Granger Causality Analysis on IP Traffic and Circuit-Level Energy Monitoring

Younghun Kim, Rahul Balani, Han Zhao, Mani B. Srivastava
Networked and Embedded Systems Lab.
Electrical Engineering Department
University of California, Los Angeles
{kimyh, rahulb, zdhzh, mbs} @ucla.edu

Abstract

Device-level energy monitoring has been increasingly proposed to understand inefficient energy use and design systematic processes for efficient building operation. Its sole use, however, is not sufficient to provide *actionable* information unless we understand the causes and context of energy use.

Fundamentally, energy consumption in a building is due to occupants' various activities. Understanding the causal relationship between occupants and their energy use is thus the key to an efficient building operation. This usually involves fine-grained sensing through intensive instrumentation of individual power outlets and/or extensive user studies that either increase the system cost or become too intrusive.

Instead, we advocate that circuit branch level energy monitoring combined with statistical Granger causality analysis is adequate to automatically understand the causal relationship. We monitor energy consumption of various zones in an office using a circuit level power monitor. IP traffic from users' PCs, obtained from a local firewall, is used to relate occupants with their energy use in each micro zone. The output is expressed in the form of causality graphs that illustrate how each individual influences energy use in different zones. We discuss the effectiveness and limitations of this causal analysis in capturing energy use patterns of the occupants in a lab environment.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous

General Terms

Experimentation, Human Factors, Measurement

Keywords

Causality, Network Traffic Analysis, Energy Consumption Profile

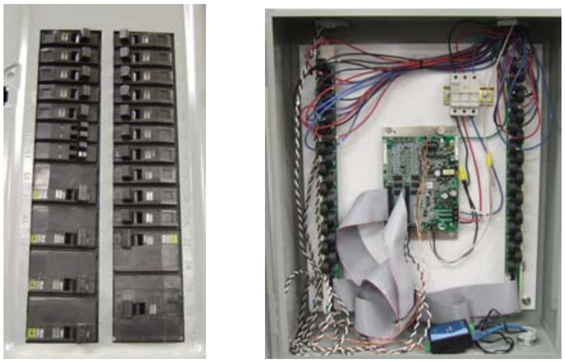
BuildSys 2010 November 2, 2010, Zurich, Switzerland.

1 Introduction

The ultimate goal of real-time fine-grained energy monitoring is to identify inefficient energy use and propose mechanisms or operational guidelines to alleviate it [11, 3, 14, 13]. Understanding the context of energy utilization, in conjunction with fine grained measurements, provides an insight into the solution of inefficient energy use. For example, it has been found that computer servers consume energy regardless of computing demands as they are always on. A simple yet effective fix is to allow the servers to sleep whenever they are not needed [16, 2]. However, with a majority of electrical appliances or devices, humans need to be involved in the energy conservation process, to a large extent, because they actually control and use the devices. Therefore, to motivate each individual, it is first critical to determine who *causes* energy consumption via different appliances.

In this paper, we consider the problem of associating office workers with energy use in their office. Unlike a residential space, office workers are much less concerned with their own energy use mainly for three reasons: (1) They do not pay the office electricity bill, (2) there's no visibility of energy use by individuals, and (3) office workers assume that they consume only a small portion of the total energy use. The total consumption information thus does not provide any stimulus (the tragedy of commons). To motivate the office workers, one can install outlet (device) level energy monitoring devices at each worker's desk [11], and the energy consumption can be further allocated with proximity-based RFID techniques [12, 10]. However, this heavily instrumented environment may increase maintenance and sensing system cost [15]. Moreover, energy use in shared appliances such as a water dispenser, or a soldering iron on a work bench, cannot be simply associated with occupants unless a proper apportionment rule is applied [10, 12].

Instead, we use a circuit level power monitoring device to monitor energy consumption in micro zones in an office (See Figure 1 and 2 for a typical electric circuit breaker and a power drain map), and associate energy use by each circuit number with each occupant. Although the circuit level power monitoring does not give an appliance level energy consumption, it gives a desk level (or micro zone level) energy consumption with a single point sensing. We use IP network traffic as a proxy for the presence or absence of an individual on his/her desk, which can be obtained easily from the existing IP infrastructure. The time series Granger causality



(a) Circuit Breaker (b) Veris E-30

Figure 1: (a) A circuit breaker for an approximately 860 sq. feet office space. (b) Each brach is monitored by an industrial grade circuit level power monitor, Veris E-30

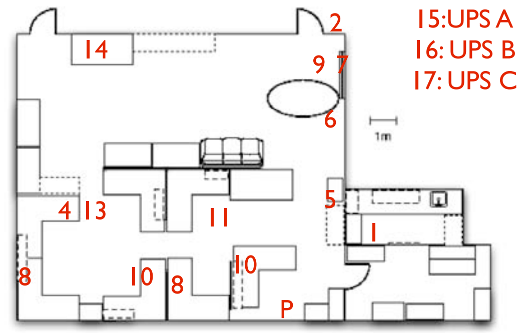


Figure 2: Power Drain Map: This figure illustrates the micro-zones corresponding to each circuit branch. The number on the floor plan indicates the branch number and 'P' shows where the power panel monitor is installed. Note that a few numbers are shown twice as electric wires are shared by a few desks. Several numbers are not shown in this figure as they go to different rooms.

(G-causality) [9] is used to measure causality between the IP traffic and energy consumption by various zones. The goal of this study is to show that this mathematical tool can be used to capture each occupant's causal relationship with energy consumption in different zones. In practice, G-causality can be used to automatically generate dynamic energy use map by individuals that relates one's identity with specific energy use (either positive or negative). Using this association, we allow a user to select energy consumption data that pertains to his/her usage, thus motivating him/her to take *actions* that enable energy efficient operation. Furthermore, we observe that the analysis captures interesting coincidence such as identifying shared work spaces and occupants interaction.

To the best of our knowledge, few attempts have been made to automatically relate an occupant's activities with energy usage in a building. Prior approaches involve intensive manual system training phase and/or require additional sensing infrastructure [12, 10, 4, 7, 5]. We believe our approach is the first attempt to automatically select a set of causes and effects on energy use. We use a simple but intuitive definition of causality: "The cause occurs before the effect, and the cause helps predict the effect." We do not intend to evaluate whether or not the causality is accurate. Rather, we discuss how the G-causality analysis captures one's energy usage patterns in terms of space and time.

2 Time Series Granger Causality

The Granger causality is a statistical concept of causality [9]. Defining causality in a mathematical formulation has been a fundamental challenge for philosophers for ages. It is a complex question with many different answers that have not been able to satisfy everyone. In 1960's, Clive W. J. Granger proposed a statistical concept of causality to understand a pair of related stochastic processes. The basic definition of Granger causality is intuitive. Suppose we have two stochastic variables X_t and Y_t. We first try to forecast X_{t+1} using only prior X_ts. Then we predict X_{t+1} using prior X_ts and Y_ts. If the latter prediction is more successful, with an appropriate metric, we can say that the Y_ts contain some information that helps predict X_{t+1}. This definition is based on the idea that "the cause occurs before the effect", which is our basic understanding of causality [9].

This concept was originally used in the context of econometrics where researchers study causal relation between economic variables such as oil price, stock market price, and so forth. Recently, G-causality was successfully adapted by researchers in neuroscience. For example, Seth et al. used this concept to simulate neural systems to understand the relationship between neuroanatomy, network dynamics, and behavior [17]. Here, they applied G-causality to understand the causal relationship among electroencephalography (EEG) signals. In addition, G-causality was also used to study the utility of magnetoencephalography (MEG) signals in detecting deceptive responses [18]. Compared to the simple lead/lagged Pearson correlation, Granger causality is generally more reliable.

Roughly speaking, while our case is in a slightly different context, electrical energy use and IP traffic in an office have a similar causal property. For instance, in most cases, IP traffic from a person's PC indicates his/her presence at the desk. In this situation, it's highly likely that he/she will be using more than a couple of appliances. Thus, IP traffic from a desktop indirectly implies energy consumption, to some extent, from non-computing appliances at the desk as well. In another scenario, temporary absence of IP traffic and simultaneous spike in power consumption from another micro-zone indicates that the person left his/her desk to work on some other appliance in a different zone. We discuss these scenarios in detail in section 4.

Now, we introduce the formal definition of time series G-causality. Let X_t and Y_t be two time series data (in our case, X_t: IP network traffic and Y_t: power consumption from a circuit branch). The Granger causal model is defined by

$$\begin{aligned} X_t &= \textstyle\sum_{j=1}^{m} a_j X_{t-j} + \sum_{j=1}^{m} b_j Y_{t-j} + \varepsilon_t \\ Y_t &= \textstyle\sum_{j=1}^{m} c_j X_{t-j} + \sum_{j=1}^{m} d_j Y_{t-j} + \nu_t \end{aligned} \qquad (1)$$

where ε_t and ν_t are two uncorrelated white-noise series, and m is the number of past data points that are used to predict the current data point at time t. Ideally m can be equal to infinity, but in practice, due to the finite length of available

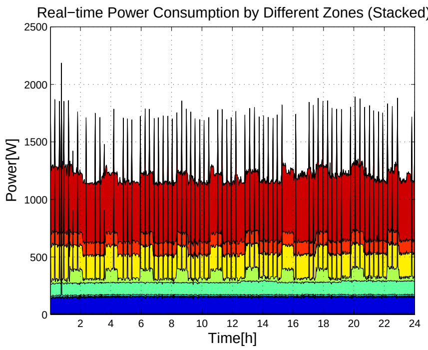


Figure 3: Real-time Power Consumption: Detailed power consumption profile can be easily obtained from the monitoring set-up. However, it is extremely difficult for general users to get a deeper insight.

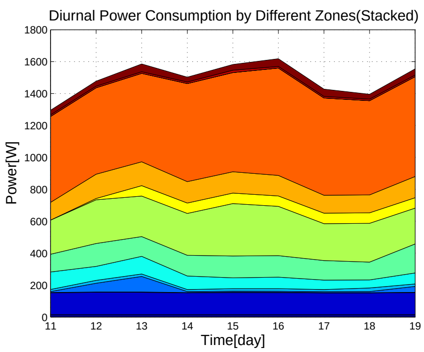


Figure 4: Diurnal Consumption Change: Average power consumption of each zone changes on a daily basis because each occupant's pattern changes. This paper exploits a tool that automatically annotates these causes and effects.

data points, m will be assumed to be finite and shorter than the given time series data.

This definition of causality implies that Y_t causes X_t provided some b_js are non-zero. Similarly, X_t is a cause of Y_t given some a_js are non-zero. This is contingent upon our intuition that the effect is the result of a cause, which denotes that the effect can be described by using a set of prior samples from the cause. The magnitude of G-causality is measured by the logarithm of the corresponding F-statistics to the causal model (Eq. 1) [8]. In short, the time-series G-causality is a measure of the F-statistics given a causal model for two stochastic processes X_t and Y_t. If the F-statistics value from Y_t to X_{t+1} is significant, we say Y_t Granger causes X_t.

Given this definition of G-causality, we now define the G-causality graph.

DEFINITION 1. *The G-causality graph is a weighted graph, $G = (V,E)$ of sets satisfying $E \subseteq (V,V,\mathbf{R})$ where V is a set of nodes such as IP addresses and electric branches, and E is a set of edges connecting these nodes with an associated weight. The edge weight is equal to the F-statistics value for a pair of nodes given the causality model in Eq. 1.*

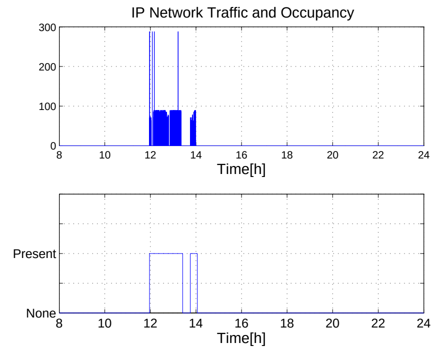


Figure 5: IP traffic is a reasonable proxy for an individual's presence. This plot shows the IP traffic generated from a user, indicating that he stayed in the lab from around Noon to 2PM with a temporary absence in the middle for a meeting.

3 Experimental Set-up

3.1 Fine grained energy monitoring set-up

We used our lab as a testbed where we currently monitor real-time energy consumption of each electrical circuit branch, and log network traffic through the firewall server. Figure 1a depicts a circuit breaker that contains 18 active power lines covering all the wall electric sockets in the lab which supply power to the server racks, desktop computers, laptops, soldering irons, table lamps, a water dispenser, and other appliances and electronic devices [1]. A spatial power drain map of the lab with corresponding electric branch numbers is shown in Figure 2. This power drain map is manually identified to see whether the G-causality analysis is contingent upon our energy use. It is interesting to note that some branches supply electricity to more than one desks that are not adjacent. This is for the phase balancing purpose.

Veris E-30 (shown in figure 1b) is a circuit breaker level power monitoring instrument that is used to monitor voltage, current, real power, reactive power, and electrical stability of these 18 active branches. It stores and updates all the measurements in its internal registers every 1.2 second. An example data set collected from Veris E-30 is shown in Figure 3. The stacked plot represents real-time power consumption of each electrical branch, from a weekday, arranged in increasing order of indexes starting from one at the bottom. Note that a few circuits are indistinguishable because they consume much less power than other circuits.

3.2 IP Network monitoring

We chose to use network traffic from an individual's IP address as a proxy for his/her presence in an office space for a couple of reasons. First, most work related tasks are done on a computer either directly or indirectly. Arguably the internet is the most integral part of an office environment today. For instance, almost everyone accesses the internet right after entering the lab. Most lab members periodically check their e-mail, use messenger services and search engines to download software and papers, and so forth. It thus provides a non-intrusive modality for sensing presence or absence of

[1]While this power panel covers most electric outlets, it does not power the ceiling lights and the HVAC unit.

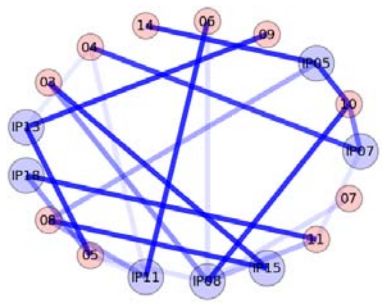


Figure 6: A complete G-causality graph from a week day.

a person.

Second, it is trivial to gather this information from most commercial or open-source firewall servers. Specifically, our lab maintains a local firewall with an open source software, pfSense [1], that provides necessary options through its web interface to log network traffic through the server. Figure 5 clearly indicates that the IP traffic data is a good indicator of presence of people in the lab.

4 Graphical G-Causality Analysis

In this section, we analyze the effectiveness and limitations of G-causality in capturing energy use patterns of the occupants in our lab. We demonstrate that, under certain assumptions, G-causality is able to successfully correlate users with their direct and indirect effect on energy consumption of circuits that power both private as well as shared zones. This is verified from user annotated diaries that were maintained during the week long experiment.

Figure 6 shows a complete G-causality graph created from a day's worth of data. The blue nodes indicate unique IP addresses corresponding to different occupants of the lab, while the red nodes denote electric circuit numbers. Edges with different opacity indicate the level of causality between connected IPs and electric circuits. Without correct context associated with the IPs and circuit numbers, it was hard to analyze the energy consumption map generated by the algorithm. Therefore, we extracted sub-graphs of all subjects and overlaid each of them on a floorplan of our lab. Figure 7 shows two such sub-graphs correspoding to IP07 and IP18. The red nodes in the sub-figures denote electric circuits that have a causal relationship with the corresponding occupant, while their size indicates the level of causality. The gray nodes are other circuits that do not have a causal relationship with the subject. Since the IPs are associated with unique users, we interchangeably refer to users with their respective IPs in this section.

4.1 Localized Energy Consumption

Our results show that G-causality correctly captures the causal relationship between an occupant's desk and the electric branch that supplies power to it. Moreover, the causality is high, which indicates that the energy consumption by individuals is localized to areas where they spend most of their time. For instance, IP07 is strongly connected to circuit number 4 (Figure 7b) that supplies electricity to the desk of corresponding occupant. Similarly, IP18 has a strong connection to circuit number 11. This reinforces our intuition described in Section 2 that correlates presence of network traffic to energy consumption from the user's desk.

4.2 Shared Spaces

G-causality goes beyond localized energy consumption to capture the causal relationship between individuals and shared appliances such as a water dispenser (with heat/cool functions) and soldering irons. This is based on the absence of network traffic from a user's desk coinciding with increased power consumption from shared devices as discussed in Section 2. Figure 2 shows that IP05 is strongly connected to circuit number 14 that powers the shared work bench used for soldering, PCB assembling and testing. This is confirmed from user annotated diaries where the corresponding occupant frequently used the soldering machine and hot air gun to fix his PCBs. Similarly, the strong connection between IP18 and circuit 5 in Figure 7a signifies the user's frequent need for hot water for her tea.

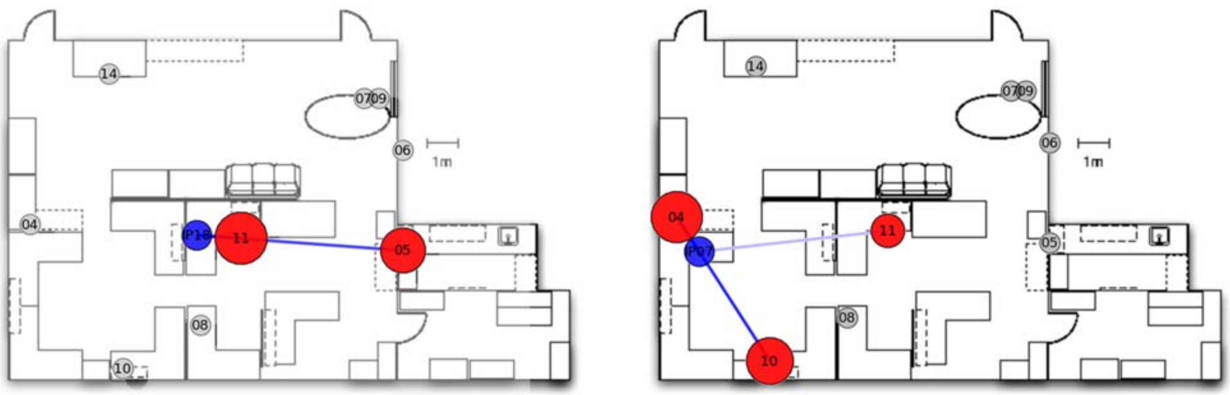
4.3 User Dynamics

An interesting but not obvious causal relationship also appears in the G-causality graphs, as is observed in Figure 7b where IP07 is connected to circuit number 10 and 11. These edges are due to interaction between IP07, and users of desks that are powered by circuits 10 and 11 respectively. On this specific day, IP07 initiated two different meetings in a shared meeting space, that is distinct from the attendees' desks, resulting in a drop in energy consumption from circuits 10 and 11. This is a particularly interesting result as it first associates IP07 with his indirect effect on circuits 10 and 11, and second, it captures the *absence* of energy consumption unlike prior work in this area.

G-causality is not only able to capture dynamics of multiple users in a single day, but a single user across multiple days too. Figure 8 illustrates diurnal changes in the G-causality graphs of a single user. Obviously, no causality appears on the graphs when user IP07 did not come to the lab (day 2, 5 and 6). IP07 has a consistent connection to his local power source, circuit number 4, though the level of causality changes every day. On day 3 and 4, IP07 is connected to circuit number 13 where a Li-Ion battery charger is connected. It is interesting to observe that the peripheral connections change slightly each day, indicating different types of causes in energy use and collaboration.

4.4 Limitations

Application of G-causality in our context is based on the assumption that the appliances consume (more) power only when they are being used. Instead, if the devices consume constant power irrespective of their utilization, the G-causal algorithm fails to attribute causes to their correct effect resulting in false negatives. For instance, we do not see any edge connected to circuit numbers 15, 16, and 17 in Figure 6. These circuits are connected to three uninterruptible power supplies (UPS) that supply power to the servers which are not designed to be energy proportional i.e. their network



(a) Causality Sub-Graph (Subject 1)

(b) Causality Sub-Graph (Subject 2)

Figure 7: G-causality Sub-graphs: Blue circles indicate IP address of each occupant. Red and gray circles indicate zone numbers superimposed on the floor plan. The size of the red circles represent the level of causality. IP traffic has strong causal relationship with nearby power sources. By looking at the sub-graph of the causality graph, we can infer the energy use map of each occupant.

load has negligible impact on their overall power consumption. However, we envision that this assumption will be satisfied in future for desktops and servers [2, 16]. For non-computing appliances directly powered on/off by the users, application of our system will serve to motivate the individuals to avoid wastage and power on the devices only when they need them.

In some other scenarios, a few false positives were reported due to comparable network activities from multiple IPs coinciding with power consumption in shared spaces. For instance, on days 3, 4 and 7 in Figure 8, IP07 is falsely connected to circuit number 6 that powers an air-pump, three test equipments and a DC power supply, despite the fact that neither IP07, nor his frequent collaborators, use these devices. Similarly, in Figure 6, IP08 is falsely connected to circuit number 7 where a television is connected. These false positives resulted in corresponding false negatives where users were not correctly attributed for their energy consumption. Since G-causality is a statistical tool, addition of learning capabilities, or supplementary information from other sensors such as infrared motion sensors and door sensors, may help address this limitation. However, we have left this as future work.

Finally, the current implementation of our system does not account for remote access to user PCs that may cause false positives when the user is not present in the lab. This gains importance in light of recent work [16, 2] that focuses on enabling remote access to computing devices while allowing them to sleep when unused. However, we believe that it is very straightforward to fix this by detecting VPN and SSH packets at the firewall and ignoring their effect on power consumption from circuits that are not directly connected to the respective PCs.

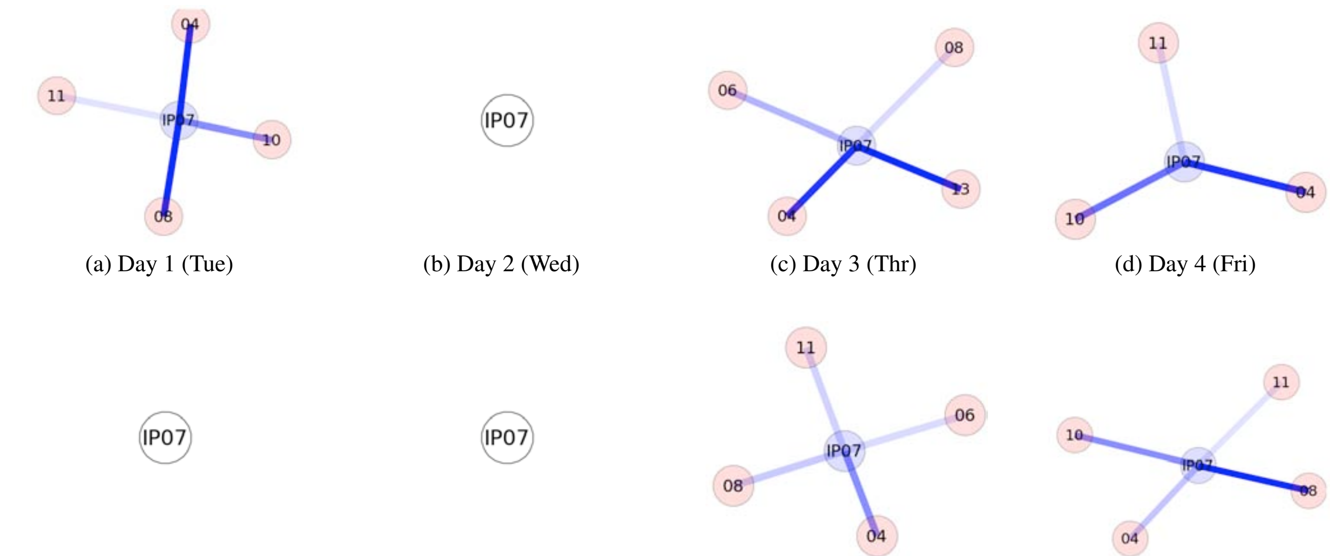
5 Related Work

The idea of motivating individuals with the right information has a positive impact on improved efficiency. Simple plots or plain consumption data alone are not effective and careful feedback design is necessary [6]. The challenge in designing an efficient eco-feedback system lies in its huge scale of data. It is evident that the scale of monitoring will become bigger and incorporate finer device-level information [11]. It will become almost impossible to manually annotate and understand implications of energy use. Unless there's a system that can automatically annotate causes and effects, associate consumption with end-users, and provide a feasible set of starting points for improved efficiency, the utility of the fine grained monitoring cannot be maximized [6].

Motivated by this, our prior work presented a prototype system that can monitor energy consumption by individuals using a proximity sensor [12]. The basic idea is that an occupant carries an active RFID tag, which is used for detecting proximity between a user and each appliance. This proximity information is then used for energy apportionment. Hay et al. also presented a case study for energy allocation where a user's occupancy information is used to assign an individual's energy footprint in a building [10]. While the authors tackle the right challenge, the system either requires users to carry active RFID tags [12] or to explicitly tap tags on RFID readers, which is cumbersome. Our preliminary G-causal analysis uses IP traffic that neither requires users to carry additional devices [10] nor depends on additional infrastructure [12].

6 Discussion and Future Work

We conducted a preliminary G-causality analysis between IP traffic and energy consumption by different zones in an office setting. Our analysis shows that it is feasible to use G-causality to start investigating personalized level energy consumption. It is evident that the analysis gives a good understanding of an individual's energy drain map as well as diurnal patterns in his/her energy consumption. The key value of our approach is that it captures an individual's energy consumption map without manual intervention or intensive instrumentation of a space and users. While device-level energy monitoring pinpoints specific energy consumption, it alone lacks the ability to capture user dynamics in energy consumption patterns. This causality analysis, on the other hand, provides personalized insights making it complemen-



(a) Day 1 (Tue) (b) Day 2 (Wed) (c) Day 3 (Thr) (d) Day 4 (Fri)

(e) Day 5 (Sat) (f) Day 6 (Sun) (g) Day 7 (Mon) (h) Day 8 (Tue)

Figure 8: Causality graphs of the subject 2 for 8 days. Black solid edges indicate strong causality and dotted lines indicate weak causality. For days when he worked from home, causality does not appear. Interestingly, the causality graph links the user to other places where his frequent collaborators sit.

tary to device-level fine grained monitoring as well (besides coarse-grained circuit-level monitoring as investigated in this text).

Although we have discussed its feasibility, this study is a first cut analysis. Further investigation is therefore required to make the analysis more insightful and usable. (1) G-causality enables an understanding of one's energy consumption. But, it neither gives 100% accurate estimate of how one is consuming energy, nor provides detailed consumption. Further analysis is required to determine if this can be used to apportion energy use by end-users. Nonetheless, it gives a set of causes (users) and effects (electric branches) that is much smaller than the whole set of occupants and zones. (2) Extending this simple causality analysis to understand consumption patterns and automatically propose remedies is a natural next step to help occupants understand implications of their energy use model. (3) Investigating the edges representing social interaction in an office could bring an interesting perspective and create a positive impact on conservation.

Acknowledgements

This material is supported in part by the National Science Foundation under grant CNS-0905580 and CCF-0820061, and the UCLA Center for Embedded Networked Sensing. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the listed funding agencies. The authors also would like to thank Lucas Wanner for the fruitful discussion.

References

[1] pfsense open source firewall distribution.

[2] Y. Agarwal, S. Savage, and R. Gupta. Sleepserver: A software-only approach for reducing the energy consumption of pcs within enterprise environments. In *USENIX*, 2010.

[3] Y. Agarwal, T. Weng, and R. K. Gupta. The energy dashboard: Improving the visibility of energy consumption at a campus-wide scale. In *BuildSys*, 2009.

[4] A. Barbato, L. Borsani, A. Capone, and S. Melzi. Home energy saving through a user profiling system based on wireless sensors. In *BuildSys*, 2009.

[5] D. T. Delaney, G. M. P. O'Hare, and A. G. Ruzzelli. Evaluation of energy-efficiency in lighting systems using sensor networks. In *BuildSys*, 2009.

[6] J. Froehlich, L. Findlater, and J. Landay. The design of eco-feedback technology. In *CHI '10*, 2010.

[7] G. Gao and K. Whitehouse. The self-programming thermostat: optimizing setback schedules based on home occupancy patterns. In *BuildSys*, 2009.

[8] J. Geweke. Measurement of linear dependence and feedback between multiple time series. *the American Statistical Association*, 1982.

[9] C. W. J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 1969.

[10] S. Hay and A. C. Rice. The case for apportionment. In *BuildSys*, 2009.

[11] X. Jiang, M. Van Ly, J. Taneja, P. Dutta, and D. Culler. Experiences with a high-fidelity wireless building energy auditing network. In *SenSys*, 2009.

[12] Y. Kim, Z. M. Charbiwala, A. Singhania, T. Schmid, and M. B. Srivastava. Spotlight: Personal natural resource consumption profiler. In *HotEmNets*, 2008.

[13] Y. Kim, T. Schmid, Y. Wang, and M. B. Srivastava. Challenges in resource monitoring for residential spaces. In *BuildSys*, 2009.

[14] A. Marchiori and Q. Han. Using circuit-level power measurements in household energy management systems. In *BuildSys*, 2009.

[15] S. N. Patel, T. Robertson, J. A. Kientz, M. S. Reynolds, and G. D. Abowd. At the flick of a switch: Detecting and classifying unique electrical events on the residential power line. In *UbiComp*, 2007.

[16] J. Reich, M. Goraczko, A. Kansal, and J. Padhye. Sleepless in seattle no longer. In *USENIX*, 2010.

[17] A. K. Seth. Causal connectivity of evolved neural networks during behavior. *Network: Computation in Neural Systems*, 2005.

[18] A. K. Seth, J. R. Iversen, and G. M. Edelman. Single trial discrimination of truthful from deceptive responses during a game of financial risk using alpha-band meg signals. *Neuroimage*, 2006.