

# Developing a Test Data Set for Electric Vehicle Applications in Smart Grid Research

Hossein Akhavan-Hejazi<sup>†</sup>, Hamed Mohsenian-Rad<sup>†</sup>, and Ali Nejat<sup>‡</sup>

<sup>†</sup>Department of Electrical Engineering, University of California, Riverside, CA, USA

<sup>‡</sup>Department of Construction Engineering, Texas Tech University, Lubbock, TX, USA

Emails: {shejazi, hamed}@ee.ucr.edu, ali.nejat@ttu.edu

**Abstract**—We analyze a detailed set of driving traces for 536 GPS-equipped taxi vehicles and combine them with the features of four different plug-in hybrid electric vehicle (PHEV) brands that currently dominate the North American market in order to develop a test data set for PHEV-related research in the field of smart grid. Our developed data set is made available to public in [1]. It consists of various information, including but not limited to per-PHEV traces of state-of-charges (SoCs), per-PHEV traces of charging loads at different carefully identified charging stations, per-PHEV information on SoC and charging deadline when the PHEV is parked at a charging station, and some information about the potential of PHEVs for vehicle-to-grid applications.

**Keywords:** Plug-in electric vehicles, publicly available data set, smart grid applications, experimental vehicle driving traces, state of charge, electric vehicle charging load, mobility patterns.

## I. INTRODUCTION

A key part of smart grid research is to use experimental data, both on the demand side and the generation side, to validate the results. There are currently various publicly available data sets for electricity prices, solar panels, wind turbines, and residential and commercial load profiles [2]–[4]. However, due to the *still-insignificant* penetration of plug-in hybrid electric vehicles (PHEVs), there currently does not exist any detailed data set for large fleets of PHEVs. This has caused major obstacles for several smart grid researchers, including the authors, who intend to investigate the challenges and opportunities that the PHEVs may introduce to power systems.

Of course, more data is expected to gradually become available as more experimental PHEV projects are conducted over the next couple of years. However, for now, one option is to use the existing major *non-PHEV* vehicular data sets and combine them with the information and features of the recently emerged commercial PHEVs in order to *synthesize* new PHEV data sets that can be used in a variety of smart grid research studies. Therefore, the central argument in this paper is that, in the absence of a detailed database of PHEV loads, a *synthetic database* that includes the key pieces of information that are needed for PHEV-related smart grid research is highly useful. Developing such practical data set is our goal in this paper.

In this paper, we analyze the driving traces for 536 GPS-equipped non-PHEV taxi vehicles in [5] that are recorded for

a duration of three weeks in San Francisco, CA. We then *combine* the results with the features and technical characteristics of four different PHEV brands that currently dominate the North American market: Chevrolet Volt [6], Honda Accord Plug-in [7], Ford Fusion Energi [8], and Toyota Prius Plug-in [9]. Our analysis has resulted in a new test data set to support PHEV-related smart grid research. It provides per-PHEV traces of states-of-charge (SoCs), per-PHEV traces of charging loads, per-PHEV information on SoC and charging deadline when the PHEV is parked at a charging station, the detailed charging load at each of our three carefully identified charging stations, and some information on vehicle-to-grid (V2G) potentials.

The dataset that is developed in this paper [1], can be used for various smart grid research projects. Some of the applications include investigating the impact of the increasing PHEV loads on power distribution feeders and substations, design, operation, and control of V2G systems based on vehicle arrival and departure data, charger sizing and design as well as optimal charger placement, and obtaining more precise statistics regarding the SoCs based on driving patterns, etc.

The rest of this paper is organized as follows. In Section II, we will process the non-PHEV driving traces in [5] to create some preliminary data to extract the parking and driving events for each car, and to identify the right locations of charging stations. In Section III, we will add the technical information from various PHEVs to create new data sets that are specific to different PHEV brands. This will provide us with data sets for the charging load profiles and the V2G potentials at different charging stations. The paper is concluded in Section IV.

## II. PROCESSING THE NON-PHEV DRIVING TRACES

Consider the non-PHEV vehicles data in [5]. This data set is widely used in over 100 vehicular research projects, ranging from vehicular movement prediction to vehicular communications, e.g., see [10]–[12]. However, the potential of this detailed vehicular movement data set has not yet been investigated in the context of electric vehicles and smart grid.

This data set includes the driving traces of 536 taxis in San Francisco, CA. The data recording starts on May 17, 2008 and ends on June, 10, 2008. For each taxi, each data record comprises a *time-stamp*, *latitude*, *longitude*, and a flag indicating whether or not the taxi has a passenger. The GPS tracking system is switched off every time that the vehicle is turned off. Recording resumes once the vehicle is turned on.

This work is supported in part by the National Science Foundation Grant ECCS 1253516 and the California Energy Commission Grant EISG 57757A/13-04TE. The corresponding author is Hamed Mohsenian-Rad.

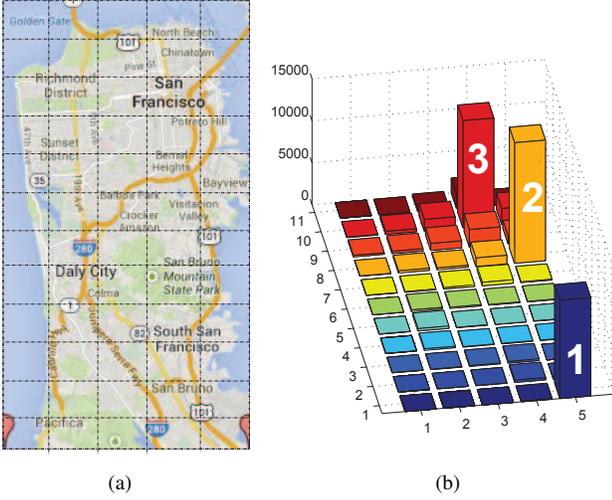


Fig. 1. By tracking the GPS locations of taxi vehicles in San Francisco, CA, we can identify the most frequent parking locations, based on our definition of a parking event: (a) The city map is divided into  $5 \times 12 = 60$  regions based on latitude and longitude. (b) The number of parking events, i.e., the frequency of parking, in each region. The regions that are marked from 1 to 3 are the most common parking locations: Airport, Taxi Headquarters, Downtown.

Since the taxis serve mainly the San Francisco area, most of the recorded GPS coordinates are within latitudes 37.6 and 37.82 and longitudes  $-122.52$  and  $-122.37$ . This rectangular area is shown in Fig. 1(a), where it is divided into  $5 \times 12 = 60$  equal-area regions, each having a rectangular shape.

As the first step in our analysis, we would like to identify the number of times that a vehicle is *parked* in each of the 60 regions in Fig. 1(a). The parking events are important as they are later interpreted, under certain conditions on their location and duration, as PHEV *plug-in* events. Of course, not every stationary behaviour of a vehicle should be interpreted as a parking event. In many cases, heavy traffic can be the cause of no movement or a very slow movement. Similarly, some temporary stops are when the taxi is waiting for a passenger, etc. Therefore, we define a parking event based on both a *time duration threshold* and a *distance threshold*. In this regard, a parking event is a scenario where the vehicle traversed a distance of 500 meters or less in 15 minutes.

As for the distance travelled between the successive GPS readings, such distance is approximated by a straight path between each two coordinates. Such approximation is reasonable for the purpose of our study since the reading intervals in [5] are fairly small. Let  $\lambda_1$  and  $\lambda_2$  denote the latitudes of two successive records of the GPS coordinates. Assume that  $\phi_1$  and  $\phi_2$  denote the longitudes of those coordinates. The direct distance traversed between the two points are calculated as

$$D_{1,2} = R \sqrt{\left( (\phi_2 - \phi_1) \cos \left( \frac{\lambda_1 + \lambda_2}{2} \right) \right)^2 + (\lambda_2 - \lambda_1)^2} \quad (1)$$

where  $R$  denotes the radius of the earth that is 6371 kilometres.

As we can see in the Fig. 1(b), most of the parking events are concentrated across *three* locations. We could identify

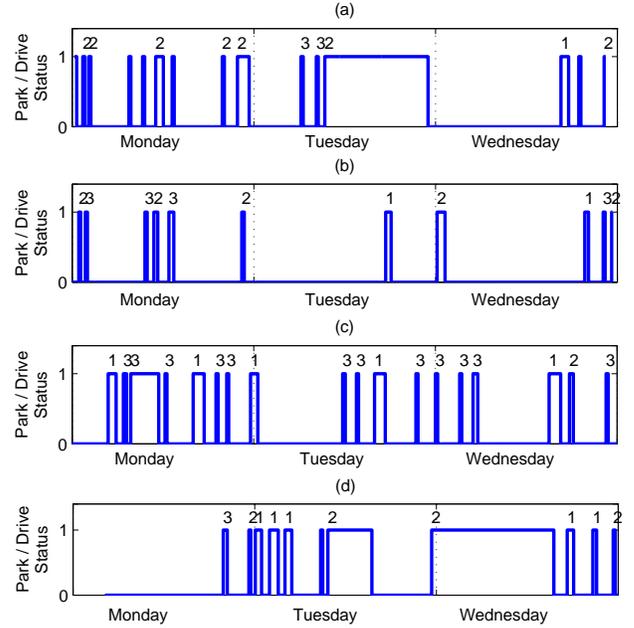


Fig. 2. Examples of drive (0) versus park (1) intervals for four sample taxis during a three-days window: (a) taxi number 1, (b) taxi number 7, (c) taxi number 10, and (d) taxi number 17. The numbers on top of each park interval range from 1 to 3 and indicate the index of the charging station.

these GPS locations on the San Francisco map in Fig. 1(a) as 1) Airport, 2) Taxi Depot and Headquarters, and 3) Downtown. Hence, it is reasonable to assume that if these taxis are replaced with PHEVs, then their charging stations must be placed in these three locations in order to provide them with charging service with minimum impact on their regular driving patterns. Accordingly, for the rest of this paper, we assume that there are indeed three *Charging Stations* for the taxis exactly at these three regions as numbered in Fig. 1(b).

Next, we record and analyze the movements and parkings of each of the 536 vehicles for the entire three weeks of data traces in [5]. The results are shown in Fig. 2 for four sample vehicles over a time window of three days. Here, the value 1 on the y-axis indicates that the vehicle is parked at one of the three charging stations; and the value of 0 means otherwise. The index number of the charging station for each parking event is shown with numbers 1 to 3 on top of the curves. We can see in Fig. 2 that different taxis have different movement and parking patterns. Another interesting observation is that, as expected, the longest duration parking events are recorded at the second charging station, i.e., the Taxi Headquarters.

By putting together the detailed driving and parking traces of *all* vehicles, we can next calculate the number of vehicles that are parked at each station at any time of interest. The results are shown in Fig. 3. Here, the resolution is one minute. That is, we have calculated the number vehicles that are parked at each charging station during every one minute interval of each day. We can see that the parking patterns are quite different across different charging stations, depending on the dynamics of vehicles movements. For example, the number of

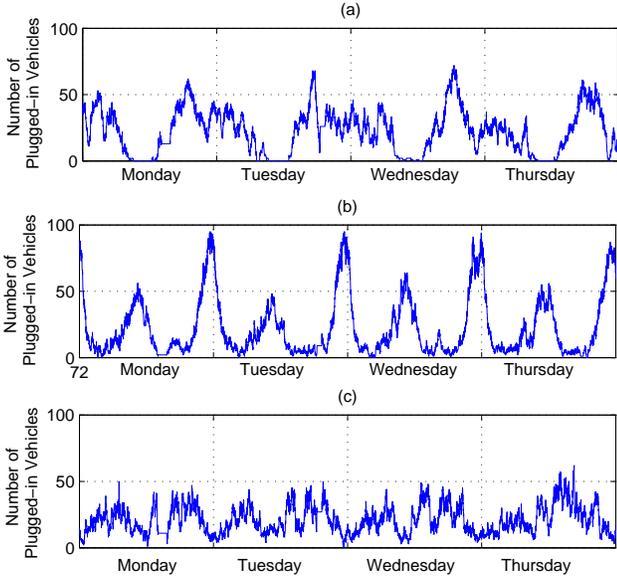


Fig. 3. The number of taxis that are parked at the three charging stations during a sample four-days window: (a) Charging station number 1, Airport, (b) Charging station number 2, Taxi Headquarters, (c) Charging station number 3, Downtown. The parking patterns are different at different charging stations.

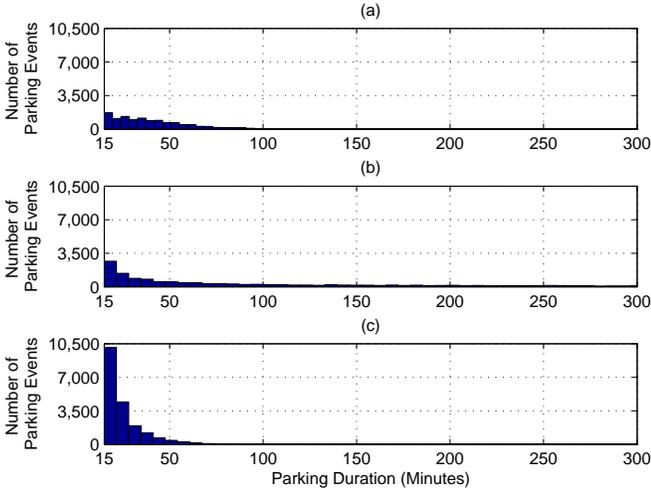


Fig. 4. The histograms of the duration of parking for each parking event: (a) Charging station number 1, Airport, (b) Charging station number 2, Taxi Headquarters, (c) Charging station number 3, Downtown.

stops vehicles that are parked at the Taxi Headquarters is at its peak during the night, when a large number of taxis return to the parking area. In contrast, there are few cars parking in the Downtown area at night. And of course, we see a completely different pattern at the Airport station, where the maximum number of vehicles are parked in late afternoon, when many flights arrive in the San Francisco area. The combination of the results in Fig. 3 with the stop duration and mileage driven by the vehicles could be the starting point for many useful information and/or applications as we will discuss next.

The histogram of the parking event durations at each charging station is shown in Fig. 4. We can see that while most

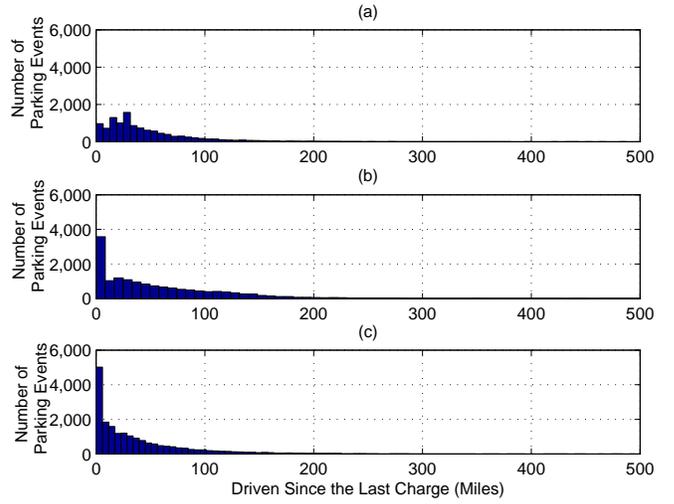


Fig. 5. The histograms of the miles driven since departing the previous charging station: (a) Charging station number 1, Airport, (b) Charging station number 2, Taxi Headquarters, (c) Charging station number 3, Downtown.

of the parking events at the Airport and Downtown stations have relatively short durations, there are several parking events that are one hour or longer at the Taxi Headquarters station. Similarly, the histograms for the distributions of the distance driven since departing the previous charging station for each vehicle that arrives at each of the three charging stations are shown in Fig. 5. This measure gives an indication about how full or empty the battery of a PHEV could be when it arrives at a charging station, allowing us to calculate the initial state-of-charge for the PHEVs that arrive at a charging stations, once we also take into account the electric consumption patterns of various PHEVs into consideration. From the results in Fig. 5, we can see that the vehicles typically drive for longer distances when they arrive at the Airport charging station.

For all the results that we have presented so far, our focus has been solely on the driving and parking patterns of various vehicles. As is, these results were already insightful in various aspects as we explained throughout this section. However, these results would be even more useful, in particular for the purpose of smart grid research, if they are combined with some more analysis based on the specific operational and technical characteristics of various PHEVs, as we will see next.

### III. ADDING THE FEATURES AND CHARACTERISTICS OF THE MOST COMMON PHEVs IN THE MARKET

In this section, we aim to combine the vehicles' movement and parking datasets that we generated in Section II with the nominal operation data of multiple PHEVs that dominate the present market. The goal is to obtain the charging patterns and SoCs of the vehicles with the same movement patterns, but different PHEV technologies. Here, we consider four PHEV brands: Chevrolet Volt, Honda Accord Plug-in, Ford Fusion Energi, and Toyota Prius. The main operational characteristics of these PHEVs are listed in Table I [13].

We can see that these vehicle brands are different in various operational characteristics. Characteristics such as available

TABLE I  
OPERATIONAL DATA OF FOUR COMMON PHEVs

Brand	Chevrolet	Honda	Ford	Toyota
Model	Volt	Accord	Fusion	Prius
Battery Capacity (KWh)	16	6.6	7.6	4.4
Available Energy (KWh)	8.8	3.8	7.1	3.2
Ave. Electric Range (M)	37	13	21	11
Max Charge Rate (KW)	3.5	6.6	3.5	3.5
Electric Consumption *	36	29	34	29
Gas Consumption **	2.7	2.2	2.3	2
Power Train index ***	D	B	D	B

\* KW/100 Miles  
\*\* Galons/100 Miles  
\*\*\* D: Charge Depleting, B: Charge Blending

energy, maximum charge rate, vehicle efficiency, and vehicle power train, etc. All such characteristics have direct impact on the charging load, state-of-charge, and V2G potentials. With respect to the power train, we have divided our selected vehicles into two groups: *charge depleting*, and *charge blending*. The PHEVs in the first group use electric power as long as there is electric energy stored in the batteries. They switch to gas power only after the batteries are depleted. However, the PHEVs in the second group may blend the power sources and use the gas engine to increase the torque in high speed movements even if the battery is not completely depleted. As a result, the SoC for the second group of PHEVs depends on both the miles travelled and also the *travelling speed*. Here, we assume that the charge blending vehicles switch from electric power to gas power at a speed threshold, set to 60 mph. Finally, it is worth clarifying that while Ford Fusion is capable of using technology, it typically runs in all electric power mode [8]. Therefore, it is categorized within the charge depleting group.

Next, we transform the driving and parking trace information of vehicles into their SoC data sets. When a vehicle is plugged-in to a charging station, it is charged by a 240 V 32 A chargers [14]. However, the *actual* charge rate for each PHEV is limited by its own charger interface, as listed in Table I. Therefore, we must calculate SoC during the charging period specifically based on the characteristics of each particular PHEV. Once a PHEV departs a charging station, its SoC will start to decrease based on its driving pattern and also its power train type, as we explained in the previous paragraph.

Four sample SoC trends for four representative taxis over a three-days time window are shown in Fig. 6. All the SoC curves in this figure are based on the Chevrolet Volt data. Recall that Chevrolet Volt has a charge depleting power train. It is interesting to compare the SoC curves in Fig. 6 with the drive and park intervals in Fig. 1. We can see that the SoC increases, following the charge rate of Chevrolet Volt, at every time that it is plugged-in to a charging station. Of course, if a charge interval is shorter than the time needed to fully charge the battery, then the vehicle leaves the charging station with a partially charged battery. Similarly, the SoC starts decreasing as the PHEV departs the charging station. We can see that, given the relatively small electric driving range of Chevrolet Volt, for a large portion of driving times,

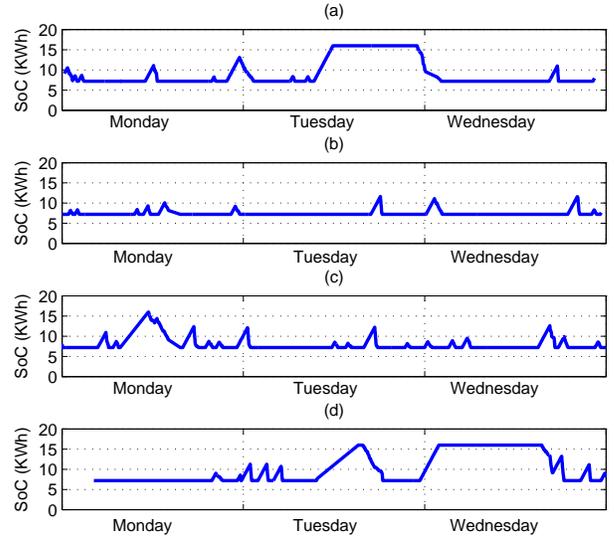


Fig. 6. The state-of-charge traces for four sample vehicles over a three-days time window: (a) taxi number 1, (b) taxi number 7, (c) taxi number 10, and (d) taxi number 17. All vehicles are assumed to be Chevrolet Volt.

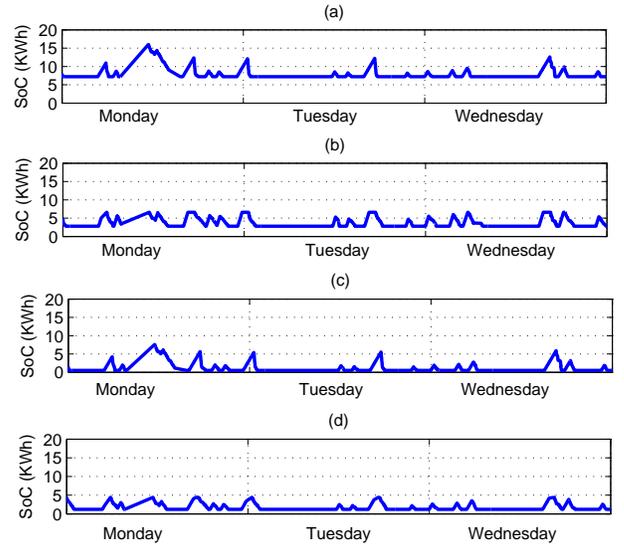


Fig. 7. The state-of-charge traces for the same taxi, taxi number 10, but based on assuming four different PHEV types: (a) Chevrolet Volt, (b) Honda Accord Plug-in, (c) Ford Fusion Energi, (d) Toyota Prius Plug-in.

the SoC is at its minimum level 7.2 kWh, i.e., battery capacity 16 kWh minus available energy 8.8 kWh, indicating that the battery is depleted and the power train is switched to gas.

Next, we compare the SoC traces of different PHEVs brands. We focus on one taxi, taxi number 10, and generate its SoC traces based on the characteristics of different PHEVs. The results are shown in Fig. 7. Note that, the traces change from one car to another, not only during the charging periods, but also during the driving periods. For example, if the taxi of interest is a Honda Accord Plug-in, then then it can almost fully charge its rather small 6.6 kWh battery in several of its short duration stops at charging stations.

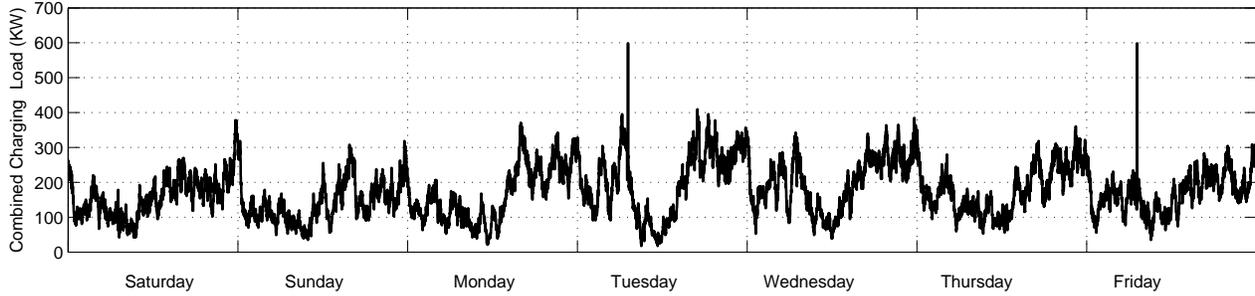


Fig. 8. The combined charging load of all 536 vehicles and across all three charging stations for a period of one week starting on May 24th. The PHEVs are a mixture of different types as explained in the text. We can see that the total charging load highly fluctuates during each day and across different days.

#### IV. AGGREGATE DATA SETS AT CHARGING STATIONS

In this section, we aggregate the data sets that we generated in Sections II and III and develop new data sets to provide useful information about the operation of charging stations. Recall from Section II that we identified three charging stations at the Airport, Taxi Headquarters, and Downtown. Our focus in this section is on calculating the *combined charging load* of all PHEVs as well as the total charging load at each charging station based on different PHEV types and during different hours of the day and different days of the week.

The combined charging load of all vehicles for a period of one week is shown in Fig. 8. The curve in this figure is the summation of the charging load of all the 536 PHEVs across all three charging stations. Therefore, it depends on not only the parking, charging, and driving patterns of each PHEV but also the exact specifications of the PHEVs. Here, we have assumed the following *mixture* of different PHEV brands:

- Chevrolet Volt: 161 vehicles,
- Honda Accord Plug-in: 125 vehicles,
- Ford Fusion Energi: 125 vehicles,
- Toyota Prius Plug-in: 125 vehicles.

We can see that the PHEV charging load curve in Fig. 8 fluctuates a lot. The average load during the considered week is 173.1 kWh, while the charging peak load is 598.5 kWh. Therefore, the *peak to average ratio* (PAR) is 3.457.

Next, we separately plot the total daily charging load at *each* of the three charging stations. The results are shown in Fig. 9. The mixture of the PHEV brands is the same as the one in Fig. 8. We can see that the load patterns across different charging stations are very different. At Airport charging station, the peak load is during the afternoon, where many taxis wait for incoming passengers. As for the Taxi Headquarters charging station, the peak load hours are at night, where several taxis are parked at the headquarters overnight.

Another interesting observation is that the load profiles differ on a weekday and a weekend. This is particularly the case at the Airport charging station and to some extent at the Downtown and Taxi Headquarters charging stations.

It is interesting also to look at the changes in the charging load at different charging stations when we change the mixture

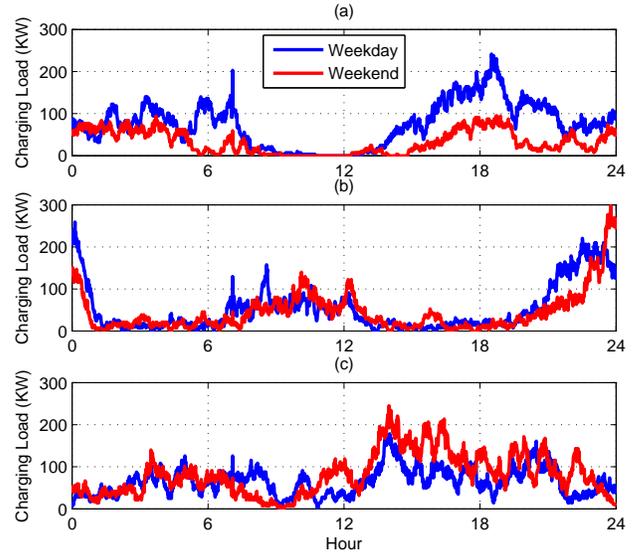


Fig. 9. The charging load at each of the three charging stations during a weekday and also during a weekend: (a) Charging station number 1, Airport, (b) Charging station number 2, Taxi Headquarters, (c) Charging station number 3, Downtown. The load curves are obtained based on a *mixture* of different PHEV brands as explained in the text.

of the PHEV brands and look at the scenarios where *all* PHEVs are of one brand. The results are shown in Fig. 10. We can see that the results are significantly different across some of the PHEV types because of the different features of these PHEVs. An interesting observation here is also about the differences between the four curves across the three charging stations. In particular, such differences are less significant at the second charging station, i.e., the Taxi Headquarters. This is because the duration of the parking events are longer at this station. Furthermore, it is notable that the curves in Fig. 10 for the cases of Chevrolet Volt and Ford Fusion are more or less similar. This can be tracked back by observing that the available charging energy and the charging rates of these vehicles are relatively close, as shown in Table I.

Another point to notice on, is that Honda Accord causes rather aggressive charging loads at short intervals, specifically at the Downtown charging station, where the parking durations

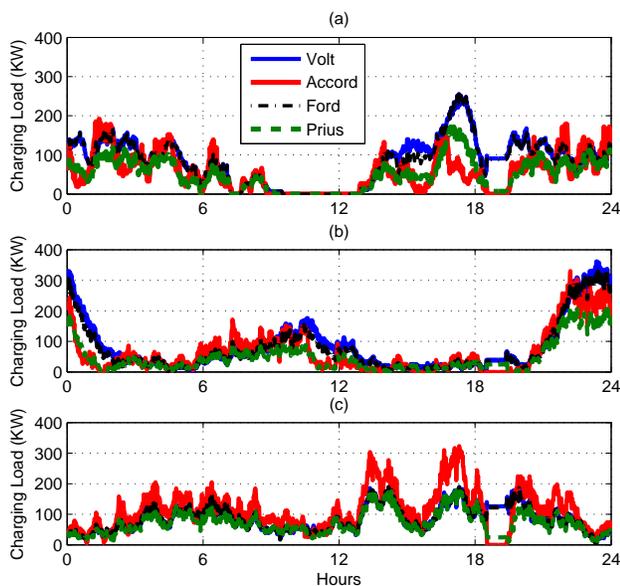


Fig. 10. The PHEV charging load at each of the three charging stations based on the operational characteristics of four different PHEV brands.

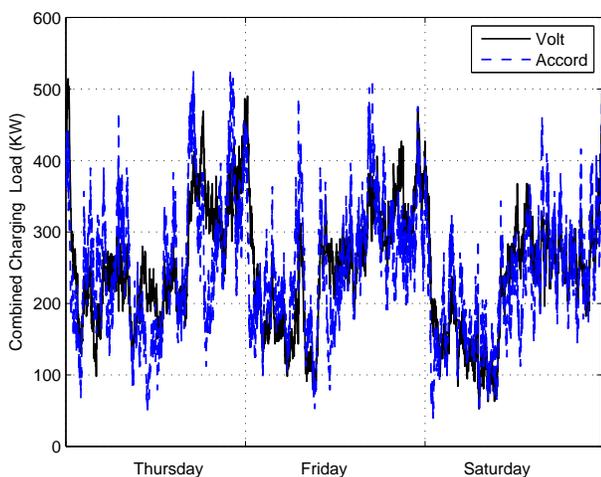


Fig. 11. A sample three-days of the combined PHEV load across all the three stations for two vehicle brands: Chevrolet Volt and Honda Accord Plug-In.

are shorter. Therefore, in order to further examine charging patterns, we look at the total EV load for the two vehicle brands, i.e. Chevrolet Volt, and Honda Accord.

Last but not least, let us look at the combined charging load across all three charging stations and compare the results for the scenario where *all* PHEVs are Chevrolet Volt with the scenario where *all* PHEVs are Honda Accord Plug-In. Recall that Chevrolet Volt has larger battery capacity but lower charging rate. In contrast, Honda Accord Plug-in has smaller battery capacity but faster charging rate. Therefore, we are comparing a long electric range and slow charging PHEV with a short electric range and fast charging PHEV. The results are shown in Fig. 11. Surprisingly, the combined charging load curves are *not* very different, even though there are considerable differences between the two PHEVs. Of course,

we do observe multiple intervals during the day where the two curves are noticeably different, such as the occasional high picks for the case of Honda Accord Plug-In, where the load spikes are twice than the average PHEV charging load.

## V. CONCLUSIONS AND FUTURE APPLICATIONS

We developed a data set, that is publicly available in [1], for a large 536 fleet of synthesized plug-in electric vehicle data to be used in PHEV-related research in the filed of smart grid. The generated data lines include the electric vehicle temporal and nodal charging loads, the minute-by-minute state-of-charge traces based on four brands of dominant PHEVs in the North American market, and the movement traces and parking analysis to identify the right locations for charging stations. In absence of a detailed measured information for large PHEV fleets, this synthesized data set is available as a useful in various smart grid studies in presence of PHEVs, such as the distribution feeder analysis, protection, planning, and optimal control as well as demand response [15]–[18]. It can be useful also for a number of charging station studies such as charging station sizing, placement, and operation. Moreover, it can also serve for evaluating many novel V2G technologies and algorithms based on the time duration of charging events.

Some of the limitations of our analysis include the estimation errors in GPS signals, the direct line distance calculation errors, and not knowing the true distance and time duration thresholds to define parking events. The characteristics considered for PHEVs are based on nominal values which are posted by PHEV manufacturers. The actual values may differ from one car to another even for the same PHEV brands.

## REFERENCES

- [1] <http://www.ee.ucr.edu/~hamed/PEVData.zip>.
- [2] <http://oasis.caiso.com>.
- [3] <http://www.windenergy.org/datasites/>.
- [4] <http://www.llnl.gov/>.
- [5] <http://crawdad.org/epfl/mobility/>.
- [6] <http://www.chevrolet.com/volt-electric-car.html>.
- [7] <http://automobiles.honda.com/accord-plug-in/>.
- [8] <http://www.ford.com/cars/fusion/trim/seenergi/>.
- [9] <http://www.toyota.com/prius-plug-in/>.
- [10] H. Saleet, O. Basir, R. Langar, and R. Boutaba, "Region-based location-service-management protocol for VANETs," *IEEE Trans. on Vehicular Technology*, vol. 59, pp. 917–931, Feb. 2010.
- [11] Y. Li, M. Zhao, and W. Wang, "Internode mobility correlation for group detection and analysis in VANETs," *IEEE Trans. on Vehicular Technology*, vol. 62, no. 9, pp. 4590–4601, Nov. 2013.
- [12] H. Luo, H. Zhang, and C. Qiao, "Efficient mobility support by indirect mapping in networks with locator/identifier separation," *IEEE Trans. on Vehicular Technology*, vol. 60, no. 5, pp. 2265–2279, June 2011.
- [13] <http://www.fueleconomy.gov/feg/label/learn-more-PHEV-label.shtml>.
- [14] [http://www.leviton.com/OA\\_HTML/SectionDisplay.jsp?section=60155](http://www.leviton.com/OA_HTML/SectionDisplay.jsp?section=60155).
- [15] S. Han, S. Han, and K. Sezaki, "Development of an optimal vehicle-to-grid aggregator for frequency regulation," *IEEE Trans. on Smart Grid*, vol. 1, no. 1, pp. 65–72, June 2010.
- [16] E. Sortomme and M. A. El-Sharkawi, "Optimal charging strategies for unidirectional vehicle-to-grid," *IEEE Trans. on Smart Grid*, vol. 2, no. 1, pp. 131–138, Mar. 2011.
- [17] C. Wu, H. Mohsenian-Rad, and J. Huang, "Vehicle-to-aggregator interaction game," *IEEE Trans. on Smart Grid*, vol. 3, no. 1, Mar. 2012.
- [18] H. Mohsenian-Rad, V. Wong, J. Jatskevich, R. Schober, and A. Leon-Garcia, "Autonomous Demand Side Management Based on Game-Theoretic Energy Consumption Scheduling for the Future Smart Grid," *IEEE Trans. on Smart Grid*, vol. 1, no. 3, pp. 320–331, Dec. 2010.