydrogen Refueling Station Locations Analysis with Scheduling and Routing Considerations of Individual Vehicles

Authors

Kang, Jee Eun
Recker, Will W.

Publication Date

2012

UCI-ITS-WP-12-2

Strategic Hydrogen Refueling Station Locations with Scheduling and Routing Considerations of Individual Vehicles

UCI-ITS-WP-12-2

Jee Eun Kang
Will W. Recker

Department of Civil Engineering and
Institute of Transportation Studies
University of California, Irvine; Irvine, CA 92697-3600, U.S.A.
jekang@uci.edu, wwrecker@uci.edu

2012

Institute of Transportation Studies
University of California, Irvine
Irvine, CA 92697-3600, U.S.A.
http://www.its.uci.edu

Strategic Hydrogen Refueling Station Locations with Scheduling and Routing Considerations of Individual Vehicles

Jee Eun Kang and Will Recker
Department of Civil and Environmental Engineering
Institute of Transportation Studies
University of California, Irvine
Irvine, CA 92697 USA

Abstract

Set Covering problems find the optimal provision of service locations while guaranteeing an acceptable level of accessibility for every demand points in a given area. Other than reliance on static, exogenously-imposed accessibility measures, these problems either exclude substantive infrastructure-vehicle interactions or only include fragmented infrastructure-vehicle interactions related to the routing considerations of households seeking refueling service as a requirement of performing routine, daily activities. Here, we address this problem by coupling a Location-Routing Problem (LRP) that uses the set covering model as a location strategy to the Household Activity Pattern Problem (HAPP) as the mixed integer scheduling and routing model that optimizes households' participation in out-of-home activities. The problem addressed includes multiple decision makers: the public/private sector as the service provider, and the collection of individual households that make their own routing decisions to perform a given set of "out-of-home activities" together with a visit to one of the service locations. A solution method that does not necessarily require the full information of the coverage matrix is developed to reduce the number of HAPPs that needs to be solved. The performance of the algorithm, as well as comparison of the results to the set covering model, is presented. Although the application is focused on identifying the optimal locations of Hydrogen Fuel Cell Vehicle (HFCV) refueling stations, this proposed formulation can be used as a facility location strategy for any service activity that is generally toured with other activities.

Key Words

Location Routing Problem, Household Activity Pattern Problem, Set Covering Problem, Location Analysis, Infrastructure Investment, Hydrogen Refueling Stations, Routing and Scheduling Considerations, Column Generation

Introduction

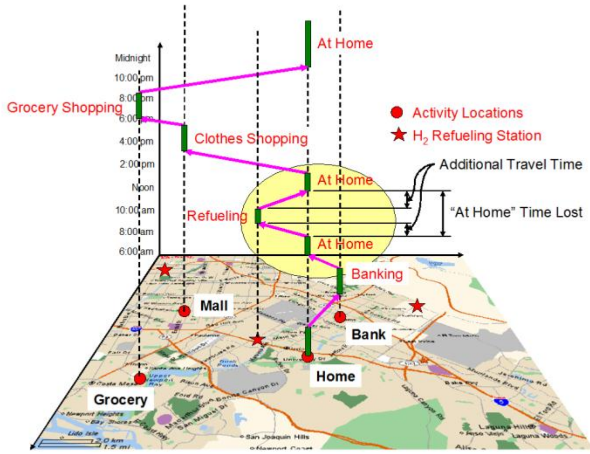
Hydrogen Fuel Cell Vehicles (HFCVs) operate on an electric engine powered by hydrogen fuel cell, producing zero emissions during vehicle operation. Their fuel energy efficiency is 40-60% while that of Internal Combustion Engine Vehicles (ICEVs) is around 20%. And, because energy sources used to produce the hydrogen vary, increased adoption of HFCVs may lead to lower fossil energy dependency, and may ultimately draw more from renewable energy sources. The main advantage of HFCVs over Battery Electric Vehicles (BEVs) or Plug-in Hybrid Electric Vehicles (PHEVs) is their similarity to conventional ICVs. Because the full driving range of HFCVs is competitive to ICEVs, the frequency of any need for refueling—the refueling time takes as low as 3 minutes which is a typical time to refuel the ICVs—will not require significant change to a driver's previous behavior.

Despite these advantages for success in the automobile market, HFCVs face a tremendous obstacle against widespread, early, adoption: an enormous investment is needed to provide the refueling infrastructure critical for HFCV operations. The so-called "chicken-and-egg" problem due to insufficient consumer demand needed to support the building of hydrogen refueling stations—according to California Fuel Cell Partnership (2009), one hydrogen refueling station is built at between $ 1.5 and 5.5 million—versus not having in place enough stations to enable the consumers to purchase the HFCVs remains largely unsolved.

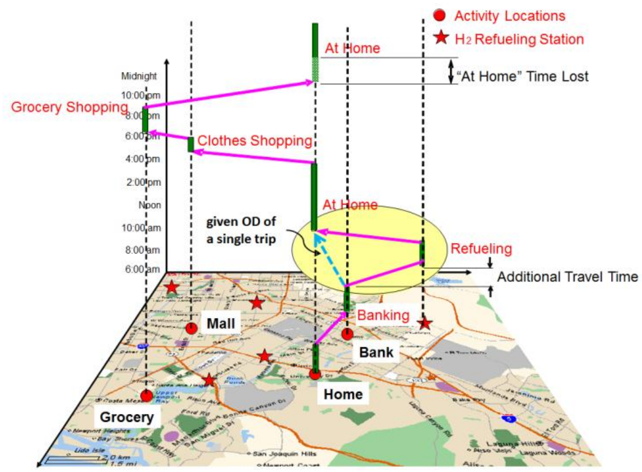
Identifying the minimum requirement for initial siting of hydrogen stations, and maximizing the effect of the public investment likely needed to "jump start" the hydrogen infrastructure that could lead to practicality of HFCVs, has received increasing attention, albeit mainly on a general scale. Melaina (2003) examined two stage initial conditions for the US road system: stage 1 to support early adopters' travel and stage 2 to support initial mass production; the minimum numbers of stations according to the criteria were estimated to be 4,500 and 17,700 stations. A number of studies have taken facility location siting approach for optimizing the initial investment of HFCVs refueling infrastructure. Nicholas and Ogden (2006) and Nicholas et al. (2004) applied the p-median (where p is the number of facilities to be located) problem to minimize the total travel times to the nearest refueling stations for a set of trips originating from each TAZ in the California metropolitan areas of Sacramento, San Francisco, Los Angeles and San Diego. From various scenarios, the density of stations to achieve certain levels of average driving time to nearest station and the population density are shown to have an inverse relationship. For example, the Los Angeles metropolitan area requires 6.8 % of current gasoline stations to provide hydrogen refueling service to achieve an average of 3 minutes driving time while the Sacramento metropolitan area requires 15.8%.. Stephens-Romero et al. (2010) and Stephens-Romero et al. (2011) applied a set covering problem for a series of early adoption communities in Southern California targeted by auto-manufacturers, and found the percentages of current gas stations "refitted for hydrogen refueling" that would guarantee the tolerable travel time to all nodes in the areas. Kuby and Lim (2005) developed a Flow-Refueling Location Problem from the Flow Capturing Location Model (Hodgson, 1990) assuming that refueling activities can be done between the origin and the destination within the range limit[1]. Based on this model, a 'clustering and bridging' strategy of hydrogen infrastructure investment plan was suggested (Kuby et al., 2009). A recent development allows deviation from the shortest path (Kim and Kuby, 2012) of a given OD. In addition to the four different models of refueling station locations cited here, there are a number of other models, many of them extending the Flow Capturing Location Model. MirHassani and Ebrazi (2012) categorized these existing models into three different groups: node-based, arc-based, and path-based. Readers are referred to their study for a detailed literature review of up-to-date hydrogen refueling station siting studies.

Here, we argue that accessibility to refueling activities is best measured within the context of the touring of trips in which drivers commonly insert such activities in planned tours with other routine activities as the need arises (typically with a frequency of once in several days, depending on driving behavior). From a travel behavior modeling perspective, we can visualize such accessibility measurements as in Figure 1, which illustrates vehicle-infrastructure interactions associated with different types of models, together with routing and scheduling interactions. Figure 1-(a) represents models using a single point, and Figure 1-(b) using a single trip (OD). Although the representation of demands in the latter case works well for long distance trips (Kuby et al., 2009; Kim and Kuby, 2012), about 70-80% of daily cumulative travel distance is less than 40 miles (Kang and Recker, 2009), which is not anywhere near the full range of HFCVs. In order to investigate daily travel comprised of a several tours of rather short distance trips, here we expand the temporal scope of the problem to a full day and address the "optimal" selection of the refueling activity (an individual decision based on the particular travel-activity pattern of that individual), together with the "optimal" siting of the refueling locations (a decision of the provider based on the collective decisions of the individuals in executing their travel-activity patterns) as in Figure 1-(c).

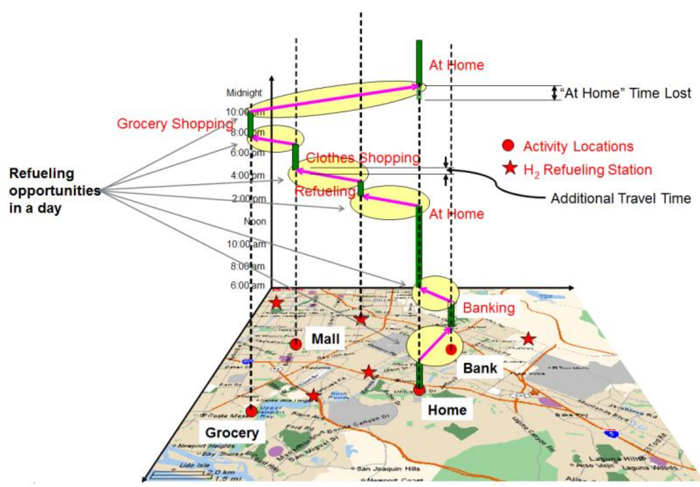
[1] Chevy Equinox FC (190 miles), Honda FCX Clarity (240 miles), Toyota FCHV-adv (Fuel Cell Hybrid, 431 miles)



(a) Refueling trip starting from home ending at home



(b) Refueling trip within a given trip



(c) Refueling trip opportunities throughout the day

Figure 1. Alternative HFCV Usage Patterns for a Single Day

Specifically, based on the concept that refueling trips are mostly linked to other primary activities, we propose a facility location problem with full-day household scheduling and routing considerations. This is in line with Location-Routing Problems, where decisions of facility locations are influenced by possible vehicle routings. The model we propose takes the Set Covering Model (with the "Set" comprising households with HFCVs) as a location strategy, and the Household Activity Pattern Problem as the scheduling and routing algorithm for performance of the households' activity schemes.

The proposed model solves the location problem simultaneously with multiple routing problems that include visitation to one of the available locations for refueling. Due to the computational complexity of the problem, a solution algorithm developed specific to the model is also described in this paper.

Location-Routing Problem and Location-HAPP Problem

The Location-Routing Problem (LRP) refers to locational problems that consider optimal locations of facilities from a vehicle routing perspective. It evolved from the management point of view that the location of distribution centers and the routing of vehicles to visit all customers from each distribution center are closely interrelated. Location Routing Problems take possible routing patterns around these depots, and the costs associated with them, into account at the time of locating distribution centers. Typically, an LRP formulation includes three parts: location, routing, and allocation. During the last few decades, the Location-Routing Problem has been studied widely, resulting in various problem formulations and numerous methodological advances (Min et al., 1998; Nagy and Salhi, 2007). The practical applications addressed by LRPs are not limited to the private sector. While most cases involve the shipping industry or decision making of private firms related to product/goods distribution or plant locations, several papers present applications, such as medicine (Or and Pierskalla, 1979; Chan et al, 2001), and waste/hazardous materials (List et al., 1991), of interest to the public sector.

To categorize by structure, the standard LRP minimizes the overall cost, comprised of depot cost, which follows median type problems, and vehicle routing cost, which includes tour planning by which a set of vehicles (one or more) traverses customer locations from/to the depots. Many previous works, both in the public sector as well as in the private sector, have this structure. Some of the non-standard structures include waste/hazardous materials applications (List et al., 1991) that often use multiple objective functions and replace tour problems with transportation problems, and the many-to-many LRP of the shipping industry that includes customers sending goods to others (Nagy and Salhi, 1998). Although an important aspect for many public sector infrastructure problems (and even for the private sector), there apparently has not been substantial effort with a focus on covering problems.

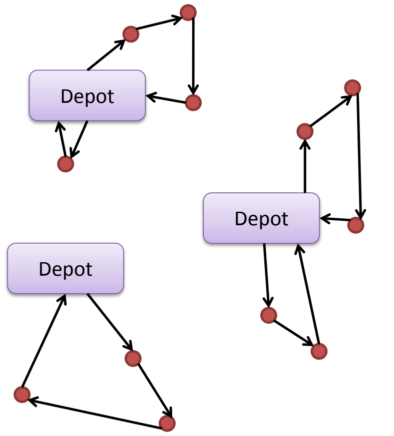
Here, we formulate a Location–Household Activity Pattern Problem (Location-HAPP) that can be identified as an LRP in a broad sense. The goal of the problem is to help identify the minimum hydrogen refueling service infrastructure—which may be either fully provided or subsidized by the public sector—that is necessary to support the initial phases of HFCV growth, i.e., before a sustainable private sector market is generated. The Location-HAPP takes the set covering problem as the basis for the location part in the LRP and couples it to an activity-based modeling framework to estimate the basic coverage of refueling stations needed to support early deployment of HFCVs. Rather than defining basic coverage by a set covering problem based on the range of HFCVs, which is about 190-430 miles, the term "basic coverage" here is defined in terms of "the maximum tolerable inconvenience" level to drivers for hydrogen refueling—a requirement likely to result in much denser packing than one that supports only maximum travel to a station. In addition, it is stressed that coverage in this context is not merely in terms of the direct distance (cost function) between home and the service location, but rather in terms of an additional distance within the tours of existing activities for any given day.

For the routing part of standard LRPs, the Traveling Salesman Problem or its variations are often used. Rather, for the hydrogen refueling stations case, a completely different routing structure needs to be used. By the property assumed that the refueling activity of customers (drivers) tends to tour with other activities when visiting the service location, the routing part of this problem needs to describe how each customer visits the service location within the confines of their daily schedules and within the constraints imposed by the touring of other multiple activities. We use the Household Activity Pattern Problem (Recker, 1995) as a tool that optimizes personal and household travel behaviors within the context of a Vehicle Routing Problem (VRP). This routing structure of HAPP is the reason we categorize the problem as an LRP rather than an extension from the path-based models to a tour-based model. A sequence of tours comprised of trips is actually an output of a decision-making scheme; specifically, travel decisions determining the sequence of activities (trips), departure times, and locations for certain activity types – refueling in the application considered in this paper. It may be argued that, in a conventional setting (e.g., ICEV gasoline refueling) a vehicle refueling activity either may or need not influence the sequence of trips and departure times; however, that it has significant impact on travel behavior when fuel availability is limited has been verified (Kitamura and Sperling, 1987).

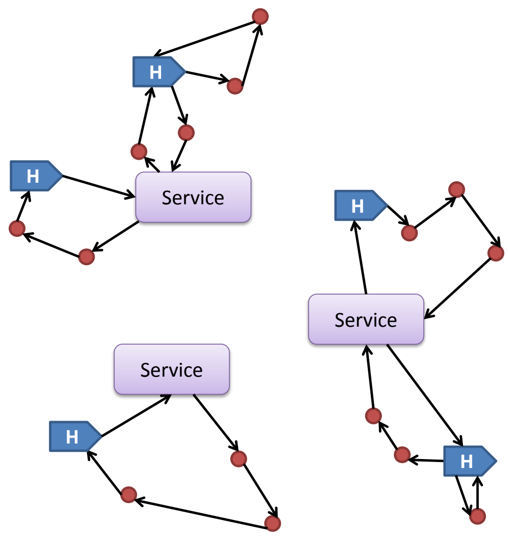
In contrast to LRPs, the Location-HAPP formulation features multiple decision makers in the problem: a public/private agency that makes facility location decisions, and a collection of individuals that make their own independent daily travel decisions but with consideration of the refueling locations common to all. Although this may well be formulated as a bi-level problem, we instead have treated it as a single problem by conveniently parameterizing the individual travel problem utilizing the structure of the set covering problem shown in the next section.[2] Additional notable differences are: 1) we are locating service locations that need to be visited by multiple customers, 2) the depots are customers' home locations in the routing problem, and 3) activity locations that are neither home (depot) nor service locations are visited as well. The inclusion of locations that are not directly connected to the facilities that are being located allows each household's location selection choice for the service type to be an output of

[2] As a side note, Kang et al. (2012) has integrated the network design problem with an activity-based travel demand via a bi-level structure in a more general manner.

interactions among other activities and schedules included in the analysis. These differences between the classical LRP models and what we propose in this paper are conceptually depicted in Figure 2.



(a) A single-stage Location-Routing Problem



(b) A single-stage Location-HAPP problem

Figure 2. Conceptual Diagram of Classical Location-Routing Problem vs Location-HAPP Problem

Problem Formulation

Consider a set of $|\mathbf{H}|$ households, $\mathbf{H} = \{h_1, h_2, ..., h, ..., |\mathbf{H}|\}$, in possession of one or more HFCVs, each of which on a particular day has an agenda, $\mathbf{A}_h = \{1, 2, ..., n_A^h\}$, comprised of n_A^h routine out-of-home activities with specific locations scheduled for completion, together with a need to refuel/service an HFCV at any one of a set of n_R candidate locations, $\mathbf{A}_R = \{1, 2, ..., n_R\}$—for example, in the case of hydrogen refueling stations, candidate locations $\mathbf{A}_R$ might logically be current gasoline stations. The problem considered finds the optimal locations of the hydrogen refueling stations both with respect to the supply side costs of the refueling stations (via the objective function of minimizing the cost) as well as with respect to the corresponding demand side optimal activity patterns of the households (via constraints that guarantee that a certain level of accessibility is ensured for everyone in the specified area). Although here we apply the Location-HAPP Problem to refueling station siting, we note that the models developed also extend to any service that requires coverage and that customers tend to travel to within a tour of other activities (e.g., post office).

Define $Z_j, j \in \mathbf{A}_R$ as the binary locational decision variable of service type R, and $C_j, j \in \mathbf{A}_R$ as the corresponding stationary cost associated with operating the service location. We assume that each household's , $h \in \mathbf{H}$, travel decisions are made so as to minimize the travel disutility subject to temporal and spatial constraints specified by an extension of the original HAPP model (Recker, 1995). In this paper, we use the formulation proposed by Kang and Recker (2012), which extended the original model to include the capability of selecting one location for one or more activity type(s) from many candidate locations. In the case of hydrogen refueling, we specify that one and only one of candidate locations for service type R needs to be visited. Then, the general form of minimizing the total disutility of all travelers $h \in \mathbf{H}$ can be represented as:

$$\min \sum_{h \in H} O^h = \sum_{h \in H} \text{Travel Disutility of Household } h = \sum_{h \in H} f(\mathbf{X}^h, \mathbf{T}^h)$$

s.t.

$$\mathbf{A}^h \begin{bmatrix} \mathbf{X}^h \\ \mathbf{T}^h \\ \mathbf{Y}^h \end{bmatrix} \leq= \mathbf{b}^h \;; \tag{1}$$

$$\mathbf{X}^h = \left[X_{u,w}^{v,h} \;, u, w \in \mathbf{N}_h, v \in \mathbf{V}_h \right],\ \mathbf{T}^h = \left[T_u^h, u \in \mathbf{P}_h^A \cup \mathbf{P}_h^R \right], \mathbf{Y}^h = \left[Y_u^h, u \in \mathbf{P}_h^A \cup \mathbf{P}_h^R \right]$$

where O^h is the travel disutility associated with the travel pattern adopted by household h, $\mathbf{N}_h$ is the set of all nodes visited by household h, (including those associated with refueling), $X_{uj}^{v,h}$ is a binary decision variable equal to unity if household h's vehicle υ travels from activity u to activity w, and zero otherwise, T_u^h is the time at which participation in activity u of household h begins, Y_u^h is the total accumulation of either sojourns[3] or time spent away from home on any tour, of household h on a particular tour immediately following completion of activity u, $\mathbf{V}_h$ is the set of vehicles available to the household (including one or more HFCVs), $\mathbf{P}_h^{A+}$ is the set of activities with predetermined locations, $\mathbf{P}_h^{R+}$ is the set of potential refueling activities (each with specific location common to all households), only one of which is to be completed, and $\mathbf{A}^h$ is a matrix of constants. (The details of Equation (1) are presented in the appendix.) We note that, in the case that the travel disutility is measured solely by the cumulated travel time, $f(\mathbf{X}^h, \mathbf{T}^h)$ is specified by the simple linear relationship

$$f(\mathbf{X}^h, \mathbf{T}^h) = \sum_{\substack{\forall u, w \in \mathbf{N}_h \\ v \in \mathbf{V}_h}} t_{u,w}^{v,h} \cdot X_{u,w}^{v,h} \tag{2}$$

[3] We have used the total accumulation of sojourns, and its maximum capacity of 4 (D = 4).

where t_{uw}^{hv} denotes the travel time from the location of activity u to the location of activity w using vehicle v.

The objective function, (1) is to minimize the sum of travel disutility of all households. However, because the objective functions and constraints are completely separable into each household problem since they do not share any of the variables, parameters or constraints across different households, this is equivalent to solving multiple household cases and combining the final objective values—because of this, the optimization can be put in terms of each household for a simple presentation of the problem that involves multiple decision makes.

The refueling service provider's objective is to minimize associated costs. Adding this objective and constraints, the Covering Location-HAPP formulation is as follows.

$$\min Z = \sum_{j \in \mathbf{A}_R} C_j \cdot Z_j \tag{3}$$

Subject to

$$X_{uj}^{v,h} \leq Z_j, u \in \mathbf{N}_h, j \in \mathbf{A}_R, v \in \mathbf{V}_h, h \in \mathbf{H} \tag{4}$$

$$\sum_{v \in \mathbf{V}_h} \sum_{u \in \mathbf{N}_h} \sum_{w \in \mathbf{N}_h} t_{uw}^{h} \cdot X_{uw}^{v,h} \leq O_{\min}^{h} + L \tag{5}$$

and conditions (A2) – (A20), contained in the Appendix[4].

Where,

$O_{\min}^{h}$: The travel disutility without visiting the service location of type R. This value can be obtained either by solving (A1) subject to (A2), (A4) – (A20), independently from the proposed problem, or the current value from existing survey data can be used.

L: The maximum tolerable inconvenience, or the minimum level of service for service type R.

The objective of the Location-HAPP to provide the infrastructure available to everyone in the region is to minimize infrastructure cost, $\sum_{j \in \mathbf{A}_R} C_j \cdot Z_j$. Equations (4) constrain that a visit to a service location can only be made at a location that provides refueling service. These are the linking constraints between the master problem of locating service centers and the sub-problem of household activity and routing decisions. Conditions (A2)-(A20), together with conditions (3) and (4) describe the scheduling

4 HAPP case 1 is used throughout this paper. HAPP case 1 is the simplest HAPP where each household member has an exclusive use of one vehicle.

and routing possibilities for each household's activities. In the original HAPP and typical location-routing problems, the objective minimizes routing cost as well, but for Location-HAPP, optimal routing is not necessarily of the model's interest. Rather, identifying whether obtaining the feasible region as constrained by condition (5) is possible or not is the key. Conditions (5) guarantee the minimum level of service to every vehicle/person or household in the analysis area. In (5), the inconvenience is specified in terms of travel times, t_{uw}^{hv}, but other travel disutility functions are equally substitutable.

(A2)-(A20) constrain the daily movement of a personal vehicle to perform a given set of out-of-home activities at the given locations, and one refueling trip at one of the refueling locations. The assumption is that the vehicle is replaced with an HFCV, but all individuals do not change their participation of the daily activities that they performed with an ICEV and a refueling trip. Because of the general unavailability of travel diaries over sufficient time periods to capture activities of the likely frequency of refueling (say, once in several days), we make certain assumptions regarding the refueling activity. In this analysis we insert one refueling trip per vehicle, under the assumption that the daily travel diaries reported in the survey represent a "typical" travel/activity day for the respondent. This, of course, in no way is meant to imply that such a refueling trip/activity occurs on every day, but rather that on the day that the refueling does occur the activity schedule of the individual is that reported in the survey. We note that, as an alternative synthetic activity synthetic pattern generation may be utilized as in Xi et al. (2012), however this approach is not taken in this paper because of the unreliability of such generations in forecasting behavior (e.g., refueling at sparsely populated locations) for which there is no empirical base. The focus here is on the determination of inconvenience of limited refueling opportunities within empirical daily routines. However, we note that information on fuel inventory can be included in the base formulation described by above without changing its basic structure simply by adding the following set of constraints:

$$X_{uw}^{v,h} = 1 \Rightarrow F_w^h = I^{v,h}, \quad u \in \mathbf{P}_h, w \in \mathbf{P}_h^{R+}, v \in \mathbf{V}_h, h \in \mathrm{H} \tag{A21}$$

$$X_{uw}^{v,h} = 1 \Rightarrow F_w^h = F_u^h - f_{u,w}^{v,h}, \quad u \in \mathbf{N}_h, \mathrm{w} \in \mathbf{P}_h^A \cup \mathbf{P}_h^{R-}, v \in \mathbf{V}_h, h \in \mathbf{H} \tag{A22}$$

$$0 \le F_u^h \le I^{v,h}, \quad u \in \mathbf{N}_h, h \in \mathbf{H}$$

(A23)

$$X_{uw}^{v,h} = 1 \Rightarrow 0 \le F_u^h - f_{u,w}^{v,h}, \quad u \in \mathbf{N}_h, w \in \mathbf{P}_h^{R+}, v \in \mathbf{V}_h, h \in \mathbf{H} \tag{A24}$$

where the variable $F_w^{v,h}$ is defined as:

$F_w^{v,h}$, $\quad u \in \mathbf{N}_h, v \in \mathbf{V}_h, h \in \mathbf{H}$ the fuel inventory of vehicle v of household h immediately following the completion of activity u.

and related parameters of

$I^{v,h}$, $\quad v \in \mathbf{V}_h, h \in \mathrm{H}$ the full fuel inventory of vehicle v of household h.

$f_{u,w}^{v,h}$, $\quad u \in \mathbf{N}_h, \mathrm{w} \in \mathbf{N}_h, v \in \mathbf{V}_h, h \in \mathbf{H}$ fuel usage from the location of activity u to the location of activity w using vehicle v of household h.

in the above, (A21) represents the refueling activity. We assume that a vehicle is refueled to the maximum. (A22) represents the fuel consumption, and updates the inventory value. (A23) constrains that the vehicle can only be operated when the fuel inventory is not empty. (A24) constrains that the fuel inventory immediately before the departure to a refueling station to be greater than the fuel consumption to the location; since (A21) directly assumes the full inventory at the refueling station, (A24) is additionally needed. If the fuel inventory constraints are employed, we can relax the constraint of one visitation to a refueling trip. A vehicle can choose not to refuel or to refuel one or more times (for long distance trips).

$$\sum_{v \in \mathbf{V}_h} \sum_{w \in \mathbf{N}_h} X_{uw}^{v,h} \geq 0, \quad u \in \mathbf{P}_h^{A+}, h \in \mathbf{H} \tag{A2-1}$$

While the formulation, (3) – (5), (A2) – (A20), represents the development from the travel behavior modeling perspective, the structure of the set covering model can be utilized for the actual computation. By introducing parameters, $\mathbf{S}_j$, the Location-HAPP can also be written as the standard covering problem formulation as follows:

$$\min Z = \sum_{j \in \mathbf{A}_R} C_j \cdot Z_j \tag{3}$$

s.t

$$\sum_{h \in \mathbf{S}_j} Z_h \geq 1, \quad j \in \mathbf{A}_R, h \in \mathbf{H} \tag{6}$$

Where $\mathbf{S}_j$ is the set of facility candidate sites j within the acceptable accessibility for household h,

$$\mathbf{S}_j = \{ j \mid O^h = \sum_{v \in \mathbf{V}_h} \sum_{u \in \mathbf{N}_h} \sum_{w \in \mathbf{N}_h} t_{uw}^h \cdot X_{uw}^{v,h} \leq O_{\min}^h + L \}, \quad j \in \mathbf{A}_R$$

Or, (6) can be substituted with

$$\sum_{j \in A_R} a_{hj} \cdot Z_h \geq 1, \quad h \in H \tag{6a}$$

Where

$$a_{hj} = \begin{cases} 1 & \text{if household } h \text{ is within the tolerable service level from refuelting location } j \\ 0 & \text{otherwise} \end{cases},$$

$$j \in \mathbf{A}_R, h \in \mathbf{H}$$

This alternative formulation specified by (3), (6a) separates each household's travel decision from each other as well as the master facility location problem. It is noted that the first formulation, (3) – (5), (A2) – (A20), and the second formulation, (3), (6a), produce the same results.

Computationally, the Location-HAPP Problem differs from the Location-Routing Problems in the following points. First, in the sub-problem of routing, each activity point (except refueling locations) is already assigned to households while in the classical LRPs' sub-problem each customer location can be visited by any vehicle from any depot. Second, the facility node is not the depot, but an intermediate location that each vehicle needs to traverse once and only once. These properties make the computational complexity of Location-HAPP much less intensive since the numbers of nodes each vehicle visits are rather limited, and the decision of choosing one of the open service locations is much easier than the decision of allocation all nodes to depots and vehicles.

Solution Algorithm

With (6a), each vehicle's routing problem has been completely isolated from those of other vehicles, as well as from the master problem. However, it requires $|\mathbf{H}| \times n_R$ number of HAPP problems in which each visits one of specified candidate locations that consistent with the parameter matrix, a_{hj}. This parameter matrix can be calculated from solving (A1)-(A20) with a condition that the household visits refueling location j. Finding this matrix becomes an issue in the Location-HAPP Problem—calculating $|\mathbf{H}| \times n_R$ sub problems that are NP-hard[5] makes the computational set-up for the master problem important.

The Set Covering and Set Partitioning problems are well-known binary combinatorial problems, and there exist many algorithms to handle the associated computational complexity (Caprara et al., 2000;

[5] However, there are many methodologies to handle HAPP or its original form of PDPTW.

Christofides and Korman, 1975). For the problem addressed in this paper, the key is to have the least number of actual a_{hj} parameters to be added in the master problem of the Location – HAPP Problem. Column Generation is a technique that does not require all variables and parameters to be in the master problem, which greatly reduces the number of a_{hj} parameters actually needed to solve the problem. Instead, column generation solves a sub problem of finding a variable that would reduce the objective function value of the master problem. The process is similar to simplex in that, at each iteration, the entering variable is selected based on reduced cost. Once an entering variable is selected, it needs to be added to the previous master problem with its parameter column. Then, the new master problem is solved, and this process is repeated until no entering variable with a negative reduced cost can be found. For efficiency, the sub problem should not be computationally challenging. Following (Lübbecke, 2011), the formulations of master – sub problems are:

$$
\begin{aligned}
&MP \\
&v(MP) := \min \sum_{j \in \mathbf{J}} c_j \lambda_j \\
&s.t \\
&\sum_{j \in \mathbf{J}} a_j \lambda_j \geq b \\
&\lambda_j \geq 0,\ j \in \mathbf{J}
\end{aligned}
\qquad\qquad
\begin{aligned}
&SP \\
&v(SP) := \min\{c_j - \pi \cdot a_j \mid j \in \mathbf{J}\} \\
&c_j - \pi \cdot a_j \leq 0
\end{aligned}
$$

Column Generation: The Column Generation Procedure

where

λ_j : the binary decision variable

c_j : cost for decision λ_j

a_j : $|\mathbf{H}| \times 1$ j th column vector of parameters

b : $n \times 1$ vector

π : $|\mathbf{H}| \times 1$ vector of dual variables, each component associated with a constraint row

This procedure is found to be efficient for some types of large combinatorial integer problems (Barnhart et al., 1998), including the Set Covering problem. Moreover, the fact that initialization is not difficult and that linear relaxation of the master problem is stable makes column generation suitable for application to the problem considered here.

For standard column generation, the sub problem is another optimization problem that finds the most negative reduced cost. However, this sub problem requires the full information of the coverage matrix, which we wish to avoid. Instead, we deploy a new search procedure that finds an entering column with a negative reduced cost, not necessarily a minimum. This way, we can keep the structure of column generation without having the full knowledge of the coverage parameters.

The overall iterative procedure we propose for the Location-HAPP is shown in Figure 3.After initialization, the master problem is solved. It then passes the dual values to the sub-problem that finds the next entering variable, j, with a negative reduced cost by our search algorithm. If no dual variable exists with negative cost, or any stations that are not in the master problem can deliver a negative dual cost, we conclude that it is at optimum. Otherwise, the entering variable and its coverage column are added to the master problem. Only the third box is changed from the standard procedure of Column Generation.

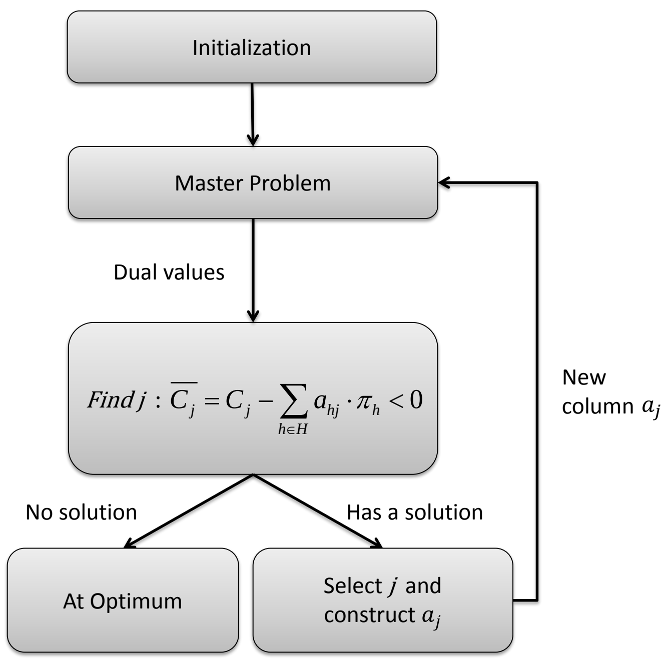


Figure 3. Schematic of iterative procedure

For the initial setup of the problem, it is first assumed that each household is covered by the refueling station closest to the home location (home is always one of the nodes in the household's pattern). If that station is not within the tolerance level, the next closest station can be tested. Although it is likely that the household is covered by a station closer to home location, this proximity rule may further be developed to be more sophisticated, and refined to be appropriate for routing considerations; for example, by the order of the distance to any of its activity nodes. Once it is decided that station j covers household h, the remaining a_{hj} parameters for all households are assumed to be 0 for station j, regardless of whether or not station j covers household h within the tolerance level. This is purely for the computational convenience that there is no practical advantage to constructing and verifying all parameters, and that the number of HAPPs to solve should be kept to the lowest possible. This way, we reduce initialization cost of the problem. For example, assume a case of 5 households and 3 refueling candidate locations, with full coverage matrix shown in (M1):

$$\begin{array}{c} \\ HH1 \\ HH2 \\ HH3 \\ HH4 \\ HH5 \end{array} \begin{array}{c} \begin{array}{ccc} c1 & c2 & c3 \end{array} \\ \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix} \end{array} \tag{M1}$$

If it is the case that the closest refueling candidate stations from households 1, 2, 3 is station candidate 1, and the closest refueling candidate stations from households 4, 5 is station candidate 2—all within the tolerance level—then the initial parameter matrix is:

$$\begin{array}{c} \\ HH1 \\ HH2 \\ HH3 \\ HH4 \\ HH5 \end{array} \begin{array}{c} \begin{array}{cc} c1 & c2 \end{array} \\ \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \end{array} \tag{M2}$$

Here, we do not check for the coverage parameters, $a_{HH4,c1}, a_{HH5,c1}, a_{HH1,c2}, a_{HH2,c2}, a_{HH3,c2}$, and eliminate the need of solving those 5 HAPPs.

However, we also note that degeneracy occurs for cases in which all actual solutions are a subset of the initial, synthetic, solution and the real coverage of those stations is highly superior than those of others not in the initial master problem. In such cases, at any iteration, stations already in the master problem cannot be the entering variable, and therefore, the initial synthetic coverage columns cannot be replaced with real ones, which are superior to the synthetic ones. For example, assume that we initialized as (M2), and the original full coverage matrix is:

$$\begin{array}{c} \\ HH1 \\ HH2 \\ HH3 \\ HH4 \\ HH5 \end{array} \begin{array}{c} \begin{array}{ccc} c1 & c2 & c3 \end{array} \\ \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix} \end{array} \tag{M3}$$

The optimal solution is building one station at site C1, which covers all households. Because our initial coverage matrix is (M2), C3 is the only column to be considered to be brought into the problem. However, the master problem would decide that C1, C2 are the solution, which will not allow a chance for real columns of those stations to be entered. This issue can be avoided simply by constructing the actual coverage matrix at initialization, but this would require calculating $|\mathbf{H}| \times (\text{number of initial columns})$ different HAPP models to be solved.

Following the standard Column Generation procedure, the sub-problem finds one station's coverage column with the most negative reduced cost to be added to the master problem, and then the master problem is solved to find the best combinations of stations to be built. As mentioned earlier, finding a station with the most negative reduced cost, full knowledge of the coverage matrix, $\mathbf{A}$ is needed. On the other hand, the sub-problem does not necessarily have to be an optimization problem since it is going to lower the objective function value in the master problem as long as the entering variable has a negative reduced cost. This may increase the number of iterations since we are not looking for the best variable, but it reduces the search cost of the column which has been the computational difficulty. Therefore, the sub-problem becomes:

$$Find\ j \quad \text{s.t. } \overline{C_j} = C_j - \sum_{h \in \mathbf{H}} a_{hj} \cdot \pi_h < 0 \tag{a}$$

Where,

$\overline{C_j}$: the marginal cost of locating a station at candidate node j

π_h: the dual variables associated with the coverage constraint of household h, $h \in \mathbf{H}$

The search method employed here to find station j with a reduced cost is as follows. First, there is no need to search for coverage parameters for constraints with zero dual values, $\pi_h = 0$; only those of positive dual values are generated by solving HAPP. Therefore, we traverse households with positive dual values, and update the cumulative term, $\sum_h a_{hj} \cdot \pi_h$ as the search progresses. And, as soon as

$\sum_h a_{hj} \cdot \pi_h > C_j$ has been established, there is no need to keep searching since $a_{hj} \cdot \pi_h$ is non-negative for any household constraint, so it can be concluded that j would be included in the master problem. Once it has been decided that j is to be the entering variable, all column parameters of a_j need to be generated. The most efficient order of searching is to check the parameters from the highest π_h to the lowest, for instance, using a priority queue data structure. In the example provided in the next section, a uniform unit cost of 1 is used; so, all dual variable values are either 1 or 0. If $\sum_h a_{hj} \cdot \pi_h \leq C_j$, even after all household coverage constraints have positive dual values, that variable is discarded and a new station is tested.

When choosing a station j to test, one random household with positive dual values is selected. Then, a priority queue that stores all candidate locations in an increasing order of the direct distance to that household is created. (We note, however, that this procedure can be developed to be more sophisticated; for example, by the order of the distance to any of activity nodes for all households with positive dual values.)

With the following notations defined for each iteration, i, this search method can be summarized as the following:

$\mathbf{Z}^i$: the set of stations that are in the solution from the master problem of previous iteration. This is the subset of $\mathbf{A}_R$

$\mathbf{M}^i$: the set of households with coverage constraints of dual values from previous iteration that are greater than 0.

$\mathbf{I}_j^i$: the subset of $\mathbf{M}^i$ of which the coverage parameter a_{hj} has been checked for the possibility of station j being the next variable. If $I^i = \mathbf{M}^i$, all components of $\mathbf{M}^i$ have been checked.

If $\mathbf{M}^i = \emptyset$, it is at optimum.

Else,

For all $j \in \overline{\mathbf{Z}^i}$ (loop 1)

For all $h \in \mathbf{M}^i$, (loop 2)

If, $a_{hj} = 1$, update $\sum_{h \in \mathbf{I}_j^i} a_{hj} \cdot \pi_h \leftarrow \sum_{h \in \mathbf{I}_j^i} a_{hj} \cdot \pi_h + a_{hj} \cdot \pi_h$, $\mathbf{I}_j^{\,i} \leftarrow \mathbf{I}_j^{\,i} \cup h$

If $\sum_{h \in \mathbf{I}_j^i} a_{hj} \cdot \pi_h > C_j$, select j as the entering variable and break both (loop 1) and (loop 2)

Else if, $\mathbf{M}^i \neq \mathbf{I}_j^{\,i}$, select a different $h \in \mathbf{M}^i$, $h \notin \mathbf{I}_j^{\,i}$ and continue (loop 2)

Else, Select a different $j \in \overline{\mathbf{Z}^i}$ and initiate a new (loop 2)

If $\sum_{h \in \mathbf{I}_j^i = \mathbf{M}^i} a_{hj} \cdot \pi_h \leq C_j$ for all $j \in \overline{\mathbf{Z}^i}$, it is at optimum.

Search Algorithm: The sub-problem of finding an entering variable with a negative reduced cost

Case Study: Area, Data and Results

Southern California is anticipated to be one of the early adoption areas of HFCVs. It is the location of major auto-manufacturers' headquarters, and three target areas (Torrance, Santa Monica, Irvine and Newport Beach) in Southern California have been identified as early adoption communities (CaFCP, 2009; CaFCP, 2010). As of 2011, sixteen refueling stations are under operation and 40 more stations are planned to be built, and the number of HFCVs deployed in this area is expected to be in the thousands by 2013, and tens of thousands by 2016 (CaFCP; 2009, 2010).

In the case study presented here, we focus on Irvine and Newport Beach as the study area. In the example, we presume that the candidate sites for future hydrogen refueling stations are drawn only from existing gasoline stations—there currently are 34 gasoline stations in the area. (Existing or planned hydrogen stations are not considered in this analysis[6]) Further, when there are two or more stations at an intersection, they are considered as one for simplicity. These existing gasoline stations are the only candidate sites.

[6] Two current hydrogen refueling stations in the area are: National Fuel Cell Research Center at University of California, Irvine (2012), and Shell (Los Angeles Times, 2012) which is annexed to a gasoline station

The analysis is based on a subset of household samples that are in the study region drawn from the Travel Diary, California Statewide Household Travel Survey (CalTrans, 2001). The Travel Survey contains the daily travel activities, and their full location data. Each trip contains information on departure and arrival times, trip/activity durations, and geo-coded information on longitude/latitude of the activity locations. A suitable subset of data was selected from the California Statewide Household Travel Survey that has complete location information and HFCV substitutable vehicle types (i.e., excludes such vehicles as motorcycles, bicycles, and etc.). Based on these data, person-based trip chains and activity ordering are converted to equivalent vehicle-based chains in order to simulate vehicle routing patterns.

The total number of full-day vehicle patterns (households) applicable for the analysis is 134; the total number of sample households in the California Statewide Household Travel Survey residing in Orange County is 500 (CalTrans, 2001). We assume that this sample is representative of travel behavior in the county. The study area and home/activity locations of the sample households are displayed in Figure 4.

For each vehicle/household, a travel time matrix is generated associated with all possible combinations of the $X_{u,w}^{v,h}$ for the out-of-home activities they performed. In the formulation, for every household, the first 34 nodes, 1 – 34, are specified as candidate locations for refueling—therefore universal for the entire sample—while nodes labeled 35 and greater are specified as the nodes unique to each individual household's out-of-home activities that were reported as performed. All-to-all travel time matrices are generated by calling *MapQuest API*.

Our principal behavioral assumption is that, with replacement of households' current conventional vehicles by HFCVs, each new vehicle will continue to perform activities that were reported in the survey, plus an additional refueling activity at one of the candidate locations. This vehicle usage model is equivalent to Case 1 in Recker (1995). For the construction of time windows, the method in Recker and Parimi (1999) is adopted. Work, school, meals between two work activities, and pick-up/drop-off activity types are constrained to have exact times as reported. Other activity types' open/close windows are specified as the minimum/maximum of respondent's reported activity start/end time and sample mean activity start/end time for the activity. For each individual, the vehicle/household start time window is the minimum of respondent's reported travel start time for his/her initial activity and mean reported travel start time for initial activity for the sample. In general, individual vehicle/household latest return to home is taken as the max of respondents' reported return-to-home time for his/her final activity and mean reported travel return-to-home time for final activity for the sample. However, when the sequence of activities performed by a household comprises only activities with exact start/end times, or in cases where the travel time generated from *MapQuest* is larger than the reported travel time, the last return home is relaxed since even the basic HAPP of reported activities becomes infeasible, since there is an additional refueling trip.

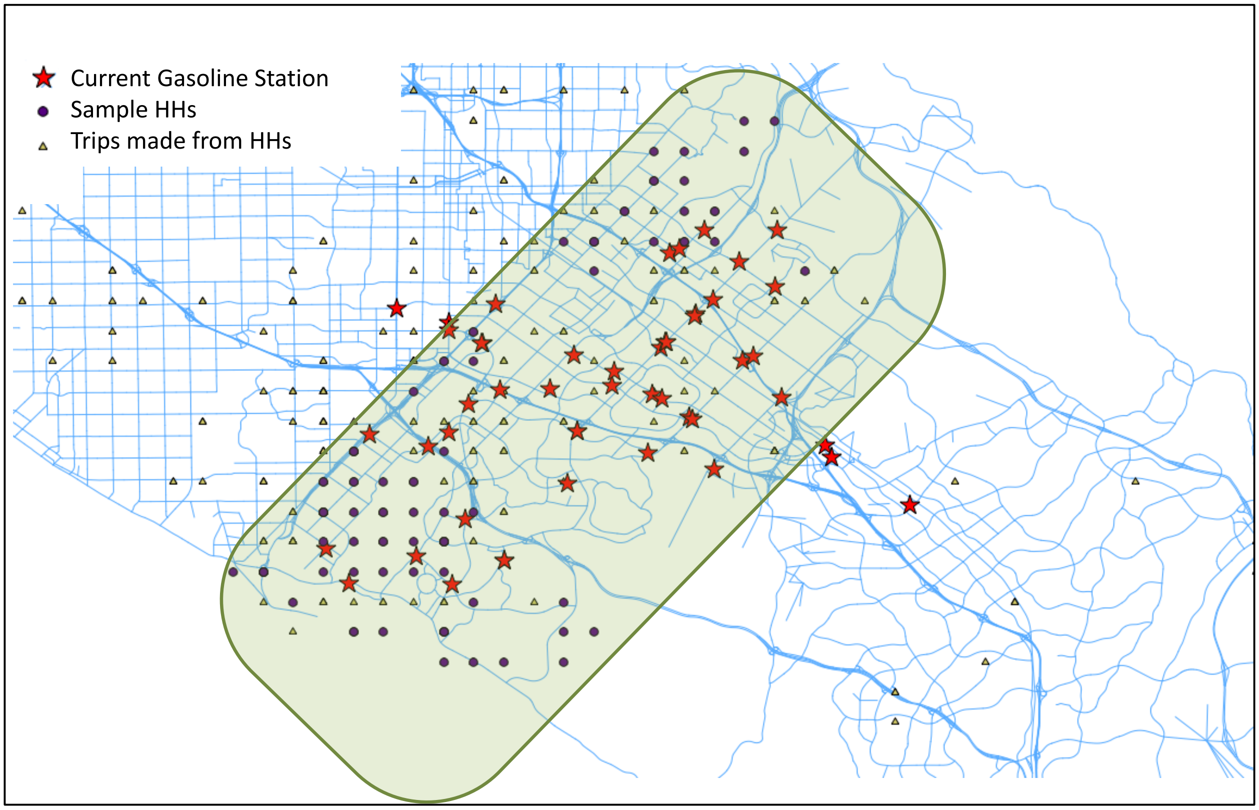


Figure 4. Study Area: Irvine and Newport Beach in California with Sample Households

The objective function for each household is assumed to be minimizing the total travel time throughout the day; i.e.,

$$\min O_h = \sum_{u \in N_h} \sum_{w \in N_h} \sum_{v \in V_h} t_{uw}^{h} \cdot X_{uw}^{v,h} \;\; ; \;\; \forall h \in H \tag{7}$$

When checking for a_{hj}, the comparison measurement of $O_{\min}^{h}$ is generated from HAPP without a visit to one of the refueling stations but with visits to all activities given by the survey responses for that household. When solving a HAPP, the exact dynamic program developed by Desrosiers et al. (1986) is used; however, direct calculation using an optimization tool, or any other suitable algorithm (e.g., branch and bound) would work as well (Cordeau and Laporte, 2003). Households in the data set not covered by any of the refueling candidate locations within the given tolerance, or households that would require violation of given constraints to visit any of the candidate locations, are omitted.

The detailed results for the base case involving 122 households with a refueling activity subject to the additional travel time tolerance of $L = 0.2hr$ are shown in Table 1. Optimum is reached after 7 iterations, solving 966 / 4,148 HAPP cases, to find the a_{hj} parameters needed to perform the proposed search algorithm. It is noted that the specific stations that each household visits are not kept track of, since any station j that has corresponding $a_{hj} = 1$ is acceptable. If needed, such data can be stored separately, and the station that gives the minimum objective value (including interactions with scheduling and routing of other activities) is presumed to be visited by the household.

Table 1. Results of Location – HAPP Problem ($L = 0.2hr$)

	Initialization $i=0$	$i=1$	$i=2$	$i=3$	$i=4$	$i=5$	$i=6$	$i=7$
Columns in MP	10	11	9	8	5	5	4	NA
Optimal Solution	$\sum_{j \in A_R} Z_j = 10$	$\sum_{j \in A_R} Z_j = 8$	$\sum_{j \in A_R} Z_j = 7$	$\sum_{j \in A_R} Z_j = 4$	$\sum_{j \in A_R} Z_j = 4$	$\sum_{j \in A_R} Z_j = 3$	$\sum_{j \in A_R} Z_j = 3$	NA
# of HAPPs solved [7]	122	2	2	2	6	4	3	93
# constraints with negative dual	NA	9	9	7	6	3	2	2

Four different distance tolerance level cases are tested and compared in Table 2. Cases 2, 3 and 4 are run using the proposed initialization method. For case 1, there is a station that is highly superior and covers all households, which is also one of the initial stations. Therefore, degeneracy occurred and it is necessary to construct the full initial columns. This increases the number of HAPP models that needed to be solved to 35% of total HAPPs. The best case we tested was case 4, for which only 7 % of total number of possible HAPP computations were required.

These cases are also compared to the Set Covering Problem— the "Set" refers to "home" locations—that guarantees the same levels of access based on the households' home locations;. The

[7] This is counting only HAPPs solved to find the next entering variable, or to find that it is at optimum. Once an entering variable is found, 122 additional HAPPs are solved to construct the parameter vector, a_j of the entering parameter column at each iteration.

coverage matrices are significantly sparser, leading to a larger number of stations for a particular level of access, when compared to the Location-HAPP formulation. From the results, it is argued that the Set Covering Problem may significantly over-estimate the number of stations required. Considering the cost of building and operating a station, it can be concluded that by including routing and scheduling consideration as a part of “accessibility”, the overall cost of the infrastructure supply can be lowered significantly as seen from the results of the Location-HAPP model.

Table 2. Results of Location-HAPP Problem vs. Set Covering Problem

	Tolerance level	**Case 1** ($L = 0.15hr$)	**Case 2** ($L = 0.2hr$)	**Case 3** ($L = 0.3hr$)	**Case 4** ($L = 0.4hr$)
	# households in Analysis	116	122	124	124
LRP – HAPP	**Optimal Solution**	$\sum_{j \in A_R} Z_j = 5$	$\sum_{j \in A_R} Z_j = 3$	$\sum_{j \in A_R} Z_j = 2$	$\sum_{j \in A_R} Z_j = 1$
	% of coverage of the full parameter matrix[8], A	1,498 / 3,944 (37.98 %)	2,048/4,148 (49.37 %)	2,888/4,216 (68.50 %)	3,108/4,216 (74.93 %)
	# HAPPs solved[9]	1,385[10] / 3,944 (35.11 %)	966 / 4,148 (22.29 %)	1,077 / 4,148 (25.96 %)	285 / 4,148 (6.87 %)
Set Covering Problem (SCP)	**Optimal Solution**	Infeasible	Infeasible	$\sum_{j \in A_R} Z_j = 5$	$\sum_{j \in A_R} Z_j = 2$
	% of coverage of the full parameter matrix[11], A	96 / 3,944 (2.43 %)	264 / 4,148 (6.36 %)	856 / 4,216 (20.30 %)	1768 / 4,216 (41.94 %)

[8] This is calculated by solving all HAPPs for parameters a_{hj}

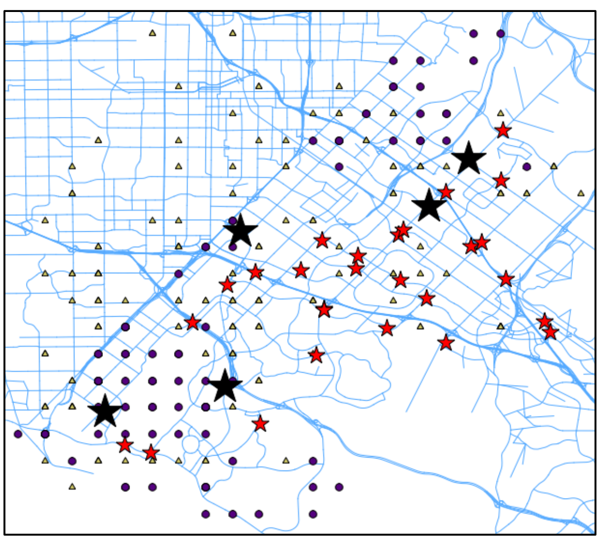
[9] This includes the construction of each a_j at each iteration. Therefore, it has double counted some of a_{hj} that have been constructed both for search and for entering variable.

[10] Due to degeneracy, all initial column parameters a_{hj} are checked.

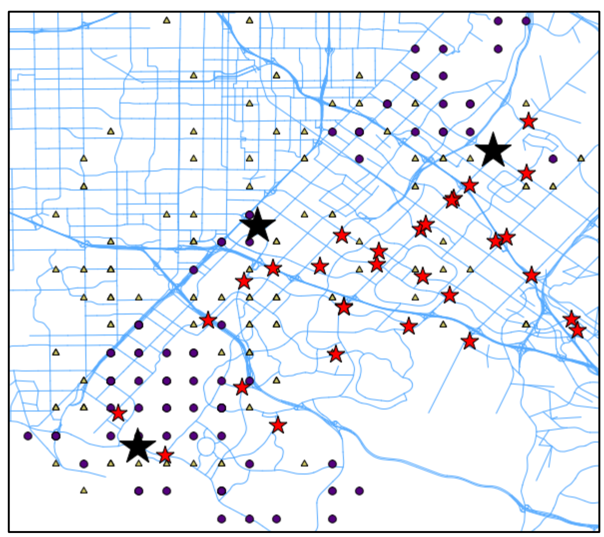
[11] This is based on direct distance from home. Accounting for a round trip, tolerance/2 is used measure the coverage.

Since we used uniform cost, there are a number of different combinations of locations that satisfy the optimality conditions for this particular example. The results for one set of optimal refueling locations are shown in Figure 5 for four different additional travel time tolerances, together with a comparison to the solution to the classical Set Covering Problem (for the case of 0.3 hr and 0.4 hr additional travel time) for this application. The results indicate that, based on the reported daily activity agendas (assumed repetitive) of households in the study area, Location-HAPP identified that full coverage within a 0.4 hr additional travel time window could be achieved with only one refueling station—tolerances of 0.15 hr, 0.2 hr, and 0.3 hr could be achieved with 2, 3 and 5 stations, respectively; ignoring these patterns of travel and based only on the home locations of the residents, full coverage within a 0.4 hr additional travel time required 2 refueling stations and 0.3 hr required 5 refueling stations —there are no feasible solutions from Set Covering Problem for tolerances of 0.15 hr and 0.2 hr.

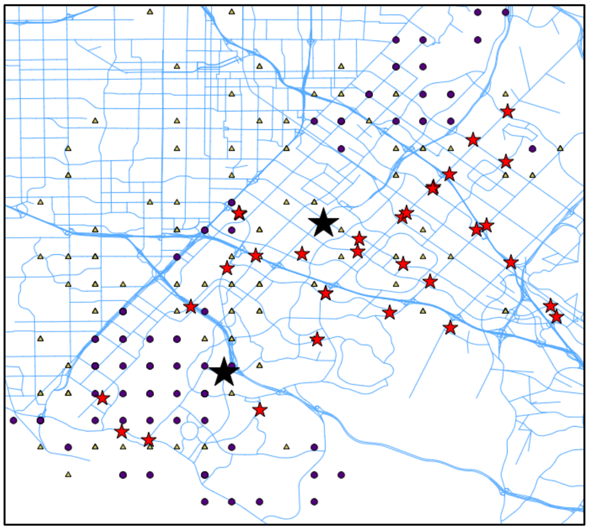
Another observation is that the Location-HAPP favors areas of high volumes of activities (such as central business district or shopping centers) whereas the Set Covering Model favors the residential areas. By the definitions, we only include home locations to be covered in the Set Covering Model vs. one of all activity locations to be covered in the Location-HAPP. If there exist certain activity/travel patterns induced by land use, it is reasonable that the master problem in the Location –HAPP would favor a high activity volume area in a collective sense.



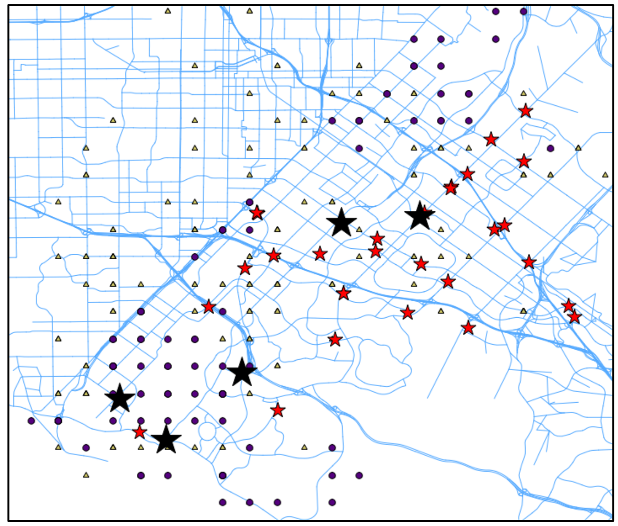
(a) Case 1 ($L = 0.15hr$)



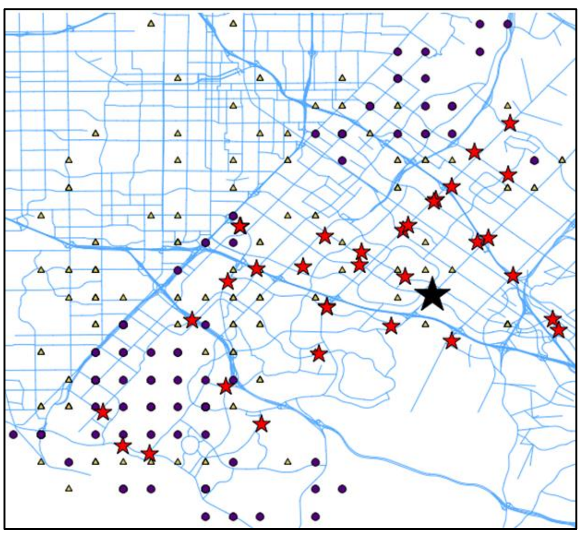
(b) Case 2 ($L = 0.2hr$)



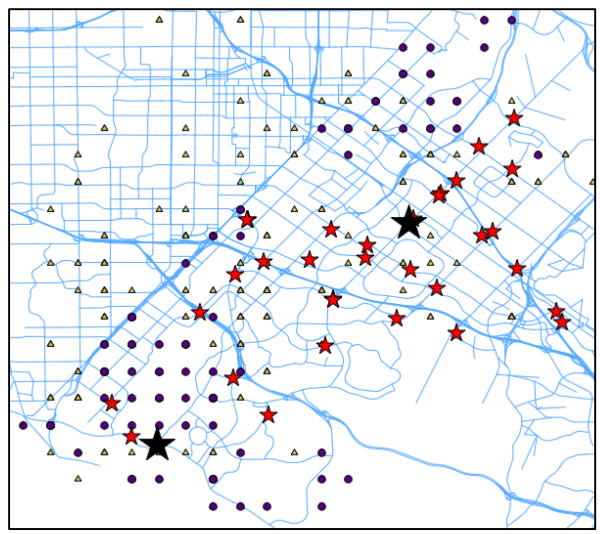
(c) Case 3 ($L = 0.3hr$)



(d) SCP: Case 3($L = 0.3hr$)



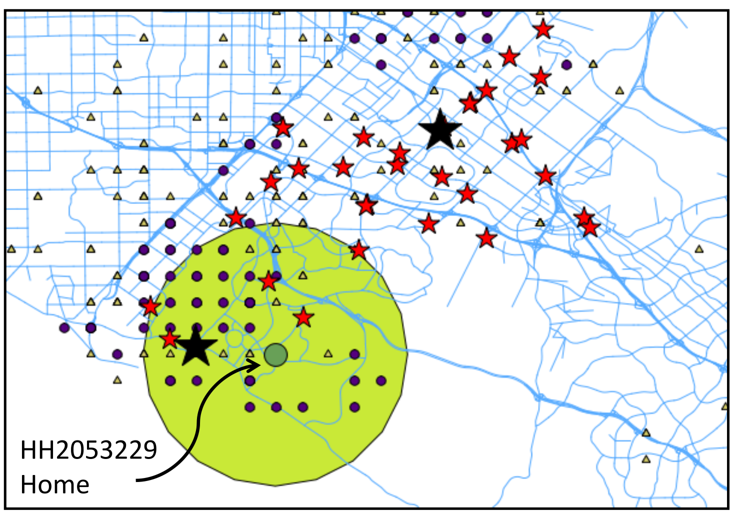
(e) Case 4 ($L = 0.4hr$)



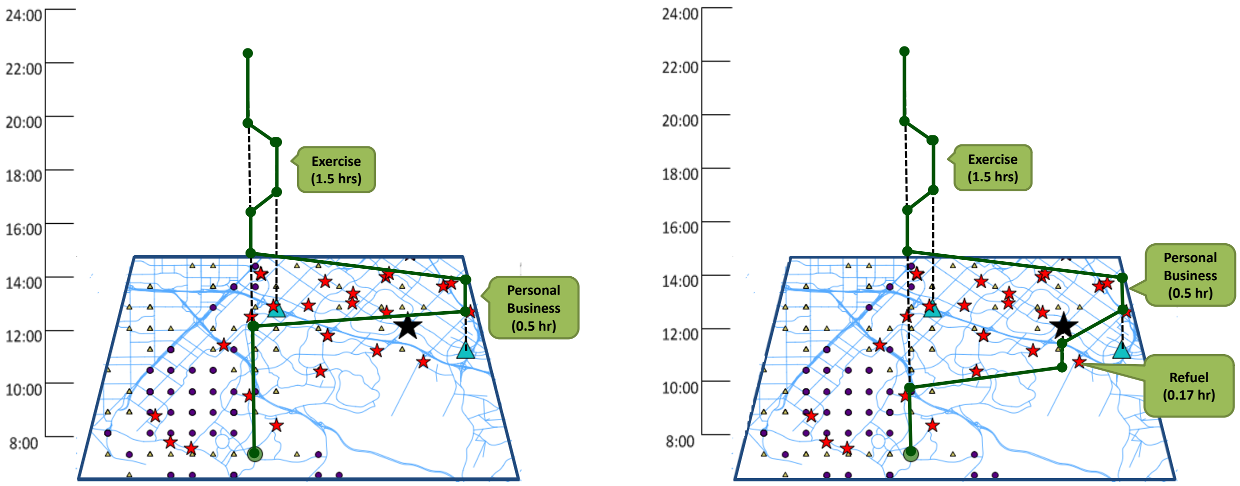
(f) SCP: Case 4 ($L = 0.4hr$)

Figure 5. Location– HAPP and Set Covering Problem Results

We can also conceptualize the accessibility in the Location-HAPP with an example. Because the Set Covering Model focuses only on the accessibility of given points (for our case, home locations), the accessibility range is limited to the home location, as seen in Figure 6a, for the case of household #253229 in the Travel Survey, under the presumption that refueling will take place at the station located within the shaded area. In the Location-HAPP model, accessibility is not limited to home but includes accessibility along the vehicle movements over the space throughout the day, expanding the area of coverage, albeit subject to the temporal constraints imposed by the available time windows for completion of each household's activity agenda. This also presumes grouping/touring refueling with other compulsory activities is an acceptable alternative. Although there appears to be evidence that drivers' tend to prefer to refuel "near home" (Kitamura and Ogden, 1987), and some studies assume this preference (Nicolas and Ogden, 2006; Nicolas et al., 2004), behavioral changes regarding refueling activity have been observed at the early adoption stage where there are limited refueling opportunities (Kelly and Kuby, 2012). For household #253229, the reported travel pattern is that shown in Figure 6b. The optimal (minimization of travel time) travel/activity pattern for this particular household obtained from HAPP is shown in Figure 6b. Since one of the compulsory activities is near one of the refueling candidate locations, this household completes the refueling activity on the way to "personal business" activity at the refueling station that is located close to that "personal business" activity location (Figure 6c), rather than at the station closest to home. Were the reported activity agenda for this household on the day of the survey repetitive (representative of this particular household's daily routine), the service provider's need to locate a facility near its home location could been substituted by a refueling location that would have been not considered as accessible by the Set Covering Model constraints.



(a) Accessibility of HH 2053229 within 0.4 hr Using Direct Distance Measurement



(b) Reported Travel Pattern (c) Travel Pattern with an added Refueling Trip

Figure 6. Refueling Pattern of HH 2053229 (Case 4 = 0.4 hr)

Conclusions and Future Research

In this paper, we have developed a facility location problem with full-day scheduling and routing considerations. This is in the category of Location-Routing Problems (LRPs), where the decisions of facility location models are influenced by possible vehicle routings. The model we propose takes the classic coverage model as a location strategy, and the Household Activity Pattern Problem (HAPP) as the scheduling and routing tool. The Location-HAPP includes multiple decision makers: the public/private sector as the refueling service provider, and each individual that makes his/her own routing decisions, including a visit to one of the service locations. It can be classified as a LRP that addresses public/private sector's refueling service provision using a non-standard structure that uses an exact solution method via decomposition. It also is an extension from node-based and path-based models to a tour-based model—to be exact, a special case of a tour-based model in a way that it allows changes in the sequence within the specified time windows.

The proposed formulation isolates each vehicle's routing problem from those of other vehicles as well as the master set covering problem. However, its coverage matrix requires the solving of $|H| \times n_R$ (number of households x number of candidate locations) HAPPs. A modified column generation that finds a column with a negative reduced price, but not necessarily the most negative, is developed. This way, only partial knowledge of the full coverage matrix is needed. A search method is developed for finding such columns in order to reduce the number of HAPPs to be solved. The performance of the methodology is described by the percentage of HAPPs that are actually solved to the total number of HAPPs in the full coverage matrix.

Although general to the location problem of any service facility that can be considered as ancillary to the spatio-temporal movement of households as they complete their daily routines, the specific application developed here relates to the incubation of the minimum refueling infrastructure that might be required to support early adoption of Hydrogen Fuel Cell Vehicles (HFCVs). The proposed model and methodology are applied to a case study of HFCV refueling stations in Irvine/Newport Beach community—one of four early hydrogen vehicle adoption communities targeted by auto manufacturers—subject to three different values of accessibility measured in terms of tolerances to added travel time. Under optimal conditions, refueling trips are found to be mostly toured with other activities, and this traveling behavior is captured by HAPP. We have tested four different levels of service provisions, and the suggested method shows that only 6% – 35% of sub problems are needed to be solved, compared to direct calculation. More importantly, from the results and the coverage matrices, there is evidence that excluding such vehicle-infrastructure interactions as well as routing and scheduling interactions can result in over-estimation of minimum facility requirement.

There has been a major simplification of imposing one refueling trip per day per a vehicle mainly due to limitations imposed by the scarcity of travel diary data covering more than a single day. For many activity-based travel behavior modeling studies, data limitation has been one of the major issues. This may be overcome using simulations to synthesize activity/travel patterns as they are developed more fully; it will certainly help in aggregating results to a whole population, as well as better analyze the refueling needs based on fuel inventory. Also, congestion effects and possible queuing of refueling service need to be evaluated. Another shortcoming relative to travel behavior is that we have ignored the intra-household interactions of travel decisions. It is expected to be the case that if an AFV is one of the vehicle sets in a household, some of the activities may be shifted to other vehicles depending on the length, property, and exchangeability of activities; such an extension could logically be based on HAPP cases 2 – 5.

Acknowledgement

This research was supported, in part, by grants from the University of California Transportation Center and the ITS Multi-Campus Research Program and Initiative on Sustainable Transportation. Their support is gratefully acknowledged. The authors would also like to thank Hang Liu, PhD Candidate at Institute of Transportation Studies, University of California, Irvine, for providing a program of finding all-to-all travel times using *MapQuest API*, and modifying it to meet our specific data needs.

References

Barnhart, C., Johnson, E. L., Nemhauser, G. L., Savelsbergh, M. W. P., Vance, P. H. (1998), Branch-and-Price: Column Generation For Solving Huge Integer Programs, *Operations Research*, 46, pp. 316-329

California Fuel Cell Partnership (CaFCP) (2009), Hydrogen Fuel Cell Vehicle and Station Deployment Plan: A Strategy for Meeting the Challenge Ahead, Action Plan, West Sacramento, CA

California Fuel Cell Partnership (CaFCP) (2010), Hydrogen Fuel Cell Vehicle and Station Deployment Plan: A Strategy for Meeting the Challenge Ahead, Progress and Next Steps, West Sacramento, CA

CalTrans (2001), 2000-2001 California Statewide Household Travel Survey

Caprara, A., Toth, P., Fischetti, M. (2000), Algorithms for the Set Covering Problem, *Annals of Operations Research*, 98, pp. 353-371

Chan, Y., Carter, W. B., Burnes, M. D. (2001), A multiple-Depot ,multiple-Vehicle, Location-Routing Problem with Stochastically Processed Demands, *Computers and Operations Research*, 28, pp. 803-826

Christofides, N. and Korman, S. (1975), A Computational Survey of Methods For The Set Covering Problem, *Management Science*, 21, pp. 591-599

Cordeau, J. F. and Laporte, G. (2003), The Dial-a-Ride Problem: Variants, modeling issues and algorithms, *4OR: A Quarterly Journal of Operations Research*, 1-2, pp. 89-101

Desrosiers, D., Dumas, Y., Soumis, F. (1986), A dynamic programming solution of the large-scale single-vehicle dial-a-ride problem with time windows, *American Journal of Mathematical and Management Sciences*, 6, pp. 301-325

Hodgson, M. J. (1990), A Flow Capturing Location-Allocation Model, *Geographical Analysis*, 22, pp. 270–279

Kang, J. E. and Recker, W. W. (2009), An activity-based assessment of the potential impacts of plug-in hybrid electric vehicles on energy and emissions using 1-day travel data, *Transportation Research Part D: Transport and Environment*, 14, pp. 541–556

Kang, J. E. and Recker, W. W. (2012), The Location Selection Problem for the Household Activity Pattern Problem, *ITS Working Paper Series, UCI-ITS-WP-12-1, University of California Irvine.*

Kang, J. E., Chow, J. Y. J., and Recker, W. W. (2012), On Activity-Based Network Design Problems, *the 20th International Symposium on Transportation and Traffic Theory, Accepted, To appear in Transportation Research Series*

Kim, J-G. and Kuby, M. (2012), The Deviation-Flow Refueling Location Model for Optimizing a Network of Refueling Stations, *International Journal of Hydrogen Energy*, 37, pp. 5406-5420

Kitamura, R. and Sperling, D. (1987), Refueling Behavior of Automobile Drivers, *Transportation Research Part A: Policy and Practice*, 3, pp. 235-245

Kuby, M. and Lim, S. (2005), The Flow-Refueling Location Problem for Alternative-Fuel Vehicles, *Socio-Economic Planning Sciences*, 39, pp. 125-148

Kuby, M. et al. (2009), Optimization of Hydrogen Stations in Florida Using the Flow-Refueling Location Model, *International Journal of Hydrogen Energy*, 34, pp. 6045-6064

List, G. F., Mirchandani, P. B., Turnquist, M. A. and Zografos, K. G. (1991), Modeling and Analysis for Hazardous Materials Transportation: Risk Analysis, Routing/Scheduling and Facility Location, *Transportation Science*, 25, pp. 100-114

Los Angeles Times (accessed in 2012), http://articles.latimes.com/2012/aug/01/business/la-fi-autos-hydrogen-20120802

Lübbecke, M. E. (2011), Column Generation, *Wiley Encyclopedia of Operations Research and Management Science*, DOI: 10.1002/9780470400531.eorms0158

Melaina, M. (2003), Initiating Hydrogen Infrastructures: Preliminary Analysis of a Sufficient Number of Initial Hydrogen Stations in the US, *International Journal of Hydrogen Energy*, 28-7, pp. 743-755

Min, H., Jayaraman, V. and Srivastava, R. (1998), Combined Location-Routing Problems: A Synthesis and Future Research Directions, *European Journal of Operations Research*, 108-1, pp. 1-15

MirHassani S. A. and Ebrazi, R. (2012), A Flexible Reformulation of the Refueling Station Location Problem, *Transportation Science*, doi: 10.1287/trsc.1120.0430

Nagy, G. and Salhi, S. (1998), The Many-to-Many Location-Routing Problem, *TOP*, 6, pp. 261-275

Nagy, G. and Salhi, S. (2007), Location-Routing: Issues, Models and Methods, *European Journal of Operations Research*, 177-25, pp. 649-672

National Fuel Cell Research Center (NFCRC), University of California, Irvine (accessed in 2012), http://www.nfcrc.uci.edu/2/ACTIVITIES/PROJECTS/hydrogen/HydrogenFuelingStations.aspx#top

Nicholas, M. A. and Handy, S., and Sperling, D. (2004), Using Geographic Information Systems to Evaluate Siting and Networks of Hydrogen Stations, *Transportation Research Record*, 1880, pp. 126-134

Nicholas, M. A. and Ogden, J. (2006), Detailed Analysis of Urban Station Siting for California hydrogen Highway Network, *Transportation Research Record*, 1883, pp. 121-128

Or, I. and Pierskalla, W. P. (1979), A Transportation Location-Allocation Model for Regional Blood Banking, *AIIE Transactions*, 11, pp. 86-94

Recker, W.W. (1995), The Household Activity Pattern Problem: General Formulation and Solution, *Transportation Research*, 29B, 1, pp. 61-77

Recker, W. W and Parimi, A. (1999), Development of a Microscopic Activity-Based Framework for Analyzing the Potential Impacts of Transportation Control Measures on Vehicle Emissions, *Transportation Research Part D: Transport and Environment*, 4, pp. 357-378

Kelly S., and Kuby M. (2012), Spatial refueling patterns of compressed natural gas (CNG) drivers in Southern California. In: Informatics rising: INFORMS 2012 annual meeting, Phoenix, USA, 14-17 October; 2012; p. 195.

Stephens-Romero, S. D., Brown, T. M., Kang, J. E., Recker, W. W., and Samuelsen, G. S. (2010), Systematic Planning to Optimize Investments in Hydrogen Infrastructure Deployment, *International Journal of Hydrogen Energy*, 35-10, pp. 4652-4667

Stephens-Romero, S. D., Brown, T. M., Carreras-Sospedra, M., Kang, J. E., Brouwer, J., Dabdub, D., Recker, W. W., and Samuelsen, G. S. (2011), Projecting Full Build-out Environmental Impacts and Roll-out Strategies Associated with Viable Hydrogen Fueling Infrastructure Strategies, *International Journal of Hydrogen Energy*, 36, pp. 14309-14323

Xi, X., Sioshansi, R., and Marano, V. (2012), A Simulation-Optimization Model for Locating of a Public Electric Vehicle Charging Infrastructure, *Transportation Science, Under Review*

Appendix

Define the following sets that are specific for each household, $h \in \mathbf{H}$

$\mathbf{V}_h = \{1, 2, ..., |\mathbf{V}_h|\}$: set of vehicles available to travelers in household h to complete their scheduled activities, one or more of which is a HFCV.

$\mathbf{P}_h^{R+} = \{1, 2, ..., n_R\}$: set designating candidate location at which each service type R can be performed. Index numbers and physical locations of this set are identical for all households, but the set is defined specifically for each household since index numbers of "return home" nodes from this set are different across households.

$\mathbf{P}_h^{A+} = \{n_R + 1, n_R + 2, ..., n_h = n_R + n_A^h\}$: set designating location at which each assigned activity is performed for household, h. Each activity and the physical location is different for each household.

$\mathbf{P}_h^{+} = \mathbf{P}_h^{R+} \cup \mathbf{P}_h^{A+} = \{1, 2, ..., n_R, n_R + 1, n_R + 2, ..., n_h = n_R + n_A^h\}$: set designating location at which each activity is performed for household h.

$\mathbf{P}_h^{R-} = \{n_h + 1, n_h + 2, ..., n_h + n_R\}$: set designating the ultimate destination of the "return to home" trip from candidate locations of service type R. Physical locations of this set are identical for all households, but the set is defined specifically for each household since index numbers are different across households. (It is noted that the physical location of each element of $\mathbf{P}_h^{R-}$ is "home".)

$\mathbf{P}_h^{A-} = \{n_h + n_R + 1, n_R + 2, ..., 2n_h = 2(n_R + n_A^h)\}$: set designating the ultimate destination of the "return to home" trip from out-of-home activities to be completed by travelers in household, h. (It is noted that the physical location of each element of $\mathbf{P}_h^{A-}$ is "home".)

$\mathbf{P}_h^{-} = \mathbf{P}_h^{R-} \cup \mathbf{P}_h^{A-} = \{n_h + 1, n_h + 2, ..., n_h + n_R, n_h + n_R + 1, n_R + 2, ..., 2n_h = 2(n_R + n_A^h)\}$: set designating the ultimate destination of the "return to home" trip for each activity for household h. (It is noted that the physical location of each element of $\mathbf{P}_h^{-}$ is "home".)

$\left[a_i^h, b_i^h\right]$: time window of available start times for activity i for household h. (Note: b_i^h must precede the closing of the availability of activity i of household h, by an amount equal to or greater than the duration of the activity.)

$\left[a_{n_h+i}^h, b_{n_h+i}^h\right]$: time windows for the "return home" arrival from activity i of household h.

$\left[a_0^h, b_0^h\right]$: departure window for the beginning of the travel day for household h.

$\left[a_{2n_h+i}^h, b_{2n_h+i}^h\right]$: arrival window by which time all members of the household h must complete their travel.

s_i^h: duration of activity i of household h.

t_{uw}^h: travel time from the location of activity u to the location of activity w.

$c_{uw}^{\upsilon,h}$: travel cost for household h, from location of activity u to the location of activity w by vehicle υ.

B_C^h: travel cost budget for household h.

$B_T^{\upsilon,h}$: travel time budget for the household h's member using vehicle υ.

$\mathbf{P}_h = \mathbf{P}_h^+ \cup \mathbf{P}_h^-$: set of nodes comprising completion of all the activities of household h.

$\mathbf{P}_h^A = \mathbf{P}_h^{A+} \cup \mathbf{P}_h^{A-}$: set of nodes comprising completion of the activities household h. This does not include trips related to service type R

$\mathbf{P}_h^R = \mathbf{P}_h^{R+} \cup \mathbf{P}_h^{R-}$: set of nodes comprising completion of service type R of household h.

$\mathbf{P}_h = \mathbf{P}_h^{A+} \cup \mathbf{P}_h^{A-} \cup \mathbf{P}_h^{R+} \cup \mathbf{P}_h^{R-} = \mathbf{P}_h^A \cup \mathbf{P}_h^R$: set of nodes comprising completion of the household's scheduled and service type R activities.

$\mathbf{N}_h = \left\{0, \mathbf{P}_h, 2n_h + 1\right\}$: set of all nodes for household h, including those associated with the initial departure and final return to home.

$$\min \sum_{h \in \mathbf{H}} O^h = \sum_{w \in \mathbf{N}_h} \text{Travel Disutility of Household } h \tag{A1}$$

$$\sum_{v \in \mathbf{V}_h} \sum_{w \in \mathbf{N}_h} X_{uw}^{v,h} = 1, \quad u \in \mathbf{P}_h^{A+}, h \in \mathbf{H} \tag{A2}$$

$$\sum_{u \in \mathbf{P}_h^{R+}} \sum_{w \in \mathbf{N}_h} X_{uw}^{v,h} = 1, \quad v \in \mathbf{V}_h, h \in \mathbf{H} \tag{A3}$$

$$\sum_{w \in \mathbf{N}_h} X_{uw}^{v,h} - \sum_{w \in \mathbf{N}_h} X_{wu}^{v,h} = 0, \quad u \in \mathbf{P}_h, v \in \mathbf{V}_h, h \in \mathbf{H} \tag{A4}$$

$$\sum_{w \in \mathbf{P}_h^+} X_{0w}^{v,h} \le 1, \quad v \in \mathbf{V}_h, h \in \mathbf{H} \tag{A5}$$

$$\sum_{u \in \mathbf{P}_h^-} X_{u,2n_h+1}^{v,h} - \sum_{u \in \mathbf{P}_h^+} X_{0u}^{v,h} = 0, \quad v \in \mathbf{V}_h, h \in \mathbf{H} \tag{A6}$$

$$\sum_{w \in \mathbf{N}_h} X_{wu}^{v,h} - \sum_{w \in \mathbf{N}_h} X_{w,n_h+u}^{v,h} = 0, \quad u \in \mathbf{P}_h^+, v \in \mathbf{V}_h, h \in \mathbf{H} \tag{A7}$$

$$T_u^h + s_u^h + t_{u,w}^h = T_{n_h+u}^h, \quad u, w \in \mathbf{P}_h^+, h \in \mathbf{H} \tag{A8}$$

$$X_{uw}^{v,h} = 1 \Rightarrow T_u^h + s_u^h + t_{u,w}^h = T_w^h, \quad u, w \in \mathbf{P}_h, v \in \mathbf{V}_h, h \in \mathbf{H} \tag{A9}$$

$$X_{0w}^{v,h} = 1 \Rightarrow T_0^{v,h} + s_u^h + t_{0w}^h = T_w^h, \quad w \in \mathbf{P}_h^+, v \in \mathbf{V}_h, h \in \mathrm{H} \tag{A10}$$

$$X_{u,2n_h+1}^{v,h} = 1 \Rightarrow T_u^h + s_u^h + t_{u,2n_h+1}^h = T_{2n_h+1}^h, \quad u \in \mathbf{P}_h^-, v \in \mathbf{V}_h, h \in \mathbf{H} \tag{A11}$$

$$a_u^h \le T_u^h \le b_u^h, \quad u \in \mathbf{P}_h, h \in \mathbf{H} \tag{A12}$$

$$a_0^h \le T_0^{v,h} \le b_0^h, \quad v \in \mathbf{V}_h, h \in \mathbf{H} \tag{A13}$$

$$a_{2n_h+1}^h \le T_{2n_h+1}^{v,h} \le b_{2n_h+1}^h, \quad v \in \mathbf{V}_h, h \in \mathbf{H} \tag{A14}$$

$$X_{uw}^{v,h} = 1 \Rightarrow Y_u^h + d_w = Y_w^h, \quad u \in \mathbf{P}_h, w \in \mathbf{P}_h^+, v \in \mathbf{V}_h, h \in \mathrm{H} \tag{A15}$$

$$X_{uw}^{v,h} = 1 \Rightarrow Y_w^h = 0, \quad u \in \mathbf{P}_h, \mathrm{w} \in \mathbf{P}_h^-, v \in \mathbf{V}_h, h \in \mathbf{H} \tag{A16}$$

$$X_{0w}^{v,h} = 1 \Rightarrow Y_0^h + d_w = Y_w^h, \quad w \in \mathbf{P}_h^+, v \in \mathbf{V}_h, h \in \mathbf{H} \tag{A17}$$

$$Y_0^h = 0, \quad 0 \le Y_u^h \le D, \quad u \in \mathbf{P}_h^+, h \in \mathbf{H} \tag{A18}$$

$$\sum_{v \in \mathbf{V}_h} \sum_{u \in \mathbf{N}_h} \sum_{w \in \mathbf{N}_h} c_{uw}^{v,h} \cdot X_{uw}^{v,h} \le B_C^{v,h}, \quad h \in \mathbf{H} \tag{A19}$$

$$\sum_{v \in \mathbf{V}_h} \sum_{u \in \mathbf{N}_h} \sum_{w \in \mathbf{N}_h} t_{uw}^h \cdot X_{uw}^{v,h} \le B_T^{v,h}, \quad v \in \mathbf{V}_h, h \in \mathbf{H} \tag{A20}$$

The household-specific decision variables in (A1)-(A20) are:

$X_{uj}^{v,h}$, $u,w \in \mathbf{N}_h, v \in \mathbf{V}_h, h \in \mathbf{H}$	binary decision variable equal to unity if vehicle υ travels from activity u to activity w, and zero otherwise.
T_u^h, $u \in \mathbf{P}_h, h \in \mathbf{H}$	the time at which participation in activity u of household h begins.
$T_0^{v,h}, T_{2n_h+1}^{v,h}$, $u \in \mathbf{P}_h, h \in \mathbf{H}$	the times at which vehicle υ from household h first departs from home and last returns to home, respectively
Y_u^h, $u \in \mathbf{P}_h, h \in \mathbf{H}$	the total accumulation of either sojourns or time (depending on the selection of D and d_u) of household h on a particular tour immediately following completion of activity u.