Real Time Monitoring and the Internet of Things: Reshaping How we Collect Data

Kevin Price, Evergreen Economics, Inc., Berkeley, CA
John Cornwell, Evergreen Economics, Inc., Portland, OR
Chad Fulton, Evergreen Economics, Inc., Portland, OR
David Thayer, Pacific Gas and Electric Company, San Francisco, CA

ABSTRACT

In the burgeoning Internet of Things, new end-use metering data collection methods are available that may revolutionize the way evaluators collect usage and behavior information. For lighting metering in particular, connected lamps may allow for a high degree of certainty regarding the on/off status and other characteristics of the lamp, such as light color or brightness, because the status data is continuously communicated to a centralized hub or controlling device. While there are concerns regarding product connectivity performance and network security, these issues are not insurmountable. This paper presents a proof of concept of the methodology and a brief overview of illustrative findings from a recently completed assessment of residential, wirelessly-controllable LEDs, as well as potential EM&V implications brought on by new, connected energy efficiency equipment.

Introduction

Evaluation, Measurement, and Verification (EM&V) frequently requires evaluators to monitor the equipment usage patterns and behaviors of end-users in order to understand the energy savings impacts of energy efficiency retrofits. Data collection typically follows a standard series of steps that include: identification of appropriate and suitable monitoring sites; acquisition and installation of monitoring equipment; monitoring period; equipment decommissioning; data download; data cleaning; and, finally, data analysis. These approaches can be very costly, labor intensive, and can create hassle for participants.

However, in the burgeoning Internet of Things, new end-use metering data collection methods are available that may revolutionize the way evaluators collect usage and behavior information. Appliances and equipment are becoming more readily available with connectivity features as well as internal monitoring functions. A variety of equipment types include features that allow users to monitor and control equipment remotely; with the appropriate permissions, evaluators can access these data without needing to purchase and install monitoring equipment. These technologies present new opportunities for deploying metering studies that are low-cost, high-volume and vastly improve the understanding of load shapes and usage patterns. Furthermore, as connected devices bring new ways of interacting with appliances, lighting, and other devices, it is important to develop understandings of how the added functionalities are used by end-users.

As with other modern metering devices that are typically capable of reporting collected data remotely to evaluators, connected equipment allows evaluators to react to data collection issues – such as malfunctioning meters or data outside of anticipated ranges – as they occur. In addition, connected device monitoring eliminates the need for site decommissioning visits altogether, as evaluators are able to download data during any point in the monitoring, and may be able to "disconnect" the monitoring remotely. This reduces the hassle for the end-user – potentially reducing nonparticipation bias, if security concerns are addressed in future studies, and requiring lower participation incentives – and thereby has the potential to reduce study costs. There are also significant financial and aesthetic benefits associated with connected equipment monitoring: no extra metering equipment to purchase (by the study sponsor or evaluators) and no external metering equipment in view of the end-user (especially important for certain technologies, such as lighting).

For lighting metering in particular, connected lamps offer a way to overcome some of the drawbacks of the currently available meters. Traditional lighting monitoring equipment relies on photocell readings to determine the on/off status of a lamp, where ambient light can create false "on" readings due to installer error. Connected lamps allow for a higher degree of certainty regarding the on/off status and other characteristics of the lamp, such as light color and brightness,[1] because the status data is continuously communicated to a centralized hub.

Despite the benefits of relying on Internet of Things devices to reveal the usage patterns of end-users identified in this paper, there are a number of concerns that should be addressed before scaling an effort to a larger population. Network security should not be compromised in an effort to obtain data, and there are unique challenges associated with the setup of the monitoring protocols we established for this study (and described in the remainder of the paper).

Background

Evergreen Economics researchers worked together with PG&E staff to conduct an emerging technologies assessment to better understand the energy savings potential of residential screw-based LED replacement lamps with embedded control features (Evergreen Economics, Inc. 2014). The study paid no attention to the primary driver of LED replacement lamp savings: the difference in wattage between the installed LEDs and the replaced lamps. Rather, PG&E's Emerging Technology Program staff wanted to investigate the potential for additional, behavior-driven savings (i.e., savings associated with how the end-user uses the lamps, not the difference in wattage between the existing lamps and the installed LEDs), and thus the study was tailored to develop an understanding of the behavioral changes associated with embedded controls, and to assess if these behavior changes amount to an additional source of energy savings (beyond the wattage reduction) in residential fixtures that currently lack sophisticated control features..

This paper turns the study on its head, focusing not on the results – how people used their connected LEDs – but how we, as evaluators, used the connected LEDs that were installed to acquire the data needed to assess their usage. The paper describes the specific method we developed and the lessons we learned implementing the procedures, including concerns regarding network security if implemented on a larger scale. Then it briefly discusses high-level implications that Internet of Things devices bring to the evaluation community. We believe that this alternative approach to data collection – once security concerns are overcome – can be applied to future studies as more energy efficiency devices become connected to home area networks as the Internet of Things moves from concept to fringe to ubiquitous.

The Studied Technology

The technology assessment was narrowly focused on one particular controls enabled LED system produced by Philips: the Hue. The Philips Hue "Starter Pack" consists of a set of three screw-in LED lamps that communicate wirelessly with an included "bridge."[2] The bridge is connected to the home's Internet router, connecting the entire system to the Internet. The lamps can be controlled by the normal fixture switch, an iOS or Android application, or via a password protected web portal. The fixture switch must be in the "on" position – with power reaching the lamp – in order to use the application or web portal controls.

---

[1] Brightness is used throughout the paper to describe the relative lumen output associated with the use of the dimming feature of the connected lamps.

[2] The bridge is the central hub for the Hue system and communicates with all connected lamps and the home's Internet router, which allows for connecting to the Internet and associated iOS and Android applications.

Embedded within the Philips Hue are the following key control variables each of which can be implemented remotely:

- On/off
- Dimming
- Color temperature (over 16 million color variations)
- Scheduling

In addition, the user can create and use additional software packages to control the Hue system based on any number of defined occurrences (e.g., flashing red LEDs to indicate a local severe weather alert, based on data from an Internet website such as the National Oceanic and Atmospheric Administration).

The Hue Starter Pack used for this study includes A19-shaped LEDs (additional LEDs can be purchased and up to 50 total can be connected to a single bridge). These LEDs are rated at 9W, and produce a variable maximum lumen output depending on the color temperature (360 lumens at 2000K, 510 lumens at 3000K, 550 lumens at 6500K, 600 lumens at 4000K).[3] They are incompatible with existing dimmer switches, but can be dimmed via the application. They are capable of producing approximately 16 million color variations and many shades of white light. Currently there are more than a dozen similar products available on the market.

Technology Assessment – Instrumentation and Metering

Evergreen Economics researchers developed a novel approach to conducting this assessment, which included a field placement of LEDs with embedded controls. The field placement took place in the homes of Evergreen Economics and PG&E staff, and intended to provide preliminary data on a small sample of sites and lamps that may be used to understand the potential for savings and changes in behavior as a result of measure installation. The sample was developed based on the need for both information-rich cases and ease of implementation, and likely does not reflect how the general population might use this technology. **Error! Reference source not found.** shows the characteristics of the study homes and occupants.

Table 1: Study Population Characteristics

Site	Home	Occupants
1	Three bedroom single-family detached	Two adults (20s), zero children
2	Three bedroom single-family detached	Two adults (40s), two children (elementary)
3	One bedroom apartment in multifamily building	Two adults (20s), zero children
4	Four bedroom single-family detached	Two adults (50s), two children (teenagers)
5	Two bedroom apartment in multifamily building	Three adults (20s), zero children
6	Three bedroom single-family attached	Two adults (30s), two children (elementary)
7	Two bedroom single-family detached	Two adults (30s), zero children
8	Four bedroom single-family detached	Two adults (40s), two children (elementary)

The metering strategy employed for this effort involved a custom software application that communicated with a user's Hue bridge to download the current setting of each controlled variable for all

[3] For additional specifications please visit: http://www2.meethue.com/en-us/the-range/hue/

connected Hue lamps. These data were collected at 15 second intervals for the duration of the study. The field placement included eight homes and lasted a total of 15 weeks.

Researchers wrote code to connect the bridge's data and a snapshot database (the process is shown visually as part of Figure 1, which is discussed in greater detail below), and these data were extracted and appended to the database in real time. One key element to the development of the database format was understanding the Philips Hue's application programming interface (API), which describes the extent and organization of the data provided by the bridge. The database tables were organized in a relational fashion, rather than as single "flat" table, so that the database could be easily queried at any level, including by study home, date, or specific lamp.

Phased Monitoring

The field placement involved two phases: the first phase mimicked baseline conditions while the second phase tested for changes in lighting usage behavior based on the capabilities of the control enabled LEDs.

The purpose of the first phase was to estimate how lamps are used during baseline conditions in rooms where the households plans to install the controls embedded LED replacement lamps. During the first phase of the placement, study participants were required to emulate a pre-Hue situation (i.e., before the controls embedded LED replacement lamps are installed). Residents of study homes were only permitted to use the pre-existing on/off switches to control the Hue lamps (we could not disable the Hue's features; the households agreed to abstain from using the features as part of the study[4]). The lamps were connected to the bridge, and thus Evergreen was receiving near real-time usage readings. This allowed for the collection of usage data associated with the fixtures under pre-Hue conditions.

During the second phase, participants were permitted full control of the Hue lamps using the included software and standard on/off switches for the fixtures. The purpose of the second phase of the study was to assess if and how lamps are used after the household is able to use the features of the control enabled LED replacement lamps.

Monitoring Setup and Specification

With an understanding of the Hue's API – and through trial and error at a test site – Evergreen staff developed process for monitoring equipment and network configuration setup involving installing the Hue system and remotely connecting to the home's Local Area Network (LAN) via port forwarding. This process is integral to the monitoring specification for the Philips Hue, but could likely be adapted for similar connected devices. There are concerns regarding security and privacy if implemented on a larger scale (these issues may be resolvable, but further investigation is required), which we discuss subsequently.

Once connectivity was established, the installer created a user account on the Hue Bridge for the data collection server. To create the account, the installer pressed a "link" button on the Hue Bridge itself (typically used to connect additional lamps to the system). At this point the monitoring setup was complete and the custom software would begin collecting data at 15-second intervals for all connected lamps.

Network Security. However unlikely, data from the Hue could be used maliciously against a household, and therefore network security is important to address. For this study, researchers maintained tight control over the sample design and limited the number of homes. Should the need for a more robust study arise (involving hundreds or thousands of homes), strict network security policies should likely be employed.

[4] Since we were able to view data in near real-time, we were able to discern during the first phase whether participants were adhering to these requirements. No participants used the features of the Hue during the first phase.

That said, vulnerabilities in the Hue system exist and privacy can be compromised given the right set of circumstances (Notra et al. 2014). First and foremost, a hacker would need to gain access to the home's network in order to implement any malicious plan involving the Hue, at which point the Hue might not be the most important concern for the household. Simple, standard home network security processes exist (i.e., using a network password) to practically ensure that hackers are unable to gain access.

Our metering protocol provided an additional, very limited opportunity for unpermitted access to the Hue system only (it would not be possible for hackers to exploit our monitoring protocol for any other purpose on the home network). We weighed the associated risk and determined that the limited scope of the installation, short timeframe, and very small risk of a hacker knowing that a particular Hue system at a particular home had a very minor vulnerability did not warrant additional security investments.

If scaled for a larger study, a number of options exist to increase the security of monitoring protocols. We considered the following two options to ensure secure data transfer: 1) creating a virtual private network (VPN) to encrypt all transmitted Hue data to ensure that, if intercepted by a malicious entity, the data and associated usernames would be unintelligible, and 2) installing a computer onsite (e.g., a small, inexpensive Raspberry Pi[5]) that would collect and aggregate data before sending the data to our external servers. The second approach would treat the data as any outgoing, encrypted data transfer, as opposed to our monitoring protocol that required permitting access into the network (via port forwarding). We believe that either of these options would sufficiently secure the data transfer, thus ensuring that the monitoring setup does not add any undue risk beyond the minor risk associated with the Hue (or other Internet of Things systems).

Data Collection, QA/QC, and Data Cleaning

The flowchart in Figure 1, below, illustrates the data collection and data cleaning process implemented as part of this technology assessment. Typical lighting monitoring studies require extensive quality control and data cleaning, and this study was no different. Data including on/off status, brightness and color profile were collected from the bridge for each lamp at 15-second intervals and stored in an interval snapshot database as a time stamped data point. At regular intervals, the snapshot database was subjected to a de-duplication processes[6] that removed all successive, redundant observations for each lamp, resulting in an event database containing time periods of common usage[7] with start and end date and time stamps. At the end of the study period, the event database went through a post-processing regime that created the "raw" analysis dataset.

The "raw" dataset was subjected to an automated cleaning script as well as manual cleaning and quality assurance analysis to identify errors or problematic readings. Primarily, the analysis investigated the frequency of on/off events per lamp per day and the duration of each on/off event to identify periods with usage patterns outside what could be considered normal usage. The analysis revealed that some lamps in the study experienced periods with high numbers of alternating on/off events of very short duration. From manual inspection of some of these periods it was determined that the likely cause of many of these events were intermittent connectivity issues during which the bridge lost communication with the lamp, resulting in false "off" reading, i.e. the bridge recorded the lamp as off when it was in reality on. .

Due to the method we employed, records associated with lost connectivity were indistinguishable from true off events with 100 percent confidence because the data signatures of these events were the same as those for events were a light was turned off at the switch. Connectivity issues were encountered during the Hue setup due to long distances between a lamp and the hub, but when those situations arose we selected a different location for the lamp (and tested to ensure connectivity). The issues we encountered during the study were for relatively short periods of time, and relatively random (there was no discernable pattern or

[5] See https://www.raspberrypi.org/ for details regarding Raspberry Pi.
[6] The readings were considered duplicative if they were successive and all monitored variables other than time were identical.
[7] "Common usage" refers to periods of time with identical data (across all variables, including on/off, brightness, color, etc.).

lamps more or less prone to connectivity issues). We did not consider altering the components contained within the Hue lamps or Bridge to reduce connectivity losses.

Evergreen staff conducted sensitivity analysis to investigate the impact on overall lamp usage when these events of differing lengths were recoded from off to on. This analysis led the study team to decide that short off events of 90 seconds or less between two on events were more likely to be connectivity issues rather than true on/off events (considering “normal” household lamp usage). The study team was less confident that events of greater than 90 seconds were not due to true on/off events. A decision was reached to create an automated cleaning protocol and script to recode from off to on, all off events of 90 seconds or less that were bounded by on events. This automated cleaning did not have a large impact on overall lamp usage over the period. Once this cleaning protocol was completed, further analysis revealed some lamps in the study that continued to have days with large numbers of on/off events. Days that were outside of what the study team deemed “normal” usage were flagged, manually inspected and removed from the analysis if deemed problematic. Seven days out of a total of approximately 4,400 lamp days were identified as abnormal and were removed.

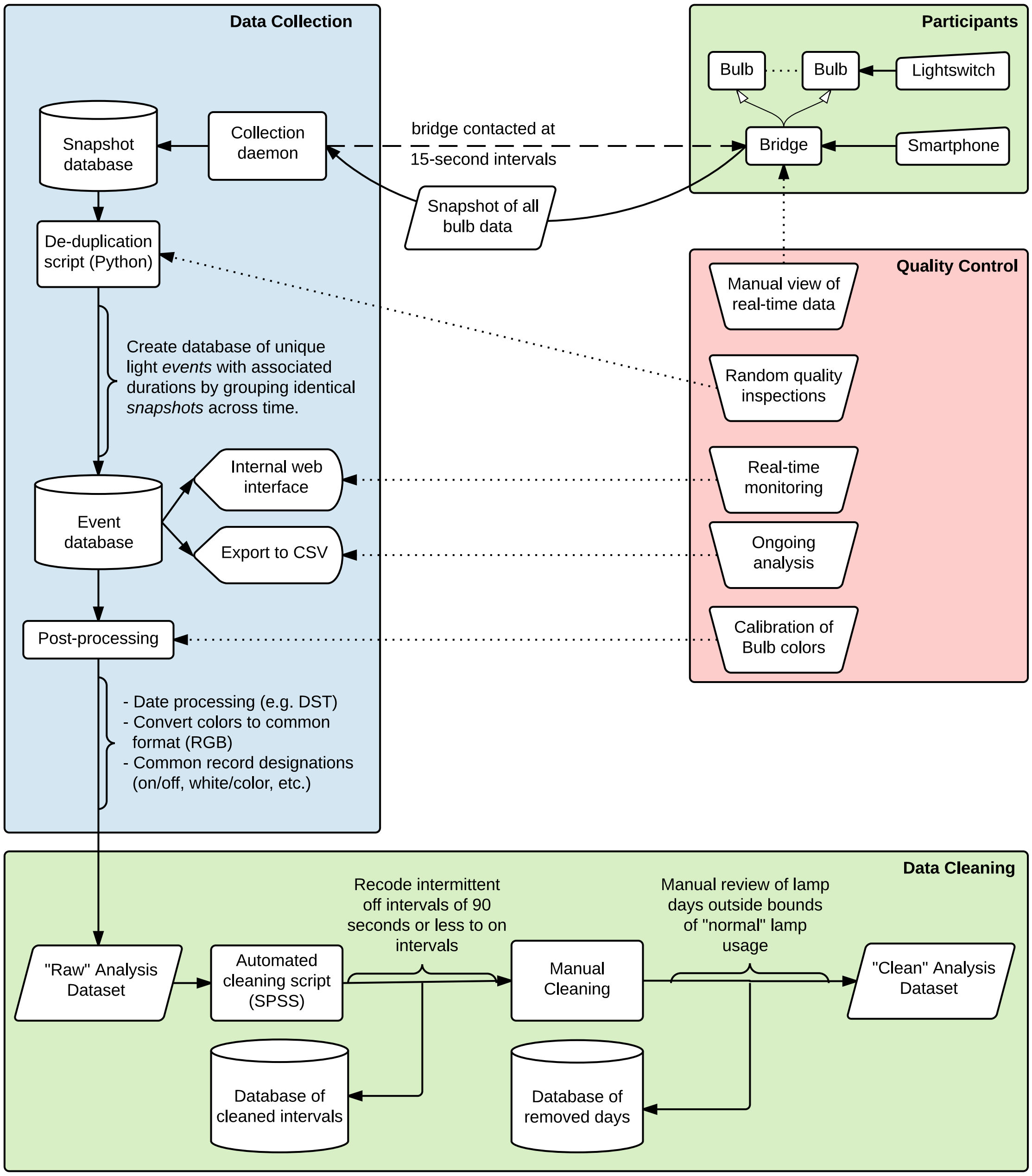

Figure 1: Data Collection and Cleaning Flowchart

Data Analysis

In this section of the paper we briefly cover the steps required to estimate the usage patterns from the previously described data.

Data Analysis – Lamp Usage and Brightness Profiles

The final cleaned dataset was analyzed using IBM SPSS and Microsoft Excel. Figure 2, below, provides an illustration of the analysis process. As the figure shows, the clean de-duplicated analysis dataset records were subjected to a database restructuring script that rounded all interval data to the nearest 15-second interval[8], and the dataset was backfilled with equal 15-second interval data. Next, we employed a database aggregation script to aggregate 15-second interval data into 15-minute periods, calculating the following metrics for each lamp in the process:[9]

- Mean percent of each 15-minute period that each lamp was "on" (versus "off") for the entire study period
- Mean percent of each 15-minute period that each lamp was "on" (versus "off") for Phase 1 and Phase 2.
- Mean percent of full lamp brightness, when "on", over the entire study period
- Mean percent of full lamp brightness, when "on", for Phase 1 and Phase 2.

Data were used to create charts to visually display the average 15-minute lamp usage profile and brightness level for each lamp. Lastly, average hours of use per day were calculated for each lamp and for each analysis category, e.g.: site, room type and overall.

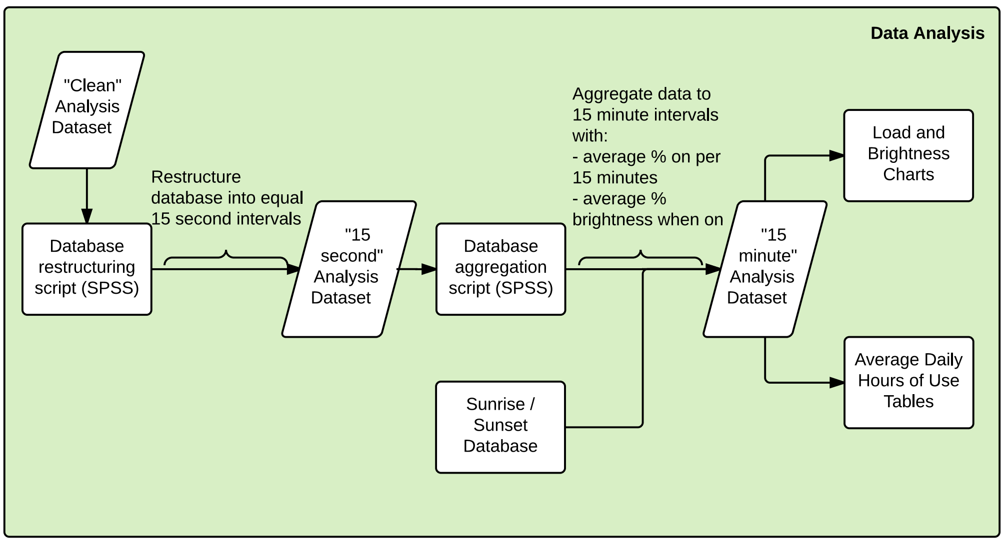

Figure 2: Lamp Usage and Brightness Data Analysis Flowchart

Data Analysis – Lamp Color Profiles

To summarize the colors used by participants, the actual colors produced by the bulbs were grouped according to the primary (red, green, and blue) and secondary (yellow, fuchsia, and cyan) colors, called reference colors. This approach was used because the full range of possible values for the hue of a given color of light range between 0 and 65,535, and the hue values for the primary and secondary colors are spaced relatively evenly across the range of possible hues. Thus, considering the distribution of colors used in terms of the reference colors provides most of the interesting data while maintaining simplicity. For each color a bulb displayed, the closest reference color was selected as the one whose hue value was the minimum distance from the recorded hue value. The color bins and associated hues are shown below, in Table 2.

---

[8] This process of rounding to the nearest 15-seconds was required to normalize the dataset, as very minor network delays resulted in some 15-second periods being slightly longer than 15 seconds.

[9] As noted in the graphic, we also created a sunrise and sunset database, but for the purpose of this paper we do not discuss this database, nor show results related to seasonality, etc.

Table 2: Non-White Color Bins (based on Hue)

Color Bin	Hue Start	Hue End
Red	0	9000
Yellow	9001	21750
Green	21751	30853
Cyan	30854	41563
Blue	41564	51510
Fuchsia	51511	60817

Although in principle the color grouping procedure could result in colors in any group (red, yellow, green, cyan, blue, or fuchsia), in practice no recorded hue values fell into either the "Green" group or the "Cyan" group. The Philips Hue hardware was designed to represent colored light, but in particular to represent whites and colors on the warm end of the spectrum. For this reason, the Hue bulbs can capture a wider range, and finer detail, in reds and yellows than they can in greens and blues.

White light was similarly binned into reference colors, but at a much finer level of detail. The reason for this is twofold; first, the Hue bulbs themselves were designed to represent more subtle variations of white light than colored light (since white light is used so much more frequently), and second because the variation in white light can essentially be represented across a single spectrum which describes white light in terms of its "color temperature". Warm light is associated with a yellow or red quality, whereas cold light is associated with blues. As shown in Table 3, white light was grouped into warmer, warm, neutral, cold, and colder. The default white light produced by the Hue bulbs is on the warm side of the spectrum, and so it is the reference color for the "warm" group. "Neutral" is the white in the middle of the color temperature spectrum, and it has the least "color", either warm or cold.

Table 3: White Color Temperature Bins (based on Mireks)

Color Temperature Bin	Mirek Start	Mirek End
Warmer	153	211
Warm	212	297
Neutral	298	347
Cold	348	434
Colder	435	500

Illustrative Findings

The following figures present the lamp use and brightness profiles, as well as the color usage profiles.

Figure 3 shows the mean percent on during both phases for all installed connected LED lamps, averaged for every fifteen minute period of the study, as well the mean brightness (when on) during the second phase (when the lamps could be dimmed). As shown, the study produced very reasonable estimations of lighting usage that was generally corroborated through discussions with site contacts and typical usage patterns associated with specific room types. Usage increased during the second phase, when the end users were able to control the test lamps (although seasonality certainly played some role[10]). However, the data also showed that households were using the dimming features of their lamps as well.

[10] The first phase went from August 13, 2014 to September 30, 2014, and the second phase went from October 1, 2014 to

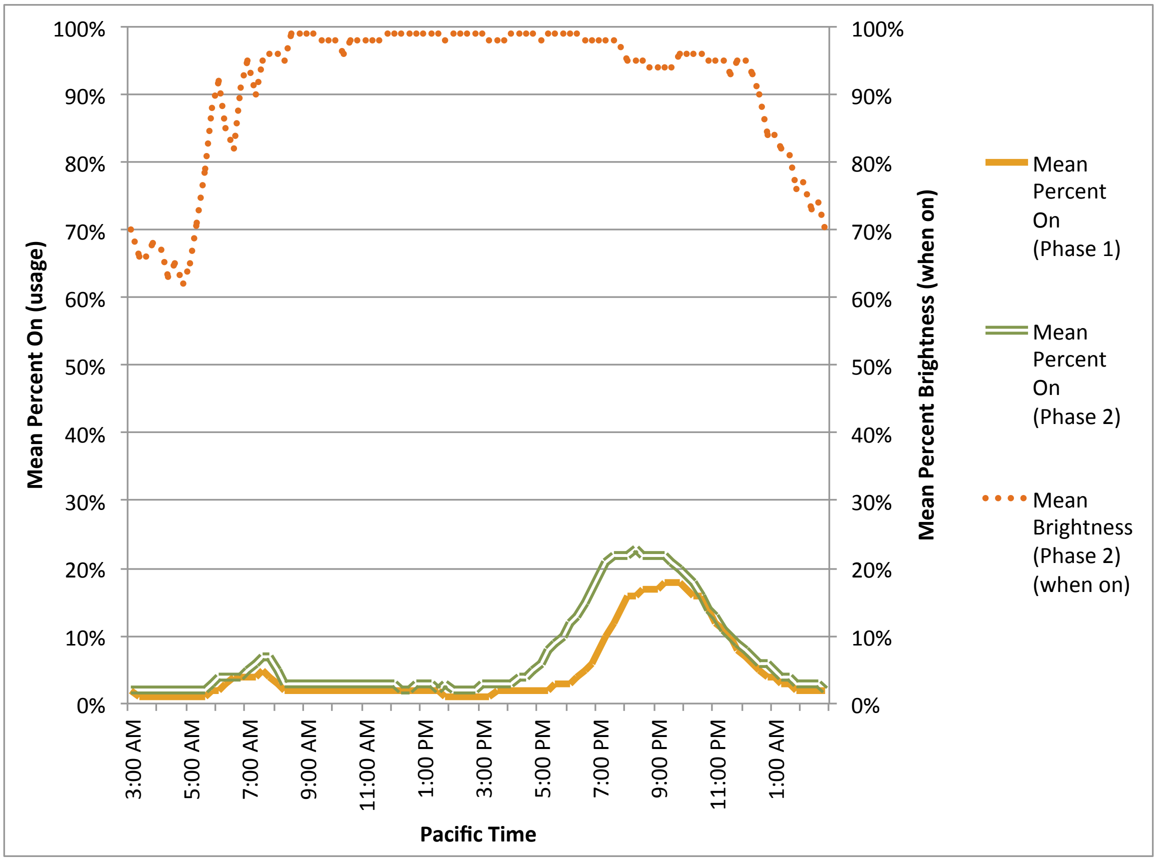

Figure 3: All Sites (Aggregate) – Lamp Use and Dimming Profile

Figure 4, below, shows morning and evening color characteristics of all lamps (based on the percent of on time at each different color bin). As shown, colors other than whites were infrequently used in the mornings (3.7% of on time) and evenings (6.7% of on time), but there is significant use of different white temperatures.

While the color aspect (blues, reds, greens, etc.) is very much a novelty component of the product (Evergreen Economics, 2014), the color temperature tuning is an important feature of connected LEDs as it allows customization within the white light spectrum. One study participant created a morning alarm clock with their lamps, where their bedroom lamps would slowly increase in brightness and change colors from a reddish hue to a warm white color; evaluators were able to distinguish this from the data, but would not have been able to do so with currently available monitoring equipment (at least not easily). This is the first study, to our knowledge, that has investigated how residential customers would use real-time color tuning of LED lamps installed in their homes.[11] With new connected products there may be other features that conventional metering equipment doesn't address, or doesn't address well. This analysis demonstrates that data driven connected devices can be leveraged to reveal usage patterns and associated behaviors of the end users controlling the devices or appliances.

December 3, 2014.

[11] This small study involved a sample of convenience to assess the potential for establishing energy savings estimates from changes in behavior and usage due to the additional control features of connected LEDs, and was not intended to be representative of the population of residential customers.

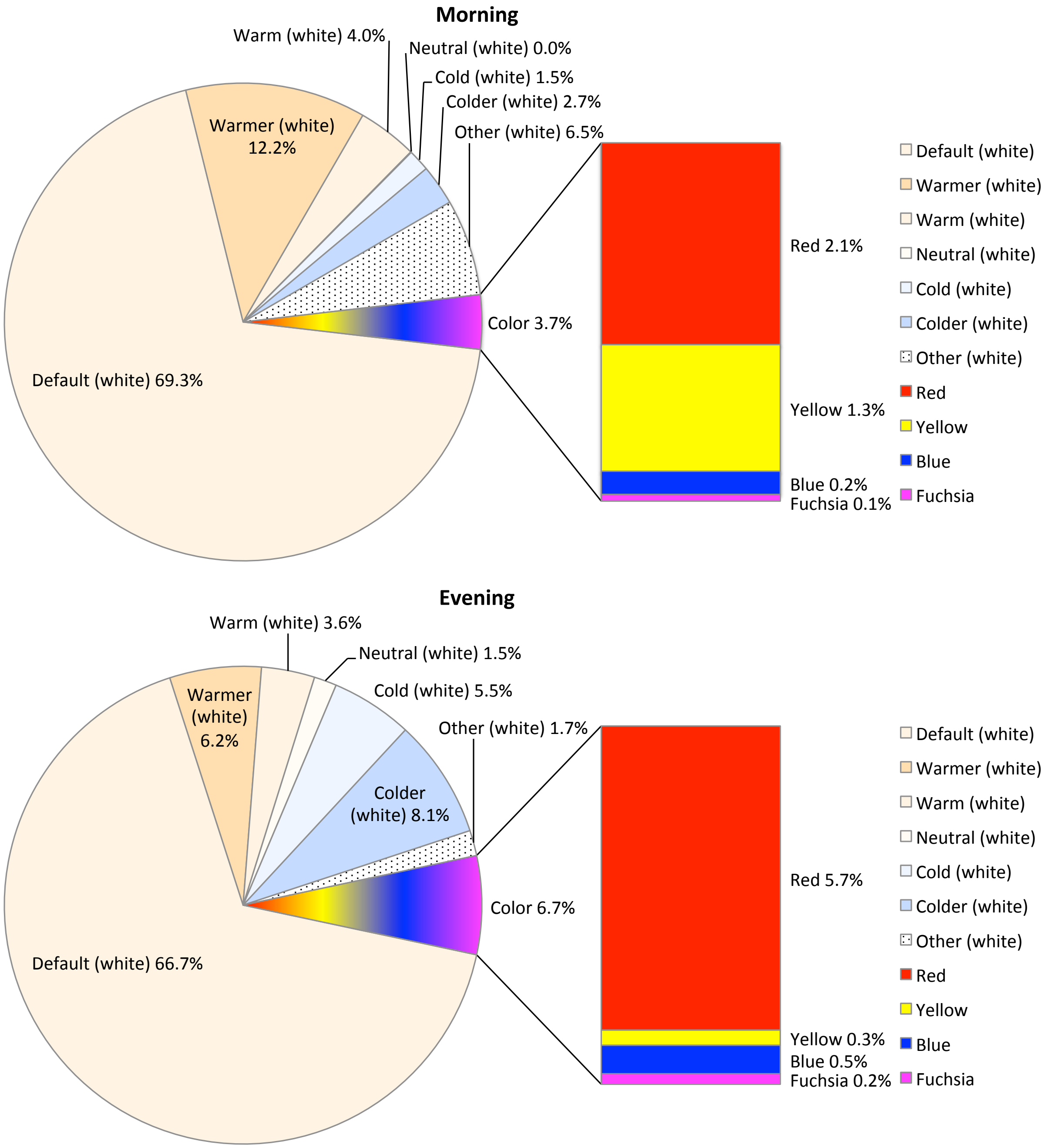

Figure 4: All Sites (Aggregate) – Color Usage

Implications and Conclusions

This study has proven that it is feasible for evaluators to leverage existing characteristics of connected equipment to better understand usage patterns among end users. There will be many opportunities to explore similar approaches as more technologies become integrated into the Internet of Things, but there will be many challenges ensuring that the resulting data are robust and accurate.

Connectivity issues are a real concern for implementing remote monitoring of equipment, especially since a monitoring protocol like the one employed for this study is subject to the studied products' connectivity performance. The connectivity issues we faced were all related to the performance of the Hue product itself, and not the monitoring strategy we employed. It is important to note that evaluators employing a similar strategy are likely constrained in similar ways, and since the products are not necessarily developed with this purpose in mind (remote monitoring for the development of usage datasets), it is critical to assess each product's connectivity performance. Ultimately, some data loss is typical of all metering studies and evaluators must balance budget constraints with the diminishing returns of extensive monitoring protocol or device testing.

However, existing metering techniques and equipment might not be designed to capture features of connected devices – such as the ability of users in the Hue study to control the color of the lamps or create personalized applications – but since the connected devices typically rely on interceptable or downloadable data it may be possible to use the devices themselves to assess new characteristics as well as standard features of the product class. With these interceptable data come network and data security risks whose challenges must be met by evaluators, but there are existing tools and strategies outlined in this paper that provide a roadmap for larger scale Internet of Things device monitoring studies.

One of the key benefits of this type of approach is cost savings and convenience. For the referenced study, there was no additional hardware purchased beyond the Philips Hue systems themselves, and no physical metering equipment to leave behind and to potentially bother the occupants of the home – existing metering equipment is often aesthetically displeasing which can be problematic in certain end use applications where physical appearance can be important to the homeowner (e.g., for lighting, but not for water heaters). Furthermore, since no metering equipment is left behind at the household, this type of strategy does not require a decommissioning site visit. These conveniences do come with a cost: the household must allow certain network permissions to the evaluators, which may not be acceptable to some households.

In conclusion, the method presented in this paper has potential limitations associated with the inflexibility of using existing equipment to effectively monitor itself. However, our study methods were ultimately successful and this paper provides a low cost framework for evaluators seeking to assess consumer behavior related to new, connected equipment, especially as new connected devices bring new untested features to market. We hope that the challenges and associated limitations of this approach will be overcome through testing and both gaining and sharing experiences conducting similar assessments as evaluators build upon and refine the methods shown in this paper.

References

Evergreen Economics, Inc., 2014. *Controls Embedded LED Replacement Lamps Technology Assessment*. Pacific Gas and Electric Company, Emerging Technologies Program. Project Number: ET13PGE1381.

Notra, S., M. Siddiqi, H.H. Gharakheili, V. Sivaraman, and R. Boreli. 2014. "An experimental study of security and privacy risks with emerging household appliances." *2014 IEEE Conference on Communications and Network Security (CNS)*. San Francisco, Calif. Oct. 2014.