

Writing 101: Towards Better Research Abstracts



Nipun Batra
July 27, 2017

Paper Structure

Gemello: Creating a Detailed Energy Breakdown from just the Monthly Electricity Bill

Nipun Batra
IIT Delhi
nipunb@iitd.ac.in

Amarjeet Singh
IIT Delhi
amarjeet@iitd.ac.in

Kamin Whitehouse
University of Virginia
whitehouse@virginia.edu

ABSTRACT

The first step to saving energy in the home is often to create an *energy breakdown*: the amount of energy used by each individual appliance in the home. Unfortunately, current techniques that produce an energy breakdown are not *scalable*: they require hardware to be installed in each and every home. In this paper, we propose a more scalable solution called *Gemello* that estimates the energy breakdown for one home by matching it with similar homes for which the breakdown is already known. This matching requires only the monthly energy bill and household characteristics such as square footage of the home and the size of the household. We evaluate this approach using 57 homes and results indicate that the accuracy of Gemello is comparable to or better than existing techniques that use sensing infrastructure in each home. The information required by Gemello is often publicly available and, as such, it can be immediately applied to many homes around the world.

1. INTRODUCTION

Buildings account for more than 30% of total energy usage around the world, of which up to 93% is due to residential buildings [41, 18, 42]. Some of this energy can be saved by producing an *energy breakdown*: the amount of energy used by each individual appliance in the home, akin to the itemised bills we get from grocery stores. With such a breakdown, utility companies can focus conservation programs on homes that have an especially inefficient appliance, such as a fridge, water heater, or air conditioner. Energy feedback can also help induce energy-saving behaviour in the occupants themselves [15, 4, 30, 43, 24].

Unfortunately, current techniques that produce an energy breakdown are not *scalable*: they require hardware to be installed in each home and/or intensive manual training of the system. One approach, called non-intrusive load monitoring (NILM), tries to infer the energy breakdown based on the aggregate power trace from a single smart meter installed in the home. This approach is perhaps the most practi-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

KDD '16, August 13 - 17, 2016, San Francisco, CA, USA

© 2016 Copyright held by the owner(s). Publication rights licensed to ACM. ISBN 978-1-4503-4232-2/16/08...\$15.00

DOI: <http://dx.doi.org/10.1145/2939672.2939735>

cal because smart meters are already being rolled out in millions of homes worldwide. However, current techniques require high resolution data (1 minute sampling frequency or higher) [39, 11, 4] while most smart meters today only support 15-minute or hourly sample rates to support time-of-use energy pricing. Even if smart meters had a higher sampling rate, most of the world does not yet have smart meters and many places do not even have plans to deploy them. Alternatives to NILM are more accurate but require specialized sensors to be installed inside the home [20], on each individual appliance [32, 26, 16], or on each circuit in the breaker box [37]. All of these solutions are limited by the need for instrumentation to be deployed in every home.

In this paper, we propose a more scalable solution called *Gemello* that produces an energy breakdown in homes without requiring new hardware to be installed in each home. Instead, Gemello estimates the energy breakdown for a home by matching it with similar homes that do have a hardware-based disaggregation solution. This matching requires only the monthly energy bill and household characteristics such as square footage of the home and the size of the household. From an energy perspective, homes in the same geographic region are often very similar because they have similar construction methods, use the same heating fuels, and contain similar fridges, washing machines, and other appliances. Gemello exploits this fact to provide an energy breakdown for many homes in a region by instrumenting only a fraction of them.

Of course, no two homes are exactly identical and finding a perfect twin is unlikely. Therefore, Gemello uses a different set of matching homes to estimate the energy usage of each individual appliance. The key to success is the ability to define 'similarity' on a per-appliance basis. For example, homes with similar seasonal trends in their monthly energy bill are expected to have similar heating/cooling energy (referred as heating, ventilation, and air conditioning [HVAC] from now on); homes with similar square footage are expected to have similar lighting loads; and homes with a similar number of occupants are expected to have similar washing machine energy usage because that is driven by the amount of clothes worn each day. The energy usage of each appliance is predicted by a different set of features, and so Gemello finds a different set of homes to predict the energy usage of each appliance.

We evaluate this approach using 57 homes from the publicly available Dataport data set [38], in which the ground truth energy breakdown is measured by metering each appliance of the home individually for one year or more. Results

show that the accuracy of Gemello is comparable to or better than established NILM techniques called FHMM and LBM, both of which require on-site, high-frequency power metering in each home [33, 45]. Gemello can achieve 76% and 69% accuracy on HVAC and fridge, compared to 61% and 39% for FHMM and 56% and 72% for LBM. Furthermore, it achieves up to 57% accuracy on washing machine, dryer, dishwasher, and lighting loads, which is higher than previously reported results. Many existing techniques are not able to disaggregate these loads at all. Our analysis shows that these results are robust with as few as 7 instrumented homes: the accuracy for HVAC loads is $\approx 69\%$, still 8% better than state-of-the-art approaches. The accuracy of Gemello becomes higher in homes with smart meters that provide power readings at 15-minute resolution.

The Gemello technique has potential for immediate impact because all of the information it requires is already available. Essentially all homes already have a standard power meter that provides total monthly energy usage for billing purposes, and in many regions this information can be downloaded online¹. Household characteristics such as the size of the home and the number of people in the household are often publicly available and are already being used by companies to match homes for other types of eco-feedback². Finally, many companies are already collecting submetering information from thousands of homes around the world³. By combining these three sources of information, Gemello could be used to provide an energy breakdown for homes around the world without the need to install new instrumentation in each location.

2. RELATED WORK

Various techniques for measuring appliance level energy consumption have been studied in the past. The simplest technique is to install appliance level sensors that monitor and report appliance energy consumption [26, 16]. Many commercial vendors are also selling appliance level sensors such as Kill-A-Watt⁴ and Hobo plug load data logger⁵. Instead of directly monitoring the appliance of interest, a few recent systems have looked into indirect sensing techniques. Kim et al. [32] develop a system called Viridiscopes that leverages the correlation amongst sensor streams, like using a vibration sensor on a fridge to tell if the compressor is running or not, and then using a model to determine fridge's power. Similarly, Clark et al. [13] develop a system called Deltaloss that employs energy harvesting sensors and performs computation on the activation of these sensors to determine appliance power draw. Jain et al. [25] install temperature sensors inside a home to estimate air conditioner energy usage. Gupta et al. [20], Chen et al. [12] and Gulati et al. [19] use the electromagnetic interference typically generated by electronic appliances to determine appliance usages. Previous work has also looked into using sensors deployed at household circuits (lesser in number than individual appliances) to infer per-appliance energy usage [37]. All these solutions are limited by the need for instrumentation to be

¹<http://www.greenbuttondata.org/>
²<https://opower.com/>
³<https://powerhouseanalytics.com/>
⁴<http://www.gointernational.com/products/p4400.html>
⁵<http://www.onsetcomp.com/products/data-loggers/ux120-018>

deployed in every home. Previous research has shown that there are various challenges beyond cost of the system that prevent such systems from being scaled [23].

NILM is often considered the most practical approach towards generating energy breakdown. However, like other approaches mentioned above, it also requires instrumentation in each home. While the effort of deploying a smart meter may be less in comparison to other related approaches, most of the world still does not have smart meter infrastructure.

Since its inception in the early 1980s by George Hart [21], the field of NILM has seen various approaches based on different machine learning techniques leveraging different features of the power trace. However, many of these approaches require submetered data to learn a model of each appliance and these models have not been shown to generalize well across homes. Even if the models did generalise, NILM approaches require high frequency power metering with resolutions of 1-minute or higher. Very high frequency approaches (>10 kHz) [11] use features such as voltage-current trajectories to detect events in aggregate power time series. However, current smart meters do not collect data at such high rates because they are designed and deployed for the purposes of time-of-use pricing and there are currently no efforts to deploy devices suitable for energy disaggregation on a large scale. Therefore, these techniques, while promising, face real practical barriers before being used at scale.

Additionally, existing approaches for energy disaggregation [33, 39, 21, 45, 35] require a model of each appliance. The main differences between these techniques are how they are created and how they are used to infer the hidden states of the appliances based on the aggregate power readings. For example, some systems model appliances as finite state machines (FSMs). However, such approaches generally show poor accuracy on complex appliances such as washing machine and other electronics, as a FSM is a poor model for such appliances. Some systems assume the model is manually generated, learned from training data [21, 33], and in rare cases learned automatically [5]. In all cases, however, the accuracy of these models depends on how well the model approximates the true appliances in the home and it has not yet been demonstrated that these model-based approaches generalize well across homes. Only recently, researchers have started looking into automatically learning arbitrarily complex appliances using deep neural networks [29]. However, as claimed by the authors themselves, the work is just scratching the surface. These challenges are likely to impact the generalisability and applicability of existing approaches towards solving the energy breakdown problem.

3. APPROACH

The goal of Gemello is to predict the energy consumption of household appliances given easy to collect information such as a single aggregate energy reading per month and static household characteristics. The key intuition behind Gemello is that if per appliance energy consumption is available for a home then any other home that is 'similar' in nature will also have the same energy consumption for that appliance. This intuition leads to the following specific requirements for Gemello to work:

1. Small set of submetered homes (SH) for which per appliance energy consumption is directly measured.
2. Easy to collect information for all the homes in the

back on an early version of this work at Sensys 2015 PhD forum. Arunchandar Vasan, Venkatesh Sarangan and Mario Berges provided useful suggestions. Jack Kelly, Oliver Parson, Mingjun Zhong, Manoj Gulati, Milan Jain provided detailed pre-submission reviews for our paper, in addition to deep technical discussions. Mingjun Zhong very kindly helped us use his LBM algorithm through Skype discussions and 100+ emails. Grant Fisher provided information about the dataset.

This work was funded in part by NSF awards 1305362, 1038271, and 0845761. The first author is supported by TCS Research and Development PhD fellowship. We also thank Department of Electronics and Information Technology (Government of India) for funding the project (Grant Number DeitY/R&D/ITEA/4(2)/2012 and Grant Number ITRA/15(57)/Mobile/HumanSense/01). We would also like to thank Google for providing support for the conference.

8. REFERENCES

- [1] J. Alcalá, O. Parson, and A. Rogers. Detecting anomalies in activities of daily living of elderly residents via energy disaggregation and cox processes. In *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*, pages 225-234. ACM, 2015.
 - [2] K. Anderson, A. Oconeau, D. Benitez, D. Carlson, A. Rowe, and M. Bergés. BLUEED: A fully labeled public dataset for Event-Based Non-Intrusive load monitoring research. In *Proceedings of 2nd KDD Workshop on Data Mining Applications in Sustainability*, pages 12-16, Beijing, China, 2012.
 - [3] P. Arjunan, H. D. Khadilkar, T. Gann, Z. M. Charbiwala, A. Singh, and P. Singh. Multi-user energy consumption monitoring and anomaly detection with partial context information. In *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*, pages 35-44. ACM, 2015.
 - [4] K. C. Armel, A. Gupta, C. Shrivastava, and A. Albert. Is disaggregation the holy grail of energy efficiency? The case of electricity. *Energy Policy*, 52:213-234, 2013.
 - [5] S. Barker, S. Kalra, D. Irwin, and P. Shenoy. Empirical characterization and modeling of electrical loads in smart homes. In *IEEE International Green Computing Conference*, pages 1-10, Arlington, VA, USA, 2013.
 - [6] N. Batra, R. Bajjal, A. Singh, and K. Whitehouse. How good is good enough? re-evaluating the bar for energy disaggregation. *arXiv preprint arXiv:1510.08718*, 2015.
 - [7] N. Batra, M. Gulati, A. Singh, and M. B. Srivastava. It's Different: Insights into home energy consumption in India. In *Proceedings of the Fifth ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*, 2013.
 - [8] N. Batra, J. Kelly, O. Parson, H. Dutta, W. Knottenbelt, A. Rogers, A. Singh, and M. Srivastava. NILMTK: An Open Source Toolkit for Non-intrusive Load Monitoring. In *Fifth International Conference on Future Energy Systems (ACM e-Energy)*, Cambridge, UK, 2014.
 - [9] N. Batra, A. Singh, and K. Whitehouse. If you
- energy disaggregation. In *Proceedings of the second ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*. ACM, 2015.
- [10] C. Beckel, W. Kleininger, R. Cicchetti, T. Staake, and S. Santini. The eco data set and the performance of non-intrusive load monitoring algorithms. In *Proceedings of the First ACM International Conference on Embedded Systems for Energy-Efficient Buildings*. ACM, 2014.
 - [11] M. E. Berges, E. Goldman, H. S. Matthews, and L. Soibelman. Enhancing electricity audits in residential buildings with nonintrusive load monitoring. *Journal of industrial ecology*, 14(5):844-858, 2010.
 - [12] K.-Y. Chen, S. Gupta, E. C. Larson, and S. Patel. Dose: Detecting user-driven operating states of electronic devices from a single sensing point. In *Pervasive Computing and Communications (PerCom), 2015 IEEE International Conference on*, pages 46-54. IEEE, 2015.
 - [13] M. Clark, B. Campbell, and P. Dutta. Deltaloss: submetering by synthesizing uncalibrated pulse sensor streams. In *Proceedings of the 5th international conference on Future energy systems*, pages 301-311. ACM, 2014.
 - [14] R. B. Cleveland, W. S. Cleveland, J. E. McRae, and I. Terpenning. Stl: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, 6(1):3-73, 1990.
 - [15] S. Darby. The effectiveness of feedback on the energy consumption. *A Review of DEFRA Literature on Metering, Billing and Direct Displays*, 2006.
 - [16] S. DeBruin, B. Ghena, Y.-S. Kuo, and P. Dutta. Powerblade: A low-profile, true-power, plug-through energy meter. In *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems*, pages 17-29. ACM, 2015.
 - [17] EIA. Household energy use in texas, 2014.
 - [18] M. Evans, B. Shui, and S. Somsundaram. *Country report on building energy codes in india*. Pacific Northwest National Laboratory, 2009.
 - [19] M. Gulati, S. S. Ram, A. Majumdar, and A. Singh. Single point conducted emi sensor with intelligent inference for detecting it appliances. In *IITD Technical Report*, 2015.
 - [20] S. Gupta, M. S. Reynolds, and S. N. Patel. Electrisense: single-point sensing using emi for electrical event detection and classification in the home. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*, pages 139-148. ACM, 2010.
 - [21] G. W. Hart. Nonintrusive appliance load monitoring. *Proceedings of the IEEE*, 80(12):1870-1891, 1992.
 - [22] T. Hastie, R. Tibshirani, J. Friedman, and J. Franklin. The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer*, 27(2):83-85, 2005.
 - [23] T. W. Hnat, V. Srivastava, J. Lu, T. I. Sookoor, R. Dawson, J. Stankovic, and K. Whitehouse. The hitchhiker's guide to successful residential sensing deployments. In *Proceedings of the 9th ACM*

Paper Structure

Title

Gemello: Creating a Detailed Energy Breakdown from just the Monthly Electricity Bill

Nipun Batra
IIT Delhi
nipunb@iitd.ac.in

Amarjeet Singh
IIT Delhi
amarjeet@iitd.ac.in

Kamin Whitehouse
University of Virginia
whitehouse@virginia.edu

ABSTRACT

The first step to saving energy in the home is often to create an *energy breakdown*: the amount of energy used by each individual appliance in the home. Unfortunately, current techniques that produce an energy breakdown are not *scalable*: they require hardware to be installed in each and every home. In this paper, we propose a more scalable solution called *Gemello* that estimates the energy breakdown for one home by matching it with similar homes for which the breakdown is already known. This matching requires only the monthly energy bill and household characteristics such as square footage of the home and the size of the household. We evaluate this approach using 57 homes and results indicate that the accuracy of Gemello is comparable to or better than existing techniques that use sensing infrastructure in each home. The information required by Gemello is often publicly available and, as such, it can be immediately applied to many homes around the world.

1. INTRODUCTION

Buildings account for more than 30% of total energy usage around the world, of which up to 93% is due to residential buildings [41, 18, 42]. Some of this energy can be saved by producing an *energy breakdown*: the amount of energy used by each individual appliance in the home, akin to the itemised bills we get from grocery stores. With such a breakdown, utility companies can focus conservation programs on homes that have an especially inefficient appliance, such as a fridge, water heater, or air conditioner. Energy feedback can also help induce energy-saving behaviour in the occupants themselves [15, 4, 30, 43, 24].

Unfortunately, current techniques that produce an energy breakdown are not *scalable*: they require hardware to be installed in each home and/or intensive manual training of the system. One approach, called non-intrusive load monitoring (NILM), tries to infer the energy breakdown based on the aggregate power trace from a single smart meter installed in the home. This approach is perhaps the most practi-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

KDD '16, August 13 - 17, 2016, San Francisco, CA, USA

© 2016 Copyright held by the owner(s). Publication rights licensed to ACM. ISBN 978-1-4503-4232-2/16/08...\$15.00
DOI: <http://dx.doi.org/10.1145/2939672.2939735>

cal because smart meters are already being rolled out in millions of homes worldwide. However, current techniques require high resolution data (1 minute sampling frequency or higher) [39, 11, 4] while most smart meters today only support 15-minute or hourly sample rates to support time-of-use energy pricing. Even if smart meters had a higher sampling rate, most of the world does not yet have smart meters and many places do not even have plans to deploy them. Alternatives to NILM are more accurate but require specialized sensors to be installed inside the home [20], on each individual appliance [32, 26, 16], or on each circuit in the breaker box [37]. All of these solutions are limited by the need for instrumentation to be deployed in every home.

In this paper, we propose a more scalable solution called *Gemello* that produces an energy breakdown in homes without requiring new hardware to be installed in each home. Instead, Gemello estimates the energy breakdown for a home by matching it with similar homes that do have a hardware-based disaggregation solution. This matching requires only the monthly energy bill and household characteristics such as square footage of the home and the size of the household. From an energy perspective, homes in the same geographic region are often very similar because they have similar construction methods, use the same heating fuels, and contain similar fridges, washing machines, and other appliances. Gemello exploits this fact to provide an energy breakdown for many homes in a region by instrumenting only a fraction of them.

Of course, no two homes are exactly identical and finding a perfect twin is unlikely. Therefore, Gemello uses a different set of matching homes to estimate the energy usage of each individual appliance. The key to success is the ability to define 'similarity' on a per-appliance basis. For example, homes with similar seasonal trends in their monthly energy bill are expected to have similar heating/cooling energy (referred as heating, ventilation, and air conditioning [HVAC] from now on); homes with similar square footage are expected to have similar lighting loads; and homes with a similar number of occupants are expected to have similar washing machine energy usage because that is driven by the amount of clothes worn each day. The energy usage of each appliance is predicted by a different set of features, and so Gemello finds a different set of homes to predict the energy usage of each appliance.

We evaluate this approach using 57 homes from the publicly available Dataport data set [38], in which the ground truth energy breakdown is measured by metering each appliance of the home individually for one year or more. Results

show that the accuracy of Gemello is comparable to or better than established NILM techniques called FHMM and LBM, both of which require on-site, high-frequency power metering in each home [33, 45]. Gemello can achieve 76% and 69% accuracy on HVAC and fridge, compared to 61% and 39% for FHMM and 56% and 72% for LBM. Furthermore, it achieves up to 57% accuracy on washing machine, dryer, dishwasher, and lighting loads, which is higher than previously reported results. Many existing techniques are not able to disaggregate these loads at all. Our analysis shows that these results are robust with as few as 7 instrumented homes: the accuracy for HVAC loads is $\approx 69\%$, still 8% better than state-of-the-art approaches. The accuracy of Gemello becomes higher in homes with smart meters that provide power readings at 15-minute resolution.

The Gemello technique has potential for immediate impact because all of the information it requires is already available. Essentially all homes already have a standard power meter that provides total monthly energy usage for billing purposes, and in many regions this information can be downloaded online¹. Household characteristics such as the size of the home and the number of people in the household are often publicly available and are already being used by companies to match homes for other types of eco-feedback². Finally, many companies are already collecting submetering information from thousands of homes around the world³. By combining these three sources of information, Gemello could be used to provide an energy breakdown for homes around the world without the need to install new instrumentation in each location.

2. RELATED WORK

Various techniques for measuring appliance level energy consumption have been studied in the past. The simplest technique is to install appliance level sensors that monitor and report appliance energy consumption [26, 16]. Many commercial vendors are also selling appliance level sensors such as Kill-A-Watt⁴ and Hobo plug load data logger⁵. Instead of directly monitoring the appliance of interest, a few recent systems have looked into indirect sensing techniques. Kim et al. [32] develop a system called Viridiscopes that leverages the correlation amongst sensor streams, like using a vibration sensor on a fridge to tell if the compressor is running or not, and then using a model to determine fridge's power. Similarly, Clark et al. [13] develop a system called Deltaloss that employs energy harvesting sensors and performs computation on the activation of these sensors to determine appliance power draw. Jain et al. [25] install temperature sensors inside a home to estimate air conditioner energy usage. Gupta et al. [20], Chen et al. [12] and Gulati et al. [19] use the electromagnetic interference typically generated by electronic appliances to determine appliance usages. Previous work has also looked into using sensors deployed at household circuits (lesser in number than individual appliances) to infer per-appliance energy usage [37]. All these solutions are limited by the need for instrumentation to be

¹<http://www.greenbuttondata.org/>
²<https://opower.com/>
³<https://powerhousesdynamics.com/>
⁴<http://www.giinterational.com/products/p4400.html>
⁵<http://www.onsetcomp.com/products/data-loggers/ux120-018>

deployed in every home. Previous research has shown that there are various challenges beyond cost of the system that prevent such systems from being scaled [23].

NILM is often considered the most practical approach towards generating energy breakdown. However, like other approaches mentioned above, it also requires instrumentation in each home. While the effort of deploying a smart meter may be less in comparison to other related approaches, most of the world still does not have smart meter infrastructure.

Since its inception in the early 1980s by George Hart [21], the field of NILM has seen various approaches based on different machine learning techniques leveraging different features of the power trace. However, many of these approaches require submetered data to learn a model of each appliance and these models have not been shown to generalize well across homes. Even if the models did generalise, NILM approaches require high frequency power metering with resolutions of 1-minute or higher. Very high frequency approaches (>10 kHz) [11] use features such as voltage-current trajectories to detect events in aggregate power time series. However, current smart meters do not collect data at such high rates because they are designed and deployed for the purposes of time-of-use pricing and there are currently no efforts to deploy devices suitable for energy disaggregation on a large scale. Therefore, these techniques, while promising, face real practical barriers before being used at scale.

Additionally, existing approaches for energy disaggregation [33, 39, 21, 45, 35] require a model of each appliance. The main differences between these techniques are how they are created and how they are used to infer the hidden states of the appliances based on the aggregate power readings. For example, some systems model appliances as finite state machines (FSMs). However, such approaches generally show poor accuracy on complex appliances such as washing machine and other electronics, as a FSM is a poor model for such appliances. Some systems assume the model is manually generated, learned from training data [21, 33], and in rare cases learned automatically [5]. In all cases, however, the accuracy of these models depends on how well the model approximates the true appliances in the home and it has not yet been demonstrated that these model-based approaches generalize well across homes. Only recently, researchers have started looking into automatically learning arbitrarily complex appliances using deep neural networks [29]. However, as claimed by the authors themselves, the work is just scratching the surface. These challenges are likely to impact the generalisability and applicability of existing approaches towards solving the energy breakdown problem.

3. APPROACH

The goal of Gemello is to predict the energy consumption of household appliances given easy to collect information such as a single aggregate energy reading per month and static household characteristics. The key intuition behind Gemello is that if per appliance energy consumption is available for a home then any other home that is 'similar' in nature will also have the same energy consumption for that appliance. This intuition leads to the following specific requirements for Gemello to work:

1. Small set of submetered homes (SH) for which per appliance energy consumption is directly measured.
2. Easy to collect information for all the homes in the

back on an early version of this work at Sensys 2015 PhD forum. Arunchandar Vasani, Venkatesh Sarangan and Mario Berges provided useful suggestions. Jack Kelly, Oliver Parson, Mingjun Zhong, Manoj Gulati, Milan Jain provided detailed pre-submission reviews for our paper, in addition to deep technical discussions. Mingjun Zhong very kindly helped us use his LBM algorithm through Skype discussions and 100+ emails. Grant Fisher provided information about the dataset.

This work was funded in part by NSF awards 1305362, 1038271, and 0845761. The first author is supported by TCS Research and Development PhD fellowship. We also thank Department of Electronics and Information Technology (Government of India) for funding the project (Grant Number DeitY/R&D/ITEA/4(2)/2012 and Grant Number ITRA/15(57)/Mobile/HumanSense/01). We would also like to thank Google for providing support for the conference.

8. REFERENCES

- [1] J. Alcalá, O. Parson, and A. Rogers. Detecting anomalies in activities of daily living of elderly residents via energy disaggregation and cox processes. In *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*, pages 225-234. ACM, 2015.
 - [2] K. Anderson, A. Oconeau, D. Benitez, D. Carlson, A. Rowe, and M. Bergés. BLUEED: A fully labeled public dataset for Event-Based Non-Intrusive load monitoring research. In *Proceedings of 2nd KDD Workshop on Data Mining Applications in Sustainability*, pages 12-16, Beijing, China, 2012.
 - [3] P. Arjunan, H. D. Khadilkar, T. Gann, Z. M. Charbiwala, A. Singh, and P. Singh. Multi-user energy consumption monitoring and anomaly detection with partial context information. In *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*, pages 35-44. ACM, 2015.
 - [4] K. C. Armel, A. Gupta, C. Shrivastava, and A. Albert. Is disaggregation the holy grail of energy efficiency? The case of electricity. *Energy Policy*, 52:213-234, 2013.
 - [5] S. Barker, S. Kalra, D. Irwin, and P. Shenoy. Empirical characterization and modeling of electrical loads in smart homes. In *IEEE International Green Computing Conference*, pages 1-10, Arlington, VA, USA, 2013.
 - [6] N. Batra, R. Bajjal, A. Singh, and K. Whitehouse. How good is good enough? re-evaluating the bar for energy disaggregation. *arXiv preprint arXiv:1510.08718*, 2015.
 - [7] N. Batra, M. Gulati, A. Singh, and M. B. Srivastava. It's Different: Insights into home energy consumption in India. In *Proceedings of the Fifth ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*, 2013.
 - [8] N. Batra, J. Kelly, O. Parson, H. Dutta, W. Knottenbelt, A. Rogers, A. Singh, and M. Srivastava. NILMTK: An Open Source Toolkit for Non-intrusive Load Monitoring. In *Fifth International Conference on Future Energy Systems (ACM e-Energy)*, Cambridge, UK, 2014.
 - [9] N. Batra, A. Singh, and K. Whitehouse. If you
- energy disaggregation. In *Proceedings of the second ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*. ACM, 2015.
- [10] C. Beckel, W. Kleininger, R. Cicchetti, T. Staake, and S. Santini. The eco data set and the performance of non-intrusive load monitoring algorithms. In *Proceedings of the First ACM International Conference on Embedded Systems for Energy-Efficient Buildings*. ACM, 2014.
 - [11] M. E. Berges, E. Goldman, H. S. Matthews, and L. Soibelman. Enhancing electricity audits in residential buildings with nonintrusive load monitoring. *Journal of industrial ecology*, 14(5):844-858, 2010.
 - [12] K.-Y. Chen, S. Gupta, E. C. Larson, and S. Patel. Dose: Detecting user-driven operating states of electronic devices from a single sensing point. In *Pervasive Computing and Communications (PerCom), 2015 IEEE International Conference on*, pages 46-54. IEEE, 2015.
 - [13] M. Clark, B. Campbell, and P. Dutta. Deltaloss: submetering by synthesizing uncalibrated pulse sensor streams. In *Proceedings of the 5th international conference on Future energy systems*, pages 301-311. ACM, 2014.
 - [14] R. B. Cleveland, W. S. Cleveland, J. E. McRae, and I. Terpenning. StI: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, 6(1):3-73, 1990.
 - [15] S. Darby. The effectiveness of feedback on the energy consumption. *A Review for DEFRA on Literature on Metering, Billing and Direct Displays*, 2006.
 - [16] S. DeBruin, B. Ghena, Y.-S. Kuo, and P. Dutta. Powerblade: A low-profile, true-power, plug-through energy meter. In *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems*, pages 17-29. ACM, 2015.
 - [17] EIA. Household energy use in texas, 2014.
 - [18] M. Evans, B. Shui, and S. Somsundaram. *Country report on building energy codes in india*. Pacific Northwest National Laboratory, 2009.
 - [19] M. Gulati, S. S. Ram, A. Majumdar, and A. Singh. Single point conducted emi sensor with intelligent inference for detecting it appliances. In *IITD Technical Report*, 2015.
 - [20] S. Gupta, M. S. Reynolds, and S. N. Patel. Electrisense: single-point sensing using emi for electrical event detection and classification in the home. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*, pages 139-148. ACM, 2010.
 - [21] G. W. Hart. Nonintrusive appliance load monitoring. *Proceedings of the IEEE*, 80(12):1870-1891, 1992.
 - [22] T. Hastie, R. Tibshirani, J. Friedman, and J. Franklin. The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer*, 27(2):83-85, 2005.
 - [23] T. W. Hnat, V. Srivastava, J. Lu, T. I. Sookoor, R. Dawson, J. Stankovic, and K. Whitehouse. The hitchhiker's guide to successful residential sensing deployments. In *Proceedings of the 9th ACM*

Paper Structure

Title

Gemello: Creating a Detailed Energy Breakdown from just the Monthly Electricity Bill

Nipun Batra
IIT Delhi
nipunb@iitd.ac.in

Amarjeet Singh
IIT Delhi
amarjeet@iitd.ac.in

Kamin Whitehouse
University of Virginia
whitehouse@virginia.edu

ABSTRACT

The first step to saving energy in the home is often to create an *energy breakdown*: the amount of energy used by each individual appliance in the home. Unfortunately, current techniques that produce an energy breakdown are not *scalable*: they require hardware to be installed in each and every home. In this paper, we propose a more scalable solution called *Gemello* that estimates the energy breakdown for one home by matching it with similar homes for which the breakdown is already known. This matching requires only the monthly energy bill and household characteristics such as square footage of the home and the size of the household. We evaluate this approach using 57 homes and results indicate that the accuracy of Gemello is comparable to or better than existing techniques that use sensing infrastructure in each home. The information required by Gemello is often publicly available and, as such, it can be immediately applied to many homes around the world.

1. INTRODUCTION

Buildings account for more than 30% of total energy usage around the world, of which up to 93% is due to residential buildings [41, 18, 42]. Some of this energy can be saved by producing an *energy breakdown*: the amount of energy used by each individual appliance in the home, akin to the itemised bills we get from grocery stores. With such a breakdown, utility companies can focus conservation programs on homes that have an especially inefficient appliance, such as a fridge, water heater, or air conditioner. Energy feedback can also help induce energy-saving behaviour in the occupants themselves [15, 4, 30, 43, 24].

Unfortunately, current techniques that produce an energy breakdown are not *scalable*: they require hardware to be installed in each home and/or intensive manual training of the system. One approach, called non-intrusive load monitoring (NILM), tries to infer the energy breakdown based on the aggregate power trace from a single smart meter installed in the home. This approach is perhaps the most practi-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or to publish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

KDD '16, August 13 - 17, 2016, San Francisco, CA, USA

© 2016 Copyright held by the owner/author(s). Publication rights licensed to ACM. ISBN 978-1-4503-4232-2/16/08...\$15.00
DOI: <http://dx.doi.org/10.1145/2939672.2939735>

cal because smart meters are already being rolled out in millions of homes worldwide. However, current techniques require high resolution data (1 minute sampling frequency or higher) [39, 11, 4] while most smart meters today only support 15-minute or hourly sampling rates to support time-of-use energy pricing. Even with higher resolution data, sampling rates, most of the work meters and many places do not have them. Alternatives to NILM are specialized sensors to be installed on each individual appliance [32], the breaker box [37]. All of these require the need for instrumentation to be installed in each home.

In this paper, we propose a new technique called *Gemello* that produces an energy breakdown in homes without requiring new hardware to be installed in each home. Instead, Gemello estimates the energy breakdown for a home by matching it with similar homes that do have a hardware-based disaggregation solution. This matching requires only the monthly energy bill and household characteristics such as square footage of the home and the size of the household. From an energy perspective, homes in the same geographic region are often very similar because they have similar construction methods, use the same heating fuels, and contain similar fridges, washing machines, and other appliances. Gemello exploits this fact to provide an energy breakdown for many homes in a region by instrumenting only a fraction of them.

Of course, no two homes are exactly identical and finding a perfect twin is unlikely. Therefore, Gemello uses a different set of matching homes to estimate the energy usage of each individual appliance. The key to success is the ability to define 'similarity' on a per-appliance basis. For example, homes with similar seasonal trends in their monthly energy bill are expected to have similar heating/cooling energy (referred as heating, ventilation, and air conditioning [HVAC] from now on); homes with similar square footage are expected to have similar lighting loads; and homes with a similar number of occupants are expected to have similar washing machine energy usage because that is driven by the amount of clothes worn each day. The energy usage of each appliance is predicted by a different set of features, and so Gemello finds a different set of homes to predict the energy usage of each appliance.

We evaluate this approach using 57 homes from the publicly available Dataport data set [38], in which the ground truth energy breakdown is measured by metering each appliance of the home individually for one year or more. Results

show that the accuracy of Gemello is comparable to or better than established NILM techniques called FHMM and LBM, both of which require on-site, high-frequency power metering in each home [33, 45]. Gemello can achieve 76% and 69% accuracy on HVAC and fridge, compared to 61% and 39% for FHMM and 56% and 72% for LBM. Furthermore, it achieves up to 57% accuracy on washing machine, dryer, dishwasher, and lighting loads, which is higher than previously reported results. Many existing techniques are not able to disaggregate these loads at all. Our analysis shows that these results are robust with as few as 7 instrumented homes: the accuracy for HVAC loads is $\approx 69\%$, still 8% better than state-of-the-art approaches. The accuracy of Gemello becomes higher in homes with smart meters that provide power readings at 15-minute resolution.

The Gemello technique has potential for immediate impact because all of the information it requires is already available. Essentially all homes already have a standard power meter that provides total monthly energy usage for billing purposes, and in many regions this information can be downloaded online¹. Household characteristics such as the size of the home and the number of people in the household are often publicly available and are already being used by companies to match homes for other types of eco-feedback². Finally, many companies are already collecting submetering information from thousands of homes around the world³. By combining these three sources of information, Gemello could be used to provide an energy breakdown for homes around the world without the need to install new instrumentation in each location.

2. RELATED WORK

Various techniques for measuring appliance level energy consumption have been studied in the past. The simplest technique is to install appliance level sensors that monitor and report appliance energy consumption [26, 16]. Many commercial vendors are also selling appliance level sensors such as Kill-A-Watt⁴ and Hobo plug load data logger⁵. Instead of directly monitoring the appliance of interest, a few recent systems have looked into indirect sensing techniques. Kim et al. [32] develop a system called Viridiscopes that leverages the correlation amongst sensor streams, like using a vibration sensor on a fridge to tell if the compressor is running or not, and then using a model to determine fridge's power. Similarly, Clark et al. [13] develop a system called Deltalflow that employs energy harvesting sensors and performs computation on the activation of these sensors to determine appliance power draw. Jain et al. [25] install temperature sensors inside a home to estimate air conditioner energy usage. Gupta et al. [20], Chen et al. [12] and Gulati et al. [19] use the electromagnetic interference typically generated by electronic appliances to determine appliance usages. Previous work has also looked into using sensors deployed at household circuits (lesser in number than individual appliances) to infer per-appliance energy usage [37]. All these solutions are limited by the need for instrumentation to be

¹<http://www.greenbuttondata.org/>

²<https://opower.com/>

³<http://powerhouseanalytics.com/>

⁴<http://www.giminternational.com/products/p4400.html>

⁵<http://www.onsetcomp.com/products/data-loggers/ux120-018>

deployed in every home. Previous research has shown that there are various challenges beyond cost of the system that prevent such systems from being scaled [23].

NILM is often considered the most practical approach towards generating energy breakdown. However, like other approaches mentioned above, it also requires instrumentation in each home. While the effort of deploying a smart meter may be less in comparison to other related approaches, most of the world still does not have smart meter infrastructure.

Since its inception in the early 1980s by George Hart [21], the field of NILM has seen various approaches based on different machine learning techniques leveraging different features of the power trace. However, many of these approaches require submetered data to learn a model of each appliance and these models have not been shown to generalize well across homes. Even if the models did generalise, NILM approaches require high frequency power metering with resolutions of 1-minute or higher. Very high frequency approaches (>10 kHz) [11] use features such as voltage-current trajectories to detect events in aggregate power time series. However, current smart meters do not collect data at such high rates because they are designed and deployed for the purposes of time-of-use pricing and there are currently no efforts to deploy devices suitable for energy disaggregation on a large scale. Therefore, these techniques, while promising, face real practical barriers before being used at scale.

Additionally, existing approaches for energy disaggregation [33, 39, 21, 45, 35] require a model of each appliance. The main differences between these techniques are how they are created and how they are used to infer the hidden states of the appliances based on the aggregate power readings. For example, some systems model appliances as finite state machines (FSMs). However, such approaches generally show poor accuracy on complex appliances such as washing machine and other electronics, as a FSM is a poor model for such appliances. Some systems assume the model is manually generated, learned from training data [21, 33], and in rare cases learned automatically [5]. In all cases, however, the accuracy of these models depends on how well the model approximates the true appliances in the home and it has not yet been demonstrated that these model-based approaches generalize well across homes. Only recently, researchers have started looking into automatically learning arbitrarily complex appliances using deep neural networks [29]. However, as claimed by the authors themselves, the work is just scratching the surface. These challenges are likely to impact the generalisability and applicability of existing approaches towards solving the energy breakdown problem.

3. APPROACH

The goal of Gemello is to predict the energy consumption of household appliances given easy to collect information such as a single aggregate energy reading per month and static household characteristics. The key intuition behind Gemello is that if per appliance energy consumption is available for a home then any other home that is 'similar' in nature will also have the same energy consumption for that appliance. This intuition leads to the following specific requirements for Gemello to work:

1. Small set of submetered homes (SH) for which per appliance energy consumption is directly measured.
2. Easy to collect information for all the homes in the

back on an early version of this work at Sensys 2015 PhD forum. Arunchandar Vasam, Venkatesh Sarangan and Mario Berges provided useful suggestions. Jack Kelly, Oliver Parson, Mingjun Zhong, Manoj Gulati, Milan Jain provided detailed pre-submission reviews for our paper, in addition to deep technical discussions. Mingjun Zhong very kindly helped us use his LBM algorithm through Skype discussions and 100+ emails. Grant Fisher provided information about the dataset.

This work was funded in part by NSF awards 1305362, 1038271, and 0845761. The first author is supported by TCS Research and Development PhD fellowship. We also thank Department of Electronics and Information Technology (Government of India) for funding the project (Grant Number DeitY/R&D/ITEA/4(2)/2012 and Grant Number ITRA/15(57)/Mobile/HumanSense/01). We would also like to thank Google for providing support for the conference.

8. REFERENCES

- [1] J. Alcalá, O. Parson, and A. Rogers. Detecting anomalies in activities of daily living of elderly residents via energy disaggregation and cox processes. In *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*, pages 225-234. ACM, 2015.
 - [2] K. Anderson, A. Oconeau, D. Benitez, D. Carlson, A. Rowe, and M. Bergés. BLUEED: A fully labeled public dataset for Event-Based Non-Intrusive load monitoring research. In *Proceedings of 2nd KDD Workshop on Data Mining Applications in Sustainability*, pages 12-16, Beijing, China, 2012.
 - [3] P. Arjunan, H. D. Khadilkar, T. Gannu, Z. M. Charbiwala, A. Singh, and P. Singh. Multi-user energy consumption monitoring and anomaly detection with partial context information. In *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*, pages 35-44. ACM, 2015.
 - [4] K. C. Armel, A. Gupta, C. Shrivastava, and A. Albert. Is disaggregation the holy grail of energy efficiency? The case of electricity. *Energy Policy*, 52:213-234, 2013.
 - [5] S. Barker, S. Kalra, D. Irwin, and P. Shenoy. Empirical characterization and modeling of electrical loads in smart homes. In *IEEE International Green Computing Conference*, pages 1-10, Arlington, VA, USA, 2013.
 - [6] N. Batra, R. Bajjal, A. Singh, and K. Whitehouse. How good is good enough? re-evaluating the bar for energy disaggregation. *arXiv preprint arXiv:1510.08718*, 2015.
 - [7] N. Batra, M. Gulati, A. Singh, and M. B. Srivastava. It's Different: Insights into home energy consumption in India. In *Proceedings of the Fifth ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*, 2013.
 - [8] N. Batra, J. Kelly, O. Parson, H. Dutta, W. Knottenbelt, A. Rogers, A. Singh, and M. Srivastava. NILMTK: An Open Source Toolkit for Non-intrusive Load Monitoring. In *Fifth International Conference on Future Energy Systems (ACM e-Energy)*, Cambridge, UK, 2014.
 - [9] N. Batra, A. Singh, and K. Whitehouse. If you
- energy disaggregation. In *Proceedings of the second ACM International Conference on Embedded Systems For Energy-Efficient Built Environments*. ACM, 2015.
- [10] C. Beckel, W. Kleininger, R. Cicchetti, T. Staake, and S. Santini. The eco data set and the performance of non-intrusive load monitoring algorithms. In *Proceedings of the First ACM International Conference on Embedded Systems For Energy-Efficient Buildings*. ACM, 2014.
 - [11] M. E. Berges, E. Goldman, H. S. Matthews, and L. Soibelman. Enhancing electricity audits in residential buildings with nonintrusive load monitoring. *Journal of industrial ecology*, 14(5):844-858, 2010.
 - [12] K.-Y. Chen, S. Gupta, E. C. Larson, and S. Patel. Dose: Detecting user-driven operating states of electronic devices from a single sensing point. In *Pervasive Computing and Communications (PerCom), 2015 IEEE International Conference on*, pages 46-54. IEEE, 2015.
 - [13] M. Clark, B. Campbell, and P. Dutta. Deltalflow: submetering by synthesizing uncalibrated pulse sensor streams. In *Proceedings of the 5th international conference on Future energy systems*, pages 301-311. ACM, 2014.
 - [14] R. B. Cleveland, W. S. Cleveland, J. E. McRae, and I. Terpenning. Stl: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, 6(1):3-73, 1990.
 - [15] S. Darby. The effectiveness of feedback on the energy consumption. *A Review of DEFRA Literature on Metering, Billing and Direct Displays*, 2006.
 - [16] S. DeBruin, B. Ghena, Y.-S. Kuo, and P. Dutta. Powerblade: A low-profile, true-power, plug-through energy meter. In *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems*, pages 17-29. ACM, 2015.
 - [17] EIA. Household energy use in texas, 2014.
 - [18] M. Evans, B. Shui, and S. Somsundaram. *Country report on building energy codes in india*. Pacific Northwest National Laboratory, 2009.
 - [19] M. Gulati, S. S. Ram, A. Majumdar, and A. Singh. Single point conducted emi sensor with intelligent inference for detecting it appliances. In *IITD Technical Report*, 2015.
 - [20] S. Gupta, M. S. Reynolds, and S. N. Patel. Electrisense: single-point sensing using emi for electrical event detection and classification in the home. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*, pages 139-148. ACM, 2010.
 - [21] G. W. Hart. Nonintrusive appliance load monitoring. *Proceedings of the IEEE*, 80(12):1870-1891, 1992.
 - [22] T. Hastie, R. Tibshirani, J. Friedman, and J. Franklin. The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer*, 27(2):83-85, 2005.
 - [23] T. W. Hnat, V. Srivastava, J. Lu, T. I. Sookoor, R. Dawson, J. Stankovic, and K. Whitehouse. The hitchhiker's guide to successful residential sensing deployments. In *Proceedings of the 9th ACM*

Paper Structure

Title

Gemello: Creating a Detailed Energy Breakdown from just the Monthly Electricity Bill

Nipun Batra
IIT Delhi
nipunb@iitd.ac.in

Amarjeet Singh
IIT Delhi
amarjeet@iitd.ac.in

Kamin Whitehouse
University of Virginia
whitehouse@virginia.edu

ABSTRACT

The first step to saving energy in the home is often to create an *energy breakdown*: the amount of energy used by each individual appliance in the home. Unfortunately, current techniques that produce an energy breakdown are not *scalable*: they require hardware to be installed in each and every home. In this paper, we propose a more scalable solution called *Gemello* that estimates the energy breakdown for one home by matching it with similar homes for which the breakdown is already known. This matching requires only the monthly energy bill and household characteristics such as square footage of the home and the size of the household. We evaluate this approach using 57 homes and results indicate that the accuracy of Gemello is comparable to or better than existing techniques that use sensing infrastructure in each home. The information required by Gemello is often publicly available and, as such, it can be immediately applied to many homes around the world.

1. INTRODUCTION

Buildings account for more than 30% of total energy usage around the world, of which up to 93% is due to residential buildings [41, 18, 42]. Some of this energy can be saved by producing an *energy breakdown*: the amount of energy used by each individual appliance in the home, akin to the itemised bills we get from grocery stores. With such a breakdown, utility companies can focus conservation programs on homes that have an especially inefficient appliance, such as a fridge, water heater, or air conditioner. Energy feedback can also help induce energy-saving behaviour in the occupants themselves [15, 4, 30, 43, 24].

Unfortunately, current techniques that produce an energy breakdown are not *scalable*: they require hardware to be installed in each home and/or intensive manual training of the system. One approach, called non-intrusive load monitoring (NILM), tries to infer the energy breakdown based on the aggregate power trace from a single smart meter installed in the home. This approach is perhaps the most practi-

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
KDD '16, August 13 - 17, 2016, San Francisco, CA, USA
© 2016 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ISBN 978-1-4503-4232-2/16/08...\$15.00
DOI: <http://dx.doi.org/10.1145/2939672.2939735>

cal because smart meters are already being rolled out in millions of homes worldwide. However, current techniques require high resolution data (1 minute sampling frequency or higher) [39, 11, 4] while most smart meters today only support 15-minute or hourly sampling rates to support time-of-use energy pricing. Even with high resolution data, sampling rate, most of the work has been done in specialized sensors to be installed in each individual appliance [32, 37]. All of this requires the need for instrumentation to be installed in each home.

In this paper, we propose a new technique called *Gemello* that produces an energy breakdown in homes without requiring new hardware to be installed in each home. Instead, Gemello estimates the energy breakdown for a home by matching it with similar homes that do have a hardware-based disaggregation solution. This matching requires only monthly energy bill and household characteristics such as square footage of the home and the size of the household. From an energy perspective, homes in the same geographic region are often very similar because they have

ability to define 'similarity' on a per-appliance basis. For example, homes with similar seasonal trends in their monthly energy bill are expected to have similar heating/cooling energy (referred as heating, ventilation, and air conditioning [HVAC] from now on); homes with similar square footage are expected to have similar lighting loads; and homes with a similar number of occupants are expected to have similar washing machine energy usage because that is driven by the amount of clothes worn each day. The energy usage of each appliance is predicted by a different set of features, and so Gemello finds a different set of homes to predict the energy usage of each appliance.

We evaluate this approach using 57 homes from the publicly available Dataport data set [38], in which the ground truth energy breakdown is measured by metering each appliance of the home individually for one year or more. Results

show that the accuracy of Gemello is comparable to or better than established NILM techniques called FHMM and LBM, both of which require on-site, high-frequency power metering in each home [33, 45]. Gemello can achieve 76% and 69% accuracy on HVAC and fridge, compared to 61% and 39% for FHMM and 56% and 72% for LBM. Furthermore, it achieves up to 57% accuracy on washing machine, dryer, dishwasher, and lighting loads, which is higher than previously reported results. Many existing techniques are not able to disaggregate these loads at all. Our analysis shows that these results are robust with as few as 7 instrumented homes: the accuracy for HVAC loads is $\approx 69\%$, still 8% better than state-of-the-art approaches. The accuracy of Gemello becomes higher in homes with smart meters that provide power readings at 15-minute resolution.

The Gemello technique has potential for immediate impact because all of the information it requires is already available. Essentially all homes already have a standard power meter that provides total monthly energy usage for billing purposes, and in many regions this information can be downloaded online¹. Household characteristics such as the size of the home and the number of people in the household are often publicly available and are already being used by companies to match homes for other types of eco-feedback². Finally, many companies are already collecting submetering information from thousands of homes around the world³. By combining these three sources of information, Gemello could be used to provide an energy breakdown for homes around the world without the need to install new instrumentation in each location.

2. RELATED WORK

Various techniques for measuring appliance level energy consumption have been studied in the past. The simplest technique is to install appliance level sensors that monitor and report appliance energy consumption [26, 16]. Many commercial vendors are also selling appliance level sensors such as Kill-A-Watt⁴ and Hobo plug load data logger⁵. Instead of directly monitoring the appliance of interest, a few recent systems have looked into indirect sensing techniques. Kim et al. [32] develop a system called Viridiscopes that leverages the correlation amongst sensor streams, like using a vibration sensor on a fridge to tell if the compressor is running or not, and then using a model to determine fridge's power. Similarly, Clark et al. [13] develop a system called Deltalaw that employs energy harvesting sensors and performs computation on the activation of these sensors to determine appliance power draw. Jain et al. [25] install temperature sensors inside a home to estimate air conditioner energy usage. Gupta et al. [20], Chen et al. [12] and Gulati et al. [19] use the electromagnetic interference typically generated by electronic appliances to determine appliance usages. Previous work has also looked into using sensors deployed at household circuits (lesser in number than individual appliances) to infer per-appliance energy usage [37]. All these solutions are limited by the need for instrumentation to be

¹<http://www.greenbuttondata.org/>
²<https://opower.com/>
³<http://powerhouseanalytics.com/>
⁴<http://www.giinterational.com/products/p4400.html>
⁵<http://www.onsetcomp.com/products/data-loggers/ux120-018>

deployed in every home. Previous research has shown that there are various challenges beyond cost of the system that prevent such systems from being scaled [23].

NILM is often considered the most practical approach towards generating energy breakdown. However, like other approaches mentioned above, it also requires instrumentation in each home. While the effort of deploying a smart meter may be less in comparison to other related approaches, most of the world still does not have smart meter infrastructure.

Since its inception in the early 1980s by George Hart [21], the field of NILM has seen various approaches based on different machine learning techniques leveraging different features of the power trace. However, many of these approaches require submetered data to learn a model of each appliance and these models have not been shown to generalize well across homes. Even if the models did generalise, NILM approaches require high frequency power metering with resolutions of 1-minute or higher. Very high frequency approaches (>10 kHz) [11] use features such as voltage-current trajectories to detect events in aggregate power time series. However, current smart meters do not collect data at such high rates because they are designed and deployed for the purposes of time-of-use pricing and there are currently no efforts to deploy devices suitable for energy disaggregation on a large scale. Therefore, these techniques, while promising, face real practical barriers before being used at scale.

Additionally, existing approaches for energy disaggregation [33, 39, 21, 45, 35] require a model of each appliance. The main differences between these techniques are how they are created and how they are used to infer the hidden states of the appliances based on the aggregate power readings. For example, some systems model appliances as finite state machines (FSMs). However, such approaches generally show poor accuracy on complex appliances such as washing machine and other electronics, as a FSM is a poor model for such appliances. Some systems assume the model is manually generated, learned from training data [21, 33], and in rare cases learned automatically [5]. In all cases, however, the accuracy of these models depends on how well the model approximates the true appliances in the home and it has not yet been demonstrated that these model-based approaches generalize well across homes. Only recently, researchers have started looking into automatically learning arbitrarily complex appliances using deep neural networks [29]. However, as claimed by the authors themselves, the work is just scratching the surface. These challenges are likely to impact the generalisability and applicability of existing approaches towards solving the energy breakdown problem.

3. APPROACH

The goal of Gemello is to predict the energy consumption of household appliances given easy to collect information such as a single aggregate energy reading per month and static household characteristics. The key intuition behind Gemello is that if per appliance energy consumption is available for a home then any other home that is 'similar' in nature will also have the same energy consumption for that appliance. This intuition leads to the following specific requirements for Gemello to work:

1. Small set of submetered homes (SH) for which per appliance energy consumption is directly measured.
2. Easy to collect information for all the homes in the

back on an early version of this work at Sensys 2015 PhD forum. Arunchandhar Vasani, Venkatesh Sarangan and Mario Berges provided useful suggestions. Jack Kelly, Oliver Parson, Mingjun Zhong, Manoj Gulati, Milan Jain provided detailed pre-submission reviews for our paper, in addition to deep technical discussions. Mingjun Zhong very kindly helped us use his LBM algorithm through Skype discussions and 100+ emails. Grant Fisher provided information about the dataset.

This work was funded in part by NSF awards 1305362, 1038271, and 0845761. The first author is supported by TCS Research and Development PhD fellowship. We also thank Department of Electronics and Information Technology (Government of India) for funding the project (Grant Number DeitY/R&D/ITEA/4(2)/2012 and Grant Number ITRA/15(57)/Mobile/HumanSense/01). We would also like to thank Google for providing support for the conference.

8. REFERENCES

- [1] J. Alcalá, O. Parson, and A. Rogers. Detecting anomalies in activities of daily living of elderly residents via energy disaggregation and cox processes. In *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*, pages 225-234. ACM, 2015.
 - [2] K. Anderson, A. Oconeau, D. Benitez, D. Carlson, A. Rowe, and M. Bergés. BLUEED: A fully labeled public dataset for Event-Based Non-Intrusive load monitoring research. In *Proceedings of 2nd KDD Workshop on Data Mining Applications in Sustainability*, pages 12-16, Beijing, China, 2012.
 - [3] P. Arjunan, H. D. Khadilkar, T. Gannu, Z. M. Charbiwala, A. Singh, and P. Singh. Multi-user energy consumption monitoring and anomaly detection with partial context information. In *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*, pages 35-44. ACM, 2015.
 - [4] K. C. Armel, A. Gupta, C. Shrivastava, and A. Albert. Is disaggregation the holy grail of energy efficiency? The case of electricity. *Energy Policy*, 52:213-234, 2013.
 - [5] S. Barker, S. Kalra, D. Irwin, and P. Shenoy. Empirical characterization and modeling of electrical loads in smart homes. In *IEEE International Green Computing Conference*, pages 1-10, Arlington, VA, USA, 2013.
 - [6] N. Batra, R. Bajjal, A. Singh, and K. Whitehouse. How good is good enough? re-evaluating the bar for energy disaggregation. *arXiv preprint arXiv:1510.08718*, 2015.
 - [7] N. Batra, M. Gulati, A. Singh, and M. B. Srivastava. It's Different: Insights into home energy consumption in India. In *Proceedings of the Fifth ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*, 2013.
 - [8] N. Batra, J. Kelly, O. Parson, H. Dutta, W. Knottenbelt, A. Rogers, A. Singh, and M. Srivastava. NILMTK: An Open Source Toolkit for Non-intrusive Load Monitoring. In *Fifth International Conference on Future Energy Systems (ACM e-Energy)*, Cambridge, UK, 2014.
 - [9] N. Batra, A. Singh, and K. Whitehouse. If you
- energy disaggregation. In *Proceedings of the second ACM International Conference on Embedded Systems For Energy-Efficient Built Environments*. ACM, 2015.
- [10] C. Beckel, W. Kleininger, R. Cicchetti, T. Stauke, and S. Santini. The eco data set and the performance of non-intrusive load monitoring algorithms. In *Proceedings of the First ACM International Conference on Embedded Systems For Energy-Efficient Buildings*. ACM, 2014.
 - [11] M. E. Berges, E. Goldman, H. S. Matthews, and L. Soibelman. Enhancing electricity audits in residential buildings with nonintrusive load monitoring. *Journal of industrial ecology*, 14(5):844-858, 2010.
 - [12] K.-Y. Chen, S. Gupta, E. C. Larson, and S. Patel. Dose: Detecting user-driven operating states of electronic devices from a single sensing point. In *Pervasive Computing and Communications (PerCom), 2015 IEEE International Conference on*, pages 46-54. IEEE, 2015.
 - [13] M. Clark, B. Campbell, and P. Dutta. Deltalaw: submetering by synthesizing uncalibrated pulse sensor streams. In *Proceedings of the 5th international conference on Future energy systems*, pages 301-311. ACM, 2014.
 - [14] R. B. Cleveland, W. S. Cleveland, J. E. McRae, and I. Terpenning. Stl: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, 6(1):3-73, 1990.
 - [15] S. Darby. The effectiveness of feedback on the energy consumption. *A Review for DEFRA on Literature on Metering, Billing and Direct Displays*, 2006.
 - [16] S. DeBruin, B. Ghena, Y.-S. Kuo, and P. Dutta. Powerblade: A low-profile, true-power, plug-through energy meter. In *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems*, pages 17-29. ACM, 2015.
 - [17] EIA. Household energy use in texas, 2014.
 - [18] M. Evans, B. Shui, and S. Somsundaram. *Country report on building energy codes in india*. Pacific Northwest National Laboratory, 2009.
 - [19] M. Gulati, S. S. Ram, A. Majumdar, and A. Singh. Single point conducted emi sensor with intelligent inference for detecting it appliances. In *IITD Technical Report*, 2015.
 - [20] S. Gupta, M. S. Reynolds, and S. N. Patel. Electrisense: single-point sensing using emi for electrical event detection and classification in the home. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*, pages 139-148. ACM, 2010.
 - [21] G. W. Hart. Nonintrusive appliance load monitoring. *Proceedings of the IEEE*, 80(12):1870-1891, 1992.
 - [22] T. Hastie, R. Tibshirani, J. Friedman, and J. Franklin. The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer*, 27(2):83-85, 2005.
 - [23] T. W. Hnat, V. Srivastava, J. Lu, T. I. Sookoor, R. Dawson, J. Stankovic, and K. Whitehouse. The hitchhiker's guide to successful residential sensing deployments. In *Proceedings of the 9th ACM*

Paper Structure

Title

Gemello: Creating a Detailed Energy Breakdown from just the Monthly Electricity Bill

Nipun Batra
IIT Delhi
nipunb@iitd.ac.in

Amarjeet Singh
IIT Delhi
amarjeet@iitd.ac.in

Kamin Whitehouse
University of Virginia
whitehouse@virginia.edu

ABSTRACT

The first step to saving energy in the home is often to create an *energy breakdown*: the amount of energy used by each individual appliance in the home. Unfortunately, current techniques that produce an energy breakdown are not *scalable*: they require hardware to be installed in each and every home. In this paper, we propose a more scalable solution called *Gemello* that estimates the energy breakdown for one home by matching it with similar homes for which the breakdown is already known. This matching requires only the monthly energy bill and household characteristics such as square footage of the home and the size of the household. We evaluate this approach using 57 homes and results indicate that the accuracy of Gemello is comparable to or better than existing techniques that use sensing infrastructure in each home. The information required by Gemello is often publicly available and, as such, it can be immediately applied to many homes around the world.

1. INTRODUCTION

Buildings account for more than 30% of total energy usage around the world, of which up to 93% is due to residential buildings [41, 18, 42]. Some of this energy can be saved by producing an *energy breakdown*: the amount of energy used by each individual appliance in the home, akin to the itemised bills we get from grocery stores. With such a breakdown, utility companies can focus conservation programs on homes that have an especially inefficient appliance, such as a fridge, water heater, or air conditioner. Energy feedback can also help induce energy-saving behaviour in the occupants themselves [15, 4, 30, 43, 24].

Unfortunately, current techniques that produce an energy breakdown are not *scalable*: they require hardware to be installed in each home and/or intensive manual training of the system. One approach, called non-intrusive load monitoring (NILM), tries to infer the energy breakdown based on the aggregate power trace from a single smart meter installed in the home. This approach is perhaps the most practical

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

KDD '16, August 13 - 17, 2016, San Francisco, CA, USA
© 2016 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ISBN 978-1-4503-4232-2/16/08...\$15.00
DOI: <http://dx.doi.org/10.1145/2939672.2939735>

cal because smart meters are already being rolled out in millions of homes worldwide. However, current techniques require high resolution data (1 minute sampling frequency or higher) [39, 11, 4] while most smart meters today only support 15-minute or hourly sampling rates to support time-of-use energy pricing. Even if the sampling rate, most of the smart meters and many places do not have them. Alternatives to NILM are specialized sensors to be installed on each individual appliance [32], or the breaker box [37]. All of these require the need for instrumentation to be installed in each home.

In this paper, we propose a novel technique called *Gemello* that produces an energy breakdown in homes without requiring new hardware to be installed in each home. Instead, Gemello estimates the energy breakdown for a home by matching it with similar homes that do have a hardware-based disaggregation solution. This matching requires only the monthly energy bill and household characteristics such as square footage of the home and the size of the household. From an energy perspective, homes in the same geographic region are often very similar because they have

Abstract

ability to define 'similarity' on a per-appliance basis. For example, homes with similar seasonal trends in their monthly energy bill are expected to have similar heating/cooling energy (referred as heating, ventilation, and air conditioning [HVAC] from now on); homes with similar square footage are expected to have similar lighting loads; and homes with a similar number of occupants are expected to have similar washing machine energy usage because that is driven by the amount of clothes worn each day. The energy usage of each appliance is predicted by a different set of features, and so Gemello finds a different set of homes to predict the energy usage of each appliance.

We evaluate this approach using 57 homes from the publicly available Dataport data set [38], in which the ground truth energy breakdown is measured by metering each appliance of the home individually for one year or more. Results

Authors

Sections

- Introduction
- Related Work
- Approach
- Evaluation
- Limitations & Future work
- Conclusions

show that the accuracy of Gemello is comparable to or better than established NILM approaches. The accuracy of Gemello is 76% and 69% for FHM and 56% and 72% for LBM. Furthermore, it achieves up to 57% accuracy on washing machine, dryer, dishwasher, and lighting loads, which is higher than previously reported results. Many existing techniques are not able to disaggregate these loads at all. Our analysis shows that these results are robust with as few as 7 instrumented homes: the accuracy for HVAC loads is $\approx 69\%$, still 8% better than state-of-the-art approaches. The accuracy of Gemello becomes higher in homes with smart meters that provide power readings at 15-minute resolution.

The Gemello technique has potential for immediate impact because all of the information it requires is already available. Essentially all homes already have a standard power meter that provides total monthly energy usage for billing purposes, and in many regions this information can be downloaded online¹. Household characteristics such as the size of the home and the number of people in the household are often publicly available and are already being used by companies to match homes for other types of eco-feedback². Finally, many companies are already collecting submetering information from thousands of homes around the world³. By combining these three sources of information, Gemello could be used to provide an energy breakdown for homes around the world without the need to install new instrumentation in each location.

2. RELATED WORK

Various techniques for measuring appliance level energy consumption have been studied in the past. The simplest technique is to install appliance level sensors that monitor and report appliance energy consumption [26, 16]. Many commercial vendors are also selling appliance level sensors such as Kill-A-Watt⁴ and Hobo plug load data logger⁵. Instead of directly monitoring the appliance of interest, a few recent systems have looked into indirect sensing techniques. Kim et al. [32] develop a system called Viridiscopes that leverages the correlation amongst sensor streams, like using a vibration sensor on a fridge to tell if the compressor is running or not, and then using a model to determine the fridge's power. Similarly, Clark et al. [18] develop a system called Deltaloc that employs energy harvesting sensors and performs computation on the activation of these sensors to determine appliance power draw. Jain et al. [25] install temperature sensors inside a home to estimate air conditioner energy usage. Gupta et al. [20], Chen et al. [12] and Gulati et al. [19] use the electromagnetic interference typically generated by electronic appliances to determine appliance usages. Previous work has also looked into using sensors deployed at household circuits (less in number than individual appliances) to infer per-appliance energy usage [37]. All these solutions are limited by the need for instrumentation to be

¹<http://www.greenbuttondata.org/>
²<https://power.com/>
³<http://powerhouseanalytics.com/>
⁴<http://www.giinternational.com/products/p4400.html>
⁵<http://www.onsetcomp.com/products/data-loggers/ux120-018>

prevent such systems from being scaled [42].

NILM has been considered the most practical approach towards generating energy breakdown. However, like other approaches mentioned above, it also requires instrumentation in each home. While the effort of deploying a smart meter may be less than comparison to other related approaches, most of the smart meters do not have smart meter infrastructure.

Since its inception in the early 1980s by George Hart [21], the field of NILM has seen various approaches based on different machine learning techniques leveraging different features of the power trace. However, many of these approaches require submetered data to learn a model of each appliance and these models have not been shown to generalize well across homes. Even if the models did generalise, NILM approaches require high frequency power metering with resolutions of 1-minute or higher. Very high frequency approaches (>10 kHz) [11] use features such as voltage-current trajectories to detect events in aggregate power time series. However, current smart meters do not collect data at such high rates because they are designed and deployed for the purposes of time-of-use pricing and there are currently no efforts to deploy devices suitable for energy disaggregation on a large scale. Therefore, these techniques, while promising, face real practical barriers before being used at scale.

Additionally, existing approaches for energy disaggregation [33, 39, 21, 45, 35] require a model of each appliance. The main differences between these techniques are how they are created and how they are used to infer the hidden states of the appliances based on the aggregate power readings. For example, some systems model appliances as finite state machines (FSMs). However, such approaches generally show poor accuracy on complex appliances such as washing machine and other electronics, as a FSM is a poor model for such appliances. Some systems assume the model is manually generated, learned from training data [21, 33], and in rare cases learned automatically [5]. In all cases, however, the accuracy of these models depends on how well the model approximates the true appliances in the home and it has not yet been demonstrated that these model-based approaches generalize well across homes. Only recently, researchers have started looking into automatically learning arbitrarily complex appliances using deep neural networks [29]. However, as claimed by the authors themselves, the work is just scratching the surface. These challenges are likely to impact the generalisability and applicability of existing approaches towards solving the energy breakdown problem.

3. APPROACH

The goal of Gemello is to predict the energy consumption of household appliances given easy to collect information such as a single aggregate energy reading per month and static household characteristics. The key intuition behind Gemello is that if per appliance energy consumption is available for a home then any other home that is 'similar' in nature will also have the same energy consumption for that appliance. This intuition leads to the following specific requirements for Gemello to work:

1. Small set of submetered homes (SH) for which per appliance energy consumption is directly measured.
2. Easy to collect information for all the homes in the

Berges provided useful suggestions. Jack Kelly, Oliver Parson, Mingjun Zhong, Manoj Gulati, Milan Jain provided detailed pre-submission reviews for our paper, in addition to deep technical discussions. Mingjun Zhong very kindly helped us use his LBM algorithm through Skype discussions and 100+ emails. Grant Fisher provided information about the dataset.

This work was funded in part by NSF awards 1305362, 1038271, and 0845761. The first author is supported by TCS Research and Development PhD fellowship. We also thank Department of Electronics and Information Technology (Government of India) for funding the project (Grant Number ITRA/15(57)/Mobile/HumanSense/01). We would also like to thank Google for providing support for the conference.

8. REFERENCES

- [1] J. Alcalá, O. Parson, and A. Rogers. Detecting anomalies in activities of daily living of elderly residents via energy disaggregation and cox processes. In *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*, pages 225-234. ACM, 2015.
- [2] K. Anderson, A. Oconeau, D. Benitez, D. Carlson, A. Rowe, and M. Bergés. BLUEED: A fully labeled public dataset for Event-Based Non-Intrusive load monitoring research. In *Proceedings of 2nd KDD Workshop on Data Mining Applications in Sustainability*, pages 12-16, Beijing, China, 2012.
- [3] P. Arjunan, H. D. Khadilkar, T. Gann, Z. M. Charbiwala, A. Singh, and P. Singh. Multi-user energy consumption monitoring and anomaly detection with partial context information. In *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*, pages 35-44. ACM, 2015.
- [4] K. C. Armel, A. Gupta, C. Shrivastava, and A. Albert. Is disaggregation the holy grail of energy efficiency? The case of electricity. *Energy Policy*, 52:213-234, 2013.
- [5] S. Barker, S. Kalra, D. Irwin, and P. Shenoy. Empirical characterization and modeling of electrical loads in smart homes. In *IEEE International Green Computing Conference*, pages 1-10, Arlington, VA, USA, 2013.
- [6] N. Batra, R. Bajjal, A. Singh, and K. Whitehouse. How good is good enough? re-evaluating the bar for energy disaggregation. *arXiv preprint arXiv:1510.08718*, 2015.
- [7] N. Batra, M. Gulati, A. Singh, and M. B. Srivastava. It's Different: Insights into home energy consumption in India. In *Proceedings of the Fifth ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*, 2013.
- [8] N. Batra, J. Kelly, O. Parson, H. Dutta, W. Knottenbelt, A. Rogers, A. Singh, and M. Srivastava. NILMTK: An Open Source Toolkit for Non-intrusive Load Monitoring. In *Fifth International Conference on Future Energy Systems (ACM e-Energy)*, Cambridge, UK, 2014.
- [9] N. Batra, A. Singh, and K. Whitehouse. If you

energy disaggregation. In *Proceedings of the second ACM International Conference on Embedded Systems For Energy-Efficient Built Environments*. ACM, 2015.

- [10] C. Beckel, W. Kleininger, R. Cicchetti, T. Staake, and S. Santini. The eco data set and the performance of non-intrusive load monitoring algorithms. In *Proceedings of the First ACM International Conference on Embedded Systems For Energy-Efficient Buildings*. ACM, 2014.
- [11] M. E. Berges, E. Goldman, H. S. Matthews, and L. Soibelman. Enhancing electricity audits in residential buildings with nonintrusive load monitoring. *Journal of industrial ecology*, 14(5):844-858, 2010.
- [12] K.-Y. Chen, S. Gupta, E. C. Larson, and S. Patel. Dose: Detecting user-driven operating states of electronic devices from a single sensing point. In *Pervasive Computing and Communications (PerCom), 2015 IEEE International Conference on*, pages 46-54. IEEE, 2015.
- [13] M. Clark, B. Campbell, and P. Dutta. Deltaloc: submetering by synthesizing uncalibrated pulse sensor streams. In *Proceedings of the 5th international conference on Future energy systems*, pages 301-311. ACM, 2014.
- [14] R. B. Cleveland, W. S. Cleveland, J. E. McRae, and I. Terpenning. Stl: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, 6(1):3-73, 1990.
- [15] S. Darby. The effectiveness of feedback on energy consumption. *A Review of DEFRA Literature on Metering, Billing and End Displays*, 2006.
- [16] S. DeBruin, B. Ghena, Y.-S. Kuo, and P. Dutta. Powerblade: A low-profile, true-power, plug-through energy meter. In *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems*, pages 17-29. ACM, 2015.
- [17] EIA. Household energy use in texas, 2014.
- [18] M. Evans, B. Shui, and S. Somsundaram. Country report on building energy codes in india. Pacific Northwest National Laboratory, 2009.
- [19] M. Gulati, S. S. Ram, A. Majumdar, and A. Singh. Single point conducted emi sensor with intelligent inference for detecting it appliances. In *IITD Technical Report*, 2015.
- [20] S. Gupta, M. S. Reynolds, and S. N. Patel. Electrisense: single-point sensing using emi for electrical event detection and classification in the home. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*, pages 139-148. ACM, 2010.
- [21] G. W. Hart. Nonintrusive appliance load monitoring. *Proceedings of the IEEE*, 80(12):1870-1891, 1992.
- [22] T. Hastie, R. Tibshirani, J. Friedman, and J. Franklin. The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer*, 27(2):83-85, 2005.
- [23] T. W. Hnat, V. Srivastava, J. Lu, T. I. Sookoor, R. Dawson, J. Stankovic, and K. Whitehouse. The hitchhiker's guide to successful residential sensing deployments. In *Proceedings of the 9th ACM*

Paper Structure

Title

Gemello: Creating a Detailed Energy Breakdown from just the Monthly Electricity Bill

Nipun Batra
IIT Delhi
nipunb@iitd.ac.in

Amarjeet Singh
IIT Delhi
amarjeet@iitd.ac.in

Kamin Whitehouse
University of Virginia
whitehouse@virginia.edu

ABSTRACT

The first step to saving energy in the home is often to create an *energy breakdown*: the amount of energy used by each individual appliance in the home. Unfortunately, current techniques that produce an energy breakdown are not *scalable*: they require hardware to be installed in each and every home. In this paper, we propose a more scalable solution called *Gemello* that estimates the energy breakdown for one home by matching it with similar homes for which the breakdown is already known. This matching requires only the monthly energy bill and household characteristics such as square footage of the home and the size of the household. We evaluate this approach using 57 homes and results indicate that the accuracy of Gemello is comparable to or better than existing techniques that use sensing infrastructure in each home. The information required by Gemello is often publicly available and, as such, it can be immediately applied to many homes around the world.

1. INTRODUCTION

Buildings account for more than 30% of total energy usage around the world, of which up to 93% is due to residential buildings [41, 18, 42]. Some of this energy can be saved by producing an *energy breakdown*: the amount of energy used by each individual appliance in the home, akin to the itemised bills we get from grocery stores. With such a breakdown, utility companies can focus conservation programs on homes that have an especially inefficient appliance, such as a fridge, water heater, or air conditioner. Energy feedback can also help induce energy-saving behaviour in the occupants themselves [15, 4, 30, 43, 24].

Unfortunately, current techniques that produce an energy breakdown are not *scalable*: they require hardware to be installed in each home and/or intensive manual training of the system. One approach, called non-intrusive load monitoring (NILM), tries to infer the energy breakdown based on the aggregate power trace from a single smart meter installed in the home. This approach is perhaps the most practical

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

KDD '16, August 13 - 17, 2016, San Francisco, CA, USA

© 2016 Copyright held by the owner/authors. Publication rights licensed to ACM. ISBN 978-1-4503-4232-2/16/08...\$15.00
DOI: <http://dx.doi.org/10.1145/2939672.2939735>

cal because smart meters are already being rolled out in millions of homes worldwide. However, current techniques require high resolution data (1 minute sampling frequency or higher) [39, 11, 4] while most smart meters today only support 15-minute or hourly sampling rates to support time-of-use energy pricing. Even if the sampling rate, most of the smart meters and many places do not have them. Alternatives to NILM are specialized sensors to be installed on each individual appliance [32], or the breaker box [37]. All of these require the need for instrumentation to be installed in each home.

In this paper, we propose a novel technique called *Gemello* that produces an energy breakdown in homes without requiring new hardware to be installed in each home. Instead, Gemello estimates the energy breakdown for a home by matching it with similar homes that do have a hardware-based disaggregation solution. This matching requires only monthly energy bill and household characteristics such as square footage of the home and the size of the household. From an energy perspective, homes in the same geographic region are often very similar because they have

Abstract

ability to define 'similarity' on a per-appliance basis. For example, homes with similar seasonal trends in their monthly energy bill are expected to have similar heating/cooling energy (referred as heating, ventilation, and air conditioning [HVAC] from now on); homes with similar square footage are expected to have similar lighting loads; and homes with a similar number of occupants are expected to have similar washing machine energy usage because that is driven by the amount of clothes worn each day. The energy usage of each appliance is predicted by a different set of features, and so Gemello finds a different set of homes to predict the energy usage of each appliance.

We evaluate this approach using 57 homes from the publicly available Dataport data set [38], in which the ground truth energy breakdown is measured by metering each appliance of the home individually for one year or more. Results

Authors

Sections

- Introduction
- Related Work
- Approach
- Evaluation
- Limitations & Future work
- Conclusions

show that the accuracy of

ter than established NILM LBM, both of which require on-site, high-frequency power metering in each home [33, 45]. Gemello can achieve 76% and 69% accuracy on HVAC and fridge, compared to 61% and 39% for FHMM and 56% and 72% for LBM. Furthermore, it achieves up to 57% accuracy on washing machine, dryer, dishwasher, and lighting loads, which is higher than previously reported results. Many existing techniques are not able to disaggregate these loads at all. Our analysis shows that these results are robust with as few as 7 instrumented homes: the accuracy for HVAC loads is $\approx 69\%$, still 8% better than state-of-the-art approaches. The accuracy of Gemello becomes higher in homes with smart meters that provide power readings at 15-minute resolution.

The Gemello technique has potential for immediate impact because all of the information it requires is already available. Essentially all homes already have a standard power meter that provides total monthly energy usage for billing purposes, and in many regions this information can be downloaded online¹. Household characteristics such as the size of the home and the number of people in the household are often publicly available and are already being used by companies to match homes for other types of eco-feedback². Finally, many companies are already collecting submetering information from thousands of homes around the world³. By combining these three sources of information, Gemello could be used to provide an energy breakdown for homes around the world without the need to install new instrumentation in each location.

2. RELATED WORK

Various techniques for measuring appliance level energy consumption have been studied in the past. The simplest technique is to install appliance level sensors that monitor and report appliance energy consumption [26, 16]. Many commercial vendors are also selling appliance level sensors such as Kill-A-Watt⁴ and Hobo plug load data logger⁵. Instead of directly monitoring the appliance of interest, a few recent systems have looked into indirect sensing techniques. Kim et al. [32] develop a system called Viridiscopes that leverages the correlation amongst sensor streams, like using a vibration sensor on a fridge to tell if the compressor is running or not, and then using a model to determine fridge's power. Similarly, Clark et al. [18] develop a system called Deltalaw that employs energy harvesting sensors and performs computation on the activation of these sensors to determine appliance power draw. Jain et al. [25] install temperature sensors inside a home to estimate air conditioner energy usage. Gupta et al. [20], Chen et al. [12] and Gulati et al. [19] use the electromagnetic interference typically generated by electronic appliances to determine appliance usages. Previous work has also looked into using sensors deployed at household circuits (lesser in number than individual appliances) to infer per-appliance energy usage [37]. All these solutions are limited by the need for instrumentation to be

¹<http://www.greenbuttondata.org/>
²<https://power.com/>
³<http://powerhouseanalytics.com/>
⁴<http://www.giinternational.com/products/p4400.html>
⁵<http://www.onsetcomp.com/products/data-loggers/ux120-018>

prevent such systems from being scaled [42].

NILM is often considered the most practical approach towards generating energy breakdown. However, like other approaches mentioned above, it also requires instrumentation in each home. While the effort of deploying a smart meter may be less than comparison to other related approaches, most of the work [11] does not have smart meter infrastructure.

Since its inception in the early 1980s by George Hart [21], the field of NILM has seen various approaches based on different machine learning techniques leveraging different features of the power trace. However, many of these approaches require submetered data to learn a model of each appliance and these models have not been shown to generalize well across homes. Even if the models did generalise, NILM approaches require high frequency power metering with resolutions of 1-minute or higher. Very high frequency approaches (>10 kHz) [11] use features such as voltage-current trajectories to detect events in aggregate power time series. However, current smart meters do not collect data at such high rates because they are designed and deployed for the purposes of time-of-use pricing and there are currently no efforts to deploy devices suitable for energy disaggregation on a large scale. Therefore, these techniques, while promising, face real practical barriers before being used at scale.

Additionally, existing approaches for energy disaggregation [33, 39, 21, 45, 35] require a model of each appliance. The main differences between these techniques are how they are created and how they are used to infer the hidden states of the appliances based on the aggregate power readings. For example, some systems model appliances as finite state machines (FSMs). However, such approaches generally show poor accuracy on complex appliances such as washing machine and other electronics, as a FSM is a poor model for such appliances. Some systems assume the model is manually generated, learned from training data [21, 33], and in rare cases learned automatically [5]. In all cases, however, the accuracy of these models depends on how well the model approximates the true appliances in the home and it has not yet been demonstrated that these model-based approaches generalize well across homes. Only recently, researchers have started looking into automatically learning arbitrarily complex appliances using deep neural networks [29]. However, as claimed by the authors themselves, the work is just scratching the surface. These challenges are likely to impact the generalisability and applicability of existing approaches towards solving the energy breakdown problem.

3. APPROACH

The goal of Gemello is to predict the energy consumption of household appliances given easy to collect information such as a single aggregate energy reading per month and static household characteristics. The key intuition behind Gemello is that if per appliance energy consumption is available for a home then any other home that is 'similar' in nature will also have the same energy consumption for that appliance. This intuition leads to the following specific requirements for Gemello to work:

1. Small set of submetered homes (SH) for which per appliance energy consumption is directly measured.
2. Easy to collect information for all the homes in the

Berges provided useful suggestions. Jack Kelly, Oliver Parson, Mingjun Zhong, Manoj Gulati, Milan Jain provided detailed pre-submission reviews for our paper, in addition to deep technical discussions. Mingjun Zhong very kindly helped us use his LBM algorithm through Skype discussions and 100+ emails. Grant Fisher provided information about the dataset.

This work was funded in part by NSF awards 1305362, 1038271, and 0845761. The first author is supported by TCS Research and Development PhD fellowship. We also thank Department of Electronics and Information Technology (Government of India) for funding the project (Grant Number Dst/Y/R&D/ITEA/4(2)/2012 and Grant Number ITRA/15(57)/Mobile/HumanSense/01). We would also like to thank Google for providing support for the conference.

8. REFERENCES

- [1] J. Alcalá, O. Parson, and A. Rogers. Detecting anomalies in activities of daily living of elderly residents via energy disaggregation and cox processes. In *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*, pages 225-234. ACM, 2015.
 - [2] K. Anderson, A. Oconeau, D. Benitez, D. Carlson, A. Rowe, and M. Bergés. BLUEED: A fully labeled public dataset for Event-Based Non-Intrusive Load monitoring research. In *Proceedings of 2nd KDD Workshop on Data Mining Applications in Sustainability*, pages 12-16, Beijing, China, 2012.
 - [3] P. Arjunan, H. D. Khadilkar, T. Gamo, Z. M. Charbiwala, A. Singh, and P. Singh. Multi-user energy consumption monitoring and anomaly detection with partial context information. In *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*, pages 35-44. ACM, 2015.
 - [4] K. C. Arred, A. Gupta, C. Shirmali, and A. Albert. Is disaggregation the holy grail of energy efficiency? The case of electricity. *Energy Policy*, 52:213-234, 2013.
 - [5] S. Barker, S. Kalra, D. Irwin, and P. Shenoy. Empirical characterization and modeling of electrical loads in smart homes. In *IEEE International Green Computing Conference*, pages 1-10, Arlington, VA, USA, 2013.
 - [6] N. Batra, R. Bajjal, A. Singh, and K. Whitehouse. How good is good enough? re-evaluating the bar for energy disaggregation. *arXiv preprint arXiv:1510.08719*, 2015.
 - [7] N. Batra, M. Gulati, A. Singh, and M. B. Srivastava. It's Different: Insights into home energy consumption in India. In *Proceedings of the Fifth ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*, 2013.
 - [8] N. Batra, J. Kelly, O. Parson, H. Dutta, W. Knottmehel, A. Rogers, A. Singh, and M. Srivastava. NILMTK: An Open Source Toolkit for Non-intrusive Load Monitoring. In *Fifth International Conference on Future Energy Systems (ACM e-Energy)*, Cambridge, UK, 2014.
 - [9] N. Batra, A. Singh, and K. Whitehouse. If you
- energy disaggregation. In *Proceedings of the second ACM International Conference on Embedded Systems For Energy-Efficient Built Environments*. ACM, 2015.
- [10] C. Beckel, W. Kleininger, R. Cicchetti, T. Staake, and S. Santini. The eco data set and the performance of non-intrusive load monitoring algorithms. In *Proceedings of the First ACM International Conference on Embedded Systems For Energy-Efficient Buildings*. ACM, 2014.
 - [11] M. E. Berges, E. Goldman, H. S. Matthews, and L. Soibelman. Enhancing electricity audits in residential buildings with nonintrusive load monitoring. *Journal of industrial ecology*, 14(5):844-858, 2010.
 - [12] K.-Y. Chen, S. Gupta, E. C. Larson, and S. Patel. Dose: Detecting user-driven operating states of electronic devices from a single sensing point. In *Pervasive Computing and Communications (PerCom), 2015 IEEE International Conference on*, pages 46-54. IEEE, 2015.
 - [13] M. Clark, B. Campbell, and P. Dutta. Deltalaw: submetering by synthesizing uncalibrated pulse sensor streams. In *Proceedings of the 5th international conference on Future energy systems*, pages 301-311. ACM, 2014.
 - [14] R. B. Cleveland, W. S. Cleveland, J. E. McRae, and I. Terpenning. Stl: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, 6(1):3-73, 1990.
 - [15] S. Darby. The effectiveness of feedback on the energy consumption. *A Review for DEFRA on Literature on Metering, Billing and End Displays*, 2006.
 - [16] S. DeBruin, B. Ghena, Y.-S. Kuo, and P. Dutta. Powerblader: A low-profile, true-power, plug-through energy meter. In *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems*, pages 17-29. ACM, 2015.
 - [17] EIA. Household energy use in texas, 2014.
 - [18] M. Evans, B. Shui, and S. Somsundaram. Country report on building energy codes in india. Pacific Northwest National Laboratory, 2009.
 - [19] M. Gulati, S. S. Ram, A. Majumdar, and A. Singh. Single point conducted emi sensor with intelligent inference for detecting it appliances. In *IITD Technical Report*, 2015.
 - [20] S. Gupta, M. S. Reynolds, and S. N. Patel. Electrisense: single-point sensing using emi for electrical event detection and classification in the home. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*, pages 139-148. ACM, 2010.
 - [21] G. W. Hart. Nonintrusive appliance load monitoring. *Proceedings of the IEEE*, 80(12):1870-1891, 1992.
 - [22] T. Hastie, R. Tibshirani, J. Friedman, and J. Franklin. The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer*, 27(2):83-85, 2005.
 - [23] T. W. Hnat, V. Srinivasan, J. Lu, T. I. Sookoor, R. Dawson, J. Stankovic, and K. Whitehouse. The hitchhiker's guide to successful residential sensing deployments. In *Proceedings of the 9th ACM*

References

Abstract

The average home in the US flushes 1000's of gallons of water down the drain each year while standing at the fixture and waiting for hot water. Some households use a pump for hot water recirculation (HWR) to ensure that hot water is always immediately available, but these systems can incur more than \$1000 per year in energy costs. In this paper, we explore the hypothesis that "just-in-time" HWR can reduce the energy footprint of HWR systems, without increasing hardware or installation costs and without increasing water waste or human annoyance. Our basic approach is to learn patterns of hot water usage in the home and to circulate hot water only when future hot water usage is highly likely. We call this approach Circulo. We evaluate Circulo by analyzing hot water usage patterns from 5 different homes over a period of 7--10 days each. Our results indicate that Circulo can reduce the energy needed for HWR by 30% while still providing households with hot water over 90% of the time.

Abstract

Context All US homes have water heaters.

The average home in the US flushes 1000's of gallons of water down the drain each year while standing at the fixture and waiting for hot water. Some households use a pump for hot water recirculation (HWR) to ensure that hot water is always immediately available, but these systems can incur more than \$1000 per year in energy costs. In this paper, we explore the hypothesis that "just-in-time" HWR can reduce the energy footprint of HWR systems, without increasing hardware or installation costs and without increasing water waste or human annoyance. Our basic approach is to learn patterns of hot water usage in the home and to circulate hot water only when future hot water usage is highly likely. We call this approach Circulo. We evaluate Circulo by analyzing hot water usage patterns from 5 different homes over a period of 7--10 days each. Our results indicate that Circulo can reduce the energy needed for HWR by 30% while still providing households with hot water over 90% of the time.

Abstract

Motivation

The average home in the US flushes 1000's of gallons of water down the drain each year while standing at the fixture and waiting for hot water.

Some households use a pump for hot water recirculation (HWR) to ensure that hot water is always immediately available, but these systems can incur more than \$1000 per year in energy costs. In this paper, we explore the hypothesis that "just-in-time" HWR can reduce the energy footprint of HWR systems, without increasing hardware or installation costs and without increasing water waste or human annoyance. Our basic approach is to learn patterns of hot water usage in the home and to circulate hot water only when future hot water usage is highly likely. We call this approach Circulo. We evaluate Circulo by analyzing hot water usage patterns from 5 different homes over a period of 7--10 days each. Our results indicate that Circulo can reduce the energy needed for HWR by 30% while still providing households with hot water over 90% of the time.

Abstract

Related Work

The average home in the US flushes 1000's of gallons of water down the drain each year while standing at the fixture and waiting for hot water.

Some households use a pump for hot water recirculation (HWR) to ensure that hot water is always immediately available, **but** these systems can incur more than \$1000 per year in energy costs. In this paper, we explore the

hypothesis that "just-in-time" HWR can reduce the energy footprint of HWR systems, without increasing hardware or installation costs and without increasing water waste or human annoyance. Our basic approach is to learn patterns of hot water usage in the home and to circulate hot water only when future hot water usage is highly likely. We call this approach Circulo. We evaluate Circulo by analyzing hot water usage patterns from 5 different homes over a period of 7--10 days each. Our results indicate that Circulo can reduce the energy needed for HWR by 30% while still providing households with hot water over 90% of the time.

Abstract

Approach

The average home in the US flushes 1000's of gallons of water down the drain each year while standing at the fixture and waiting for hot water.

Some households use a pump for hot water recirculation (HWR) to ensure that hot water is always immediately available, but these systems can incur

more than \$1000 per year in energy costs. In this paper, we explore the hypothesis that "just-in-time" HWR can reduce the energy footprint of HWR systems, without increasing hardware or installation costs and without increasing water waste or human annoyance. Our basic approach is to learn patterns of hot water usage in the home and to circulate hot water only when future hot water usage is highly likely. We call this approach

Circulo. We evaluate Circulo by analyzing hot water usage patterns from 5 different homes over a period of 7--10 days each. Our results indicate that Circulo can reduce the energy needed for HWR by 30% while still providing households with hot water over 90% of the time.

Abstract

Evaluation

The average home in the US flushes 1000's of gallons of water down the drain each year while standing at the fixture and waiting for hot water. Some households use a pump for hot water recirculation (HWR) to ensure that hot water is always immediately available, but these systems can incur more than \$1000 per year in energy costs. In this paper, we explore the hypothesis that "just-in-time" HWR can reduce the energy footprint of HWR systems, without increasing hardware or installation costs and without increasing water waste or human annoyance. Our basic approach is to learn patterns of hot water usage in the home and to circulate hot water only when future hot water usage is highly likely. We call this approach

Circulo. We evaluate Circulo by analyzing hot water usage patterns from 5 different homes over a period of 7--10 days each. Our results indicate that Circulo can reduce the energy needed for HWR by 30% while still providing households with hot water over 90% of the time.

Abstract

Results

The average home in the US flushes 1000's of gallons of water down the drain each year while standing at the fixture and waiting for hot water. Some households use a pump for hot water recirculation (HWR) to ensure that hot water is always immediately available, but these systems can incur more than \$1000 per year in energy costs. In this paper, we explore the hypothesis that "just-in-time" HWR can reduce the energy footprint of HWR systems, without increasing hardware or installation costs and without increasing water waste or human annoyance. Our basic approach is to learn patterns of hot water usage in the home and to circulate hot water only when future hot water usage is highly likely. We call this approach Circulo. We evaluate Circulo by analyzing hot water usage patterns from 5 different homes over a period of 7--10 days each. Our results indicate that Circulo can reduce the energy needed for HWR by 30% while still providing households with hot water over 90% of the time.

Abstract

Conclusions

Circulo can be a scalable cost-effective solution.

The average home in the US flushes 1000's of gallons of water down the drain each year while standing at the fixture and waiting for hot water. Some households use a pump for hot water recirculation (HWR) to ensure that hot water is always immediately available, but these systems can incur more than \$1000 per year in energy costs. In this paper, we explore the hypothesis that "just-in-time" HWR can reduce the energy footprint of HWR systems, without increasing hardware or installation costs and without increasing water waste or human annoyance. Our basic approach is to learn patterns of hot water usage in the home and to circulate hot water only when future hot water usage is highly likely. We call this approach Circulo. We evaluate Circulo by analyzing hot water usage patterns from 5 different homes over a period of 7--10 days each. Our results indicate that Circulo can reduce the energy needed for HWR by 30% while still providing households with hot water over 90% of the time.

Structure of a Good Abstract

1. Context
2. Motivation
3. Prior art
4. Approach
5. Evaluation
6. Results
7. Conclusions

Abstract Tells a Lot Paper Quality!

(or, avoiding common pitfalls!)

Context All US homes have water heaters.

The average home in the US flushes 1000's of gallons of water down the drain each year while standing at the fixture and waiting for hot water. Some households use a pump for hot water recirculation (HWR) to ensure that hot water is always immediately available, but these systems can incur more than \$1000 per year in energy costs. In this paper, we explore the hypothesis that "just-in-time" HWR can reduce the energy footprint of HWR systems, without increasing hardware or installation costs and without increasing water waste or human annoyance. Our basic approach is to learn patterns of hot water usage in the home and to circulate hot water only when future hot water usage is highly likely. We call this approach Circulo. We evaluate Circulo by analyzing hot water usage patterns from 5 different homes over a period of 7--10 days each. Our results indicate that Circulo can reduce the energy needed for HWR by 30% while still providing households with hot water over 90% of the time.

Abstract Tells a Lot Paper Quality!

Color coding for next few slides

The average home in the US flushes 1000's of gallons of water down the drain each year while standing at the fixture and waiting for hot water. Some households use a pump for hot water recirculation (HWR) to ensure that hot water is always immediately available, but these systems can incur more than \$1000 in energy costs. In this paper, we explore the hypothesis that HWR can reduce the energy footprint of HWR systems, without increasing hardware or installation costs and without increasing water waste or human annoyance. Our basic approach is to learn patterns of hot water usage in the home and to circulate hot water only when future usage is highly likely. We call this approach Circulo. We evaluate Circulo by analyzing hot water usage patterns from 5 different homes over a period of 7--10 days each. Our results indicate that Circulo can reduce the energy needed for HWR by 30% while still providing households with hot water over 90% of the time.

Good

Bad

Piece of the pie!

Context All US homes have water heaters.

The average **2% US homes have water heaters.** drain each year while standing at the fixture and waiting for hot water. Some households use a pump for hot water recirculation (HWR) to ensure that hot water is always immediately available, but these systems can incur more than \$1000 per year in energy costs. In this paper, we explore the hypothesis that "just-in-time" HWR can reduce the energy footprint of HWR systems, without increasing hardware or installation costs and without increasing water waste or human annoyance. Our basic approach is to learn patterns of hot water usage in the home and to circulate hot water only when future hot water usage is highly likely. We call this approach Circulo. We evaluate Circulo by analyzing hot water usage patterns from 5 different homes over a period of 7--10 days each. Our results indicate that Circulo can reduce the energy needed for HWR by 30% while still providing households with hot water over 90% of the time.

Piece of the pie!

Motivation

The average home in the US flushes **1000's** of gallons of water down the drain each year while standing at the fixture and waiting for hot water.

The average home in the US flushes **10's** of gallons of water down the drain each year while standing at the fixture and waiting for hot water.

systems, without increasing hardware or installation costs and without increasing water waste or human annoyance. Our basic approach is to learn patterns of hot water usage in the home and to circulate hot water only when future hot water usage is highly likely. We call this approach Circulo. We evaluate Circulo by analyzing hot water usage patterns from 5 different homes over a period of 7--10 days each. Our results indicate that Circulo can reduce the energy needed for HWR by 30% while still providing households with hot water over 90% of the time.

Is Prior Art Already Good Enough?

Prior art

Some households use a pump for hot water recirculation (HWR) to ensure that hot water is always immediately available, but these systems can incur more than \$1000 per year in energy costs.

Some households use a pump for hot water recirculation (HWR) to ensure that hot water is always immediately available, but these systems can incur \$5 per year in energy costs.

only when future hot water usage is highly likely. We call this approach Circulo. We evaluate Circulo by analyzing hot water usage patterns from 5 different homes over a period of 7--10 days each. Our results indicate that Circulo can reduce the energy needed for HWR by 30% while still providing households with hot water over 90% of the time.

Using a Weak Baseline/Bad Prior Art

Prior art

Some households use a pump for hot water recirculation (HWR) to ensure that hot water is always immediately available, but these systems can incur more than \$1000 per year in energy costs.

more than \$1000 per year in energy costs. In this paper, we explore the hypothesis that "just-in-time" HWR can reduce the energy footprint of HWR.

Some households use a 1940s technology and waste \$1000 USD.

learn patterns of hot water usage in the home and to circulate hot water only when future hot water usage is highly likely. We call this approach Circulo. We evaluate Circulo by analyzing hot water usage patterns from 5 different homes over a period of 7--10 days each. Our results indicate that Circulo can reduce the energy needed for HWR by 30% while still providing households with hot water over 90% of the time.

Weak evaluation

Evaluation

The average home in the US flushes 1000's of gallons of water down the drain each year while standing at the fixture and waiting for hot water. Some households use a pump for hot water recirculation (HWR) to ensure that hot water is always immediately available, but these systems can incur more than \$1000 per year in energy costs. In this paper, we explore the hypothesis that "just-in-time" HWR can reduce the energy footprint of HWR systems, without increasing hardware or installation costs and without increasing water waste or human annoyance. Our basic approach is to learn patterns of hot water usage in the home and to circulate hot water

We evaluate Circulo by analyzing hot water usage patterns from 5 different homes over a period of 7--10 days each.

different homes over a period of 7--10 days each. Our results indicate that

We evaluate Circulo by analyzing hot water usage patterns from 2 different homes over a period of 2 days each.

Metrics not Tying to Motivation

Results

The average home in the US flushes 1000's of gallons of water down the drain each year while standing at the fixture and waiting for hot water. Some households use a pump for hot water recirculation (HWR) to ensure that hot water is always immediately available, but these systems can incur more than \$1000 per year in energy costs. In this paper, we explore the hypothesis that "just-in-time" HWR can reduce the energy footprint of HWR systems, without increasing hardware or installation costs and without increasing water waste or human annoyance. Our basic approach is to

Our results indicate that Circulo can reduce the energy needed for HWR by 30% while still providing households with hot water over 90% of the time.

different homes over a period of 7-10 days each. Our results indicate that

Our results indicate that Circulo can predict water need with 95% RMSE while still providing households with hot water over 90% of the time.

Incomplete Metrics

Results

The average home in the US flushes 1000's of gallons of water down the drain each year while standing at the fixture and waiting for hot water. Some households use a pump for hot water recirculation (HWR) to ensure that hot water is always immediately available, but these systems can incur more than \$1000 per year in energy costs. In this paper, we explore the hypothesis that "just-in-time" HWR can reduce the energy footprint of HWR systems, without increasing hardware or installation costs and without increasing water waste or human annoyance. Our basic approach is to

Our results indicate that Circulo can reduce the energy needed for HWR by 30% while still providing households with hot water over 90% of the time.

different homes over a period of 7 - 10 days each. Our results indicate that Circulo can reduce the energy needed for HWR by 30% while still providing

Our results indicate that Circulo can reduce the energy needed for HWR by 30%.

Abstract

Motivation

The average home in the US flushes 1000's of gallons of water down the drain each year while standing at the fixture and waiting for hot water.

Some households use a pump for hot water recirculation (HWR) to ensure that hot water is always immediately available, but these systems can incur more than \$1000 per year in energy costs. In this paper, we explore the hypothesis that "just-in-time" HWR can reduce the energy footprint of HWR systems, without increasing hardware or installation costs and without increasing water waste or human annoyance. Our basic approach is to learn patterns of hot water usage in the home and to circulate hot water only when future hot water usage is highly likely. We call this approach Circulo. We evaluate Circulo by analyzing hot water usage patterns from 5 different homes over a period of 7--10 days each. Our results indicate that Circulo can reduce the energy needed for HWR by 30% while still providing households with hot water over 90% of the time.

Exercise #1 (3 mins)

Write an abstract for a smart coffee machine

1. Context
2. Motivation
3. Prior art
4. Approach
5. Evaluation
6. Results
7. Conclusions

Exercise #1 (3 mins)

Write an abstract for a smart coffee machine

1. Context: XX% of homes in the US have coffee makers
2. Motivation: YY work hours are lost annually due to late coffee
3. Prior art: Current coffee makers have timers, but...
4. Approach: 1) sensors 2) learning 3) optimization
5. Evaluation: 100 houses with both types; measured productivity at work
6. Results: Homes with smart coffee maker have ZZ% higher productivity
7. Conclusions: Companies should subsidize smart coffee maker purchases

Exercise #2 (5 mins)

Write an abstract for any research you want to do

1. Context
2. Motivation
3. Prior art
4. Approach
5. Evaluation
6. Results
7. Conclusions