Using Synchronization as an Indicator of Controllability in a Fleet of Water Heaters

Eve Tsybina*, Viswadeep Lebakula*, Justin Hill**, Jeff Munk***, Helia Zandi*
* Oak Ridge National Laboratory, Knoxville, TN, USA
** Southern Company, Birmingham, AL, USA
*** National Renewable Energy Laboratory, Golden, CO, USA
Email: {lebakulav, tsybinae, zandih} @ ornl.gov, jeff.munk@nrel.gov, jhill@southernco.com

Abstract — **Peak reduction is an important concern that can help reduce the growing stress on distribution grid and allow to defer investments in new capacity. However, the growing concern for customer privacy and comfort may impact the performance of load control for residential devices. Water heaters represent a convenient way of reducing peak due to their ability to store thermal energy for future use. In this paper, we developed a methodology to help utilities gain more insight with respect to the impact of load control efforts for shaving peak with no necessary information about the water heaters except the device status (on/off). To this end, we use a fleet of water heaters in a controlled residential neighborhood in Atlanta, GA. Our findings show that convergence in device status can serve as a proxy for peak shifting during hours of the evening peak.**

Index Terms—**Demand response, smart grid, peak reduction, water heater, synchronization**

I. Introduction

Peak shifting is one of the emerging tools that utilities use to avoid peak charges or to defer infrastructure investments. The recent changes in load control have brought into the spotlight a few considerations which were not a concern in traditional load management approaches. One is the tradeoff between shedding load and maintaining user comfort [1]-[3]. This has given rise to demand flexibility as opposed to direct load control. Instead of sending a signal to connect or disconnect, the approach allows sending a utility signal such as price [4]-[5] as a part of the device cost function. This way the device on/off status will be based on its economic and comfort. The second is a growing concern for customer privacy [6]-[7]. There is evidence among the utilities that residential customers may be unwilling to disclose detailed information about their use of air conditioning or water, for a concern that this will reveal information about their intraday routines. Finally, there is a growing awareness of the communication and computation requirements which can be very large for a large fleet of devices [8]. These considerations affect the utility's visibility into the fleet of controllable devices and the controllability of devices.

This study makes investigates the impact of demand flexibility with only the power consumed by the load. To address this issue of lack of information, we leveraged the concept of controllability of devices, a property which is unobservable. More controllable devices show a larger change of kW load during hours of more intensive control and the hours immediately following it. They also show a greater change in the operating status. Even if a utility cannot say with confidence that the occurred change in total load of a given house is due to optimization efforts. There may be a way to assess the extent of device responsiveness by looking at the operating status. If the convergence in operating status is positively correlated with change in load, device status convergence can be used as a proxy for estimating peak shifting.

The rest of this study is organized as follows. Section II discusses the experimental setup, optimization approach, and data collection. Section III presents the results. Section IV offers the conclusion and discusses further research.

Eve Tsybina, Viswadeep Lebakula, and Helia Zandi are with Oak Ridge National Laboratory, TN, USA (e-mail: {tsybinae, lebakulav, zandih} @ ornl.gov).
Justin Hill is with Southern Company, Birmingham, AL, USA (e-mail: jhill@southernco.com)
Jeff Munk is with National Renewable Energy Laboratory, CO, USA (e-mail: jeff.munk@nrel.gov).

II. Background of the Neighborhood

A. Hardware and price signal control approach

The data used for this research was collected in an occupied residential neighborhood in Atlanta, GA. The neighborhood consists of 46 townhomes, each equipped with API controlled devices that include a water heater, an HVAC system, a rooftop PV system, and two residential batteries. Identical water heaters were installed across all the homes in the neighborhood and they are capable of operating based on a heat pump, a resistance coil, or both. Each water heater is connected to a neighborhood model predictive controller which optimizes load of the water heater according to the objective function presented in Eq. 1.

$$min\left(\sum_{t=0} \rho_t^c \, P_t^{wh} + W_{mode} \sum_{t=0} Chg_t^{mode}\right) \qquad (1)$$

Where ρ_t^c is the electricity cost at time t, P_t^{wh} is the active power consumed at time t, which is calculated as a sum of heat pump power and the power of the heating element as shown in the equation below.

$$P_t^{wh} = On_t \, heat_{wh} + On_t^{element} \, heat_{element} \qquad (2)$$

$heat_{wh}$ is power consumed by the heat pump, On_t is a binary variable and takes the value of 1 if the heat pump is switched on and 0 if the heat pump is switched off. $heat_{element}$ is the power of the heating element, and On_t takes the value of 1 if the heating element is switched on. W_{mode} is weight associated with changing water heater mode, Chg_t^{mode} is a binary variable that takes the value of 1 when the water heater changes the mode. The optimization takes several operational and comfort constraints which are not reviewed here in detail. More information about the model and optimization of water heaters can be found in [9].

The important part of the optimization procedure is that behavior of water heaters is indirectly controlled through the price signal. The utility operating the neighborhood formulates a price signal based on its peak preferences and distributes it to all devices in the neighborhood.

From the price graph in Fig. 1, we can observe that the price is not the same throughout the day. There are areas of higher prices which are expected devices to shed load and lower prices which are expected to encourage preheating of water. For the Atlanta neighborhood, the utility has full visibility of the current operation of water heaters through API data and a submetering system. We use this information to inform our data analysis and validate the research results.

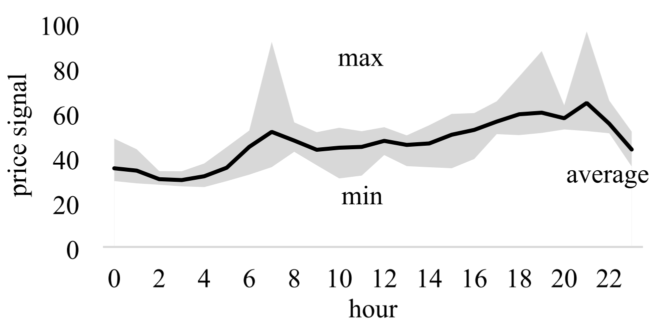

Fig. 1. Hourly price signal over the two weeks of control

B. *Experimental Setup and Data Collection*

We validate the proposed approach through the use of historical data from October-December 2022, excluding October 31. The data was collected for weekdays during the hours of optimization, from 5 am to 10 pm.

In order to address the questions discussed in Section I, we first try to establish the relationship between peak and device synchronization for a smaller period of time. To do so, we collected data on load during four weeks in September and October 2022. All water heaters in the testbed have the same technical characteristics. Therefore, the next step was to find weeks with comparable homeowner routines to develop the control and idle counterfactuals. We identified two weeks of idle water heater behavior in the second half of September as described in Table I.

TABLE I
CHARACTERISTICS OF WATER HEATER INSTANCES

INSTANCE	DATE	DESCRIPTION
1	2022/10/24-2022/11/04 5 am - 10 pm	Control period optimization hours (weekdays only)
2	2022/10/24-2022/11/05 10 pm - 12 am	Control period after the optimization (weekdays only)
3	2022/09/19-2022/09/30 5 am - 10 pm	Idle period optimization counterfactual (weekdays only)
4	2022/09/19-2022/10/01 10 pm - 12 am	Idle period after the optimization counterfactual (weekdays only)

Further, a series of changes in experiments did not allow us to use the period immediately following the idle weeks. However, we were able to find two consecutive weeks with no changes in the experiment or data collection procedures in October. While the idle week and control week are about a month apart, the average water heater load increased by only about 80 W between the idle and control periods. This allows us to use the selected weeks for further analysis. The selected idle and control periods were used to analyze the controllability of the fleet during morning, day, and evening hours. The findings are summarized in Section III.

C. *Methodology*

In this research, we adopted Ward clustering algorithm to examine the extent of synchronization among the water heaters across townhomes [10], [11]. The Ward algorithm initially assigns an individual position to every observation and then merges the closes clusters into higher tier clusters until all observations are merged into one group. The distance between two clusters is found according to Eq. 3.

$$d(u,v) = \sqrt{\frac{|v|+|s|}{T} d(v,s)^2 + \frac{|v|+|t|}{T} d(v,t)^2 - \frac{|v|}{T} d(s,t)^2} \qquad (3)$$

Where $d(u,v)$ is the distance between two clusters u and v. It is an iterative algorithm, for each iteration two clusters s and t are merged to form a cluster u each iteration. v is an unused cluster. The distance between two largest clusters shows the degree of homogeneity in data and is used to estimate the similarity behavior of water heaters.

The load data received from the vendor API for the time period mentioned in Table I is preprocessed through the following steps. The first step incudes resampling the water heaters data into 10-minute intervals. The original data is sampled every 5 minutes. If an observation is not logged at exactly 10-minute distance, the sampling takes the closest observation before the timestamp. The second step is to convert W values of instantaneous load into binary values (0 or 1). This allows to avoid one of the main shortcomings of the Ward method, namely its edge susceptance. The preprocessed data is later inputted to Ward's clustering method.

III. Results and Discussion

As a first step, we review the intraday load behavior and the extent of synchronization in the fleet. We compare the control and idle periods to see if the peak shifting efforts resulted in load change. The effect of the peak shifting is illustrated in Fig. 2. The control effort resulted in a shift of peak from peak evening hours to later evening. While the total average load increased between idle and control weeks, the comparison of demeaned data allows to evaluate the peak shaving result. During hours 18 and 19 which were main hours of interest, total load decreased by 0.4 and 0.8 kW.

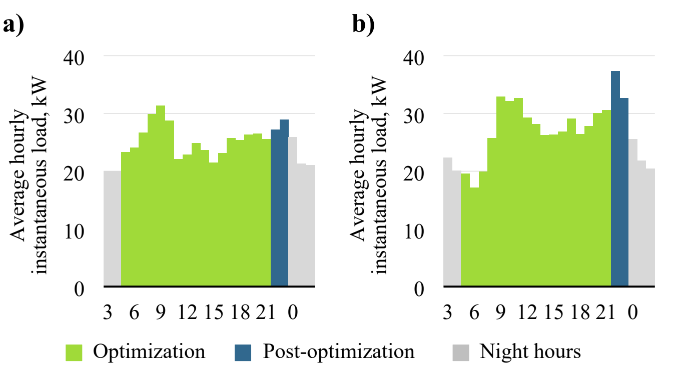

Fig. 2. Average hourly water heater load in the neighborhood during (a) idle period and (b) control period. To preserve the visibility of the evening peak, the data is presented from 3 am to 2 am rather than from hours 0 to 23.

Control effort affects the operating status of the devices. For instance, intensive water heater operation results in more devices being switched on. This, in turn, would result in the convergence of device operating patterns, defined as synchronization. This is confirmed by reviewing the extent of convergence of devices per hour of the day, shown in Fig. 3. Fig.3. shows that hours with lower cluster distance or higher convergence in water heater operation patterns correspond to the hours during or after the morning or evening peak. The intervals in the middle indicate the lower load periods of 11 am – 6 pm. Further, the convergence is observed during both idle and control weeks, indicating that some of the peak is probably natural.

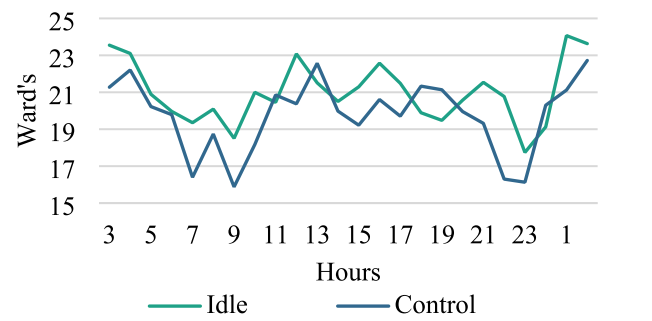

Fig. 3. Sum of distances to nearest cluster centers for the Idle and Control periods. To preserve the visibility of the evening peak, the data is presented from 3 am to 2 am rather than from hour 0 to 23.

The figure shows that hours with lower cluster distance or higher convergence in water heater operation patterns correspond to the hours during or after the morning or evening peak. The intervals in the middle indicate the lower load periods of 11 am – 6 pm. Further, the convergence is observed during both idle and control weeks, indicating that some of the peak is probably natural.

As a next step, we attempt to analyze the joint distribution of load peak and the extent of convergence of device operation. We use the entire control period from the late October until Christmas of 2022. There were 1436 hourly observations in the sample. This constitutes 92% of the total 1548 control hours between October 20 and December 22. The overall results are presented in Fig. 4.

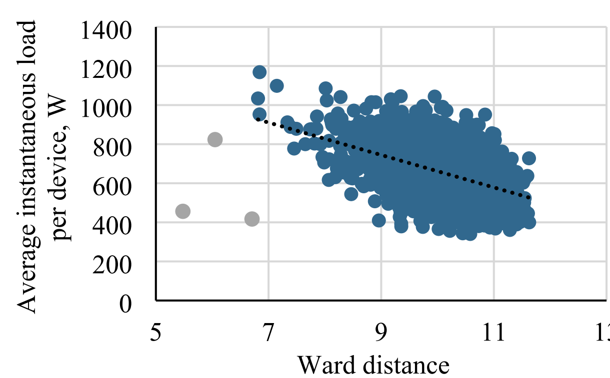

Fig. 4. Joint distribution of average load per device and Ward distance for control hours in October – December 2022.

The results cannot be used for least squares analysis because they do not have the independent identically distribute nature. But we can still conclude that there is a correlation between the two series of data, with the correlation of the main group of observations -0.466. The number drops to -0.445 when the three data points highlighted in gray are merged with the rest of the dataset. This supports the idea that the two variables may be directly dependent on one another or dependent on a third unobserved variable, which we defined as controllability.

The analysis of the results is made more difficult by the hardware. The water heaters can operate with different configurations. Depending on whether they use the heat pump, the resistance coil, or a combination of the two, their instantaneous load can be below 1 kW or above 5 kW. As a result, we expect the findings to be of approximate nature. Using water heaters with constant operating power could produce more accurate results. We design an artificial scenario when all nonzero load instances are set equal to 800 W. This would produce the results which are biased up on the horizontal axis because more hours are needed to reheat water in the 800 W mode compared to 5300 W mode. However, this reduces the variance along the vertical axis, as illustrated in Fig. 5.

The results were found to be comparable to those found in the analysis of actual observations. The correlation between the two series of data was found to be -0.449 for the main sample and -0.437 for the sample which includes the outliers highlighted in gray.

Further we investigated how the distribution changes between morning peak, day hours, and evening peak. The summary of hourly distributions is provided in Fig. 6.

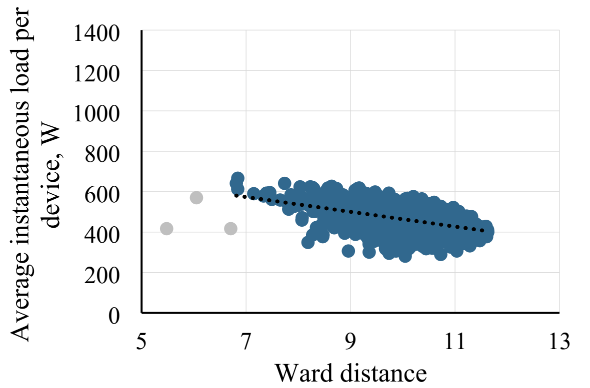

Fig. 5. Joint distribution of the average load per device calculated based on replacement values of 800 W and Ward distance for control hours in October – December 2022.

Intraday findings shows that there is strong and consistent correlation for late evening hours, indicating that the load and device controllability are indeed interrelated. The correlation between the two values for hour 22 is -0.628. The coefficient declines to -0.384 for hour 23, probably because some of the water heaters already stop reheating water and connect or disconnect more sporadically, causing an increase in Ward distance between devices.

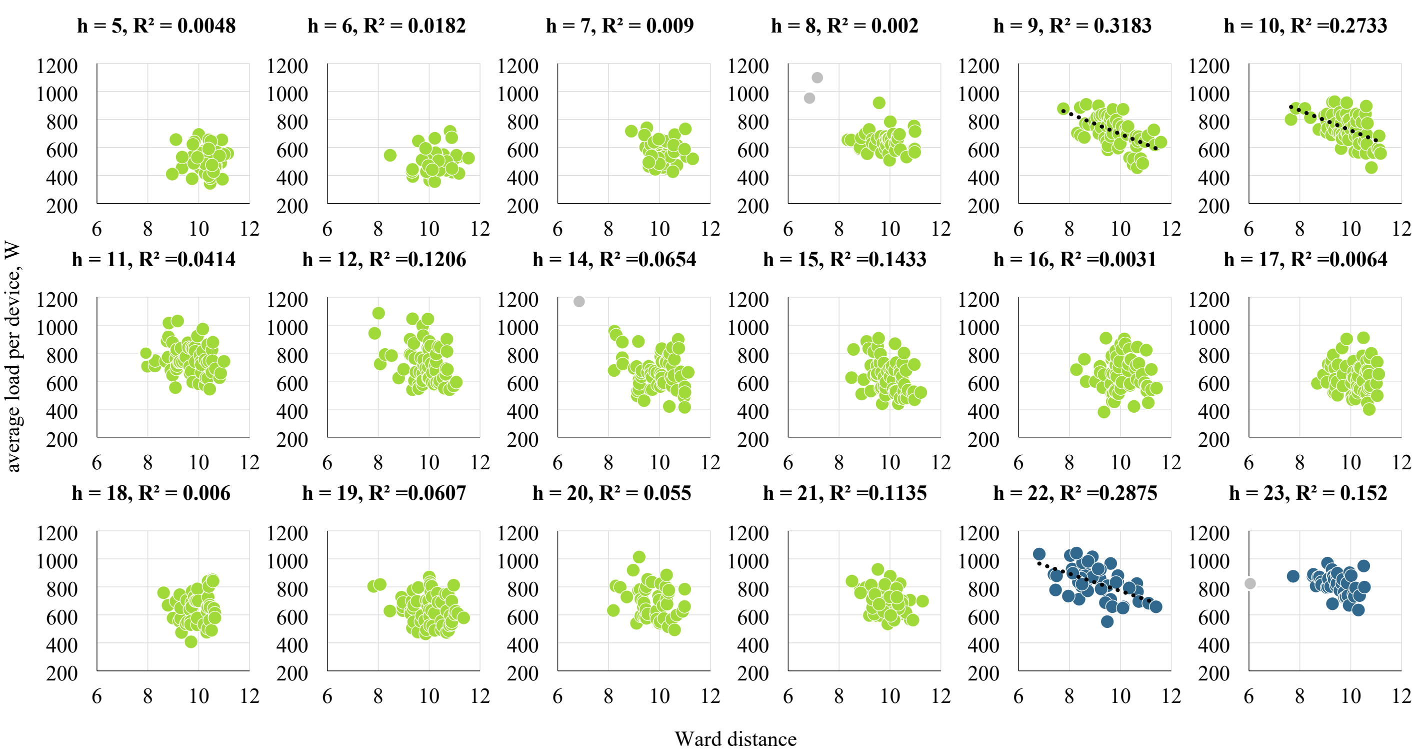

Fig. 6. Joint distribution of average load per device and Ward distance for the indicated control hours in October – December 2022 (note: hour 13 from 1 pm to 2 pm was skipped due to space constraints).

The correlation almost disappears during day hours, in line with the lower load control effort. The optimization pattern from 11 am until about 5 pm does not show any pronounced effort to shift load. The price signal levels are rather similar through those hours, which probably results in a variety of load values, as well as a varying operating patterns. We do expect more correlation between Ward distance and average load during the afternoon hours when price signal pressure grows. However, both the load and the operating patterns preserve the high variability already seen during early afternoon, probably as a results of varying homeowner behaviors.

Morning hours show a rather unexpected pattern. While price signal is designed to incentivize lower load during hour 7 with a rebound at hour 8, this is not reflected in the observations. By contrast, the effect is visible during the two subsequent hours. The price signal is increasing for hours 9, 10, 11. But the load behavior changes only for hours 9 and 10, which is not consistent with the overall trend observed for other hours. At this point we do not have a consistent explanation for such behavior. While it is possible that homeowner routines also affect the morning hours, the change in price signal on a narrow interval of hour 8 should be able to provide sufficient control incentives for all types of homeowner profiles.

IV. Conclusions

In this study, we attempted to find an approach which would help utilities to have a better understanding of the impact of the control decisions and price signal for a specific device, even if they have very little visibility into the load detail. In this study, we found that there is a positive correlation between the convergence in operating status of devices and the change in load volume. Further, we were able to find that this correlation is present for the evening peak, as expected. It is not present for the afternoon interval, which was not of interest to the utility, which was also in line with the expected operation. Contrary to our expectations, we do not find significant correlation for early morning hours. These hours were of interest to the utility, which resulted in more active change of price signal. However, the convergence only started being visible during later morning hours. This can be an evidence of role of water usage to the impact of control on fleet of water heaters.

More research is needed to understand why the morning peak does not behave in the same way as evening peak. Such research could potentially include a more detailed analysis of individual device or homeowner routines. It could also require a larger sample of devices or dedicated experimental verification with a more aggressive price signal approach. More research would also be necessary to statistically verify the proposed approach on a large sample of devices.

Acknowledgment

This material is based upon work supported by the U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy, Buildings Technologies Office, under contract number DE-AC05-00OR22725.

References

[1] D. Zhang, C. Li, S. Luo, D. Luo, M. Shahidehpour, C. Chen, and B. Zhou, “Multi-Objective Control of Residential HVAC Loads for Balancing the User’s Comfort with the Frequency Regulation Performance,” *IEEE Transactions on Smart Grid*, vol. 13, no. 5, pp. 3546-3557, 2022.

[2] S. D. Nazemi, M. A. Jafari, and E. Zaidan, “An incentive-based optimization approach for load scheduling problem in smart building communities,” *Buildings*, vol. 11, no. 6, 237, 2021.

[3] A. Belov, N. Meratnia, B. Jan van der Zwaag, and P. Havinga, “An efficient water flow control approach for water heaters in direct load control,” *Journal of Engineering & Applied Sciences*, vol. 9, no. 11, pp. 2106-2120, 2014.

[4] S. Widergren, C. Marinovici, T. Berliner, and A. Graves, “Real-time pricing demand response in operations,” in *2012 IEEE Power and Energy Society General Meeting*, pp. 1-5, 2012.

[5] Z. Csetvei, J. Østergaard, and P. Nyeng, “Controlling price-responsive heat pumps for overload elimination in distribution systems,” in *2011 2nd IEEE PES International Conference and Exhibition on Innovative Smart Grid Technologies*, pp. 1-8, 2011.

[6] L. Paull, H. Li, and L. Chang, “A novel domestic electric water heater model for a multi-objective demand side management program,” *Electric Power Systems Research*, vol. 80, no. 12, pp. 1446-1451, 2010.

[7] M. Amer, A. Naaman, N. K. M'Sirdi, and A. M. El-Zonkoly, “Smart home energy management systems survey,” in *International Conference on Renewable Energies for Developing Countries 2014*, pp. 167-173, 2014.

[8] A. Gharbi, M. Ayari, and A. E. Yahya, “Demand-Response Control in Smart Grids,” *Applied Sciences*, vol. 13, no. 4, pp. 2355, 2023.

[9] M. Starke, J. Munk, H. Zandi, T. Kuruganti, H. Buckberry, J. Hall, and J. Leverette, “Real-Time MPC for Residential Building Water Heater Systems to Support the Electric Grid,” 2020 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference, 2020, pp. 1-5.

[10] A. Javed, B. S. Lee, and D. M. Rizzo, “A benchmark study on time series clustering,” Machine Learning with Applications, vol. 1, 2020.

[11] https://docs.scipy.org/doc/scipy/reference/generated/scipy.cluster.hierarchy.linkage.html.