Towards User Identification In The Home From Appliance Usage Patterns

Daniel Garnier-Moiroux
Mines ParisTech
d.garnier.moiroux@gmail.com

Fernando Silveira
Technicolor
fernando.silveira@technicolor.com

Anmol Sheth
Technicolor
anmol.sheth@technicolor.com

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 ACM 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.
UbiComp'13 Adjunct, September 8–12, 2013, Zurich, Switzerland.
Copyright © 2013 ACM 978-1-4503-2215-7/13/09...$15.00.

DOI string from ACM form confirmation

Abstract
We explore the feasibility of identifying users from the unique patterns they exhibit when interacting with an individual electrical appliance in the home. We evaluate the effectiveness of a supervised learning based approach for user identification from a dataset of appliance usage collected across five users and three kitchen appliances over a period of eight weeks. Our results show that using appliance usage information alone provides a moderate average accuracy of 32% for group sizes of up to five users in the home. However augmenting usage information with hints about user presence can improve accuracy by 15-20%.

Author Keywords
User identification, energy meter, smart home

ACM Classification Keywords
H.5.m [Information interfaces and presentation (e.g., HCI)]: Miscellaneous.

Introduction
The digital home of the future is envisioned to be a mix of sensing and computing infrastructure that seamlessly interacts with the user to enable a wide range of personalized digital home applications and services. Examples include recommending content by identifying

who is watching television, personalizing the settings of an appliance based on who is using it, and personalizing the cooking experience based on who is performing the activity in the home. A key component of many such applications is a non-intrusive and seamless user identification and tracking technique to personalize the experience for the user.

Existing approaches for user tracking and identification are cumbersome as they are either limited to individual devices that require explicit feedback from the user or make use of invasive sensors like microphones and cameras. Approaches requiring users to log in or pick a profile are limited to a handful of devices in the home, like smart TVs and media devices, and are often from the same manufacturer. Such approaches cannot provide seamless user tracking and identification across multiple heterogeneous devices in the home. Other approaches that require the installation of sensors like cameras and microphones raise several privacy concerns and are fragile to environmental conditions like poor lighting or background noise.

In this paper we propose an alternate approach to user identification and tracking. We identify which user is using a particular electrical appliance by profiling and learning the unique appliance usage patterns across the different users in the home. Appliance usage information is obtained by monitoring the energy consumption of individual appliances in the home from smart meters [4],[5] or distributed smart plugs[1]. Unique user patterns are profiled by extracting rich features like the time of day and the duration of appliance usage. This approach has the following three main advantages over existing approaches.

[1]http://talkingplug.com

Ease of deployment. Smart plugs and energy meters are simple to install in the home and are fast becoming ubiquitous. Furthermore, these devices are inexpensive and do not require replacing batteries since they are already plugged into the power line.

Support for heterogeneous appliances. Identifying appliance usage by monitoring the energy consumption does not require any cooperation from the appliances and can seamlessly be used across heterogeneous appliances in the home.

Minimal user interaction. The approach is lightweight and does not require any explicit interaction by the user after the system is trained.

We report on the feasibility of such an approach and evaluate the accuracy of user identification. We adopt a supervised learning based approach that trains a model for each user in the home given a few days of labeled appliance usage data. We also compare the accuracy by augmenting the appliance usage data with user presence information. Our dataset is based on energy monitor data collected from the kitchen area in an office over a period of eight weeks involving five users. Our key contributions are:

- As far as we know, this is the first paper that explores the feasibility of identifying users from appliance usage information derived from electrical smart meters in the home.
- We show how usage features extracted from the power consumption of an electrical appliance are used by supervised machine learning algorithms for user identification.

- Our results show that user identification from appliance usage data alone is hard. We achieve an average accuracy of 32% for groups of up to five users. However adding hints about user presence significantly improves the accuracy of user identification by an additional 20% and provides a 2x improvement over random chance.

Even though our results are preliminary and our accuracy in identifying users are moderate, we believe this work shows a very promising avenue for future work. Namely, we show that usage data from individual appliances contains enough signal to improve identification accuracy by twice of what a random guess would perform. While it is intuitive that our daily interactions with appliances (e.g., time of day, appliance-specific settings) have unique patterns, our work shows that these patterns can be leveraged in a real setting. We believe that identification accuracy can be further improved by, e.g., combining usage data from multiple appliances. In our conclusions, we discuss these and other extensions of our approach that are left for future work.

System Design and Experimental Setup

The overall system design is based on a supervised learning approach using linear Support Vector Machines (SVM) that *learn* the per-appliance usage patterns of users in the home. We also evaluate the accuracy of user identification by augmenting the appliance usage data with information about whether the user is present and how long ago did the user arrive. Intuitively, these features should improve identification accuracy by ruling out users who are not present and by capturing appliance usage events that happen immediately after the user arrives home. In the rest of this section we provide an overview of the deployment and outline the four main components of our approach.

System Deployment

We deployed the system in the kitchen area of our lab. We use the *Watts Up? Pro ES* smart plugs [2] and collect energy consumption data of three appliances (a coffee machine, a refrigerator, and a microwave oven) shared by five users over a period of eight weeks. In this paper, we present results for the coffee machine which, during our measurement period, generated 133 usage events. Figure 1 shows the distribution of these usage events across the duration of the experiment (the x-axis) and the five users (the y-axis) participating in the study. Each circle's radius is proportional to the number of events involving a given user on a given day.

Figure 1: Experimental data for the coffee machine showing the distribution of events across the duration of the experiment

Data Acquisition

The energy monitors report power consumption every second for each of the three appliances to a central database. The power readings are within 0.2-0.3%

accuracy and can support electrical loads across a wide range. This high accuracy and fine-grained resolution of the power meters makes it possible to differentiate between the different power states of the appliance. For example, we can differentiate between the opening of the main door and freezer door of the refrigerator as the two lights bulbs vary in power consumption by 5 W.

In addition to energy usage information we also collect user presence information from logs generated by a system of electronic key fobs used to provide secure access to the office. Every time a key fob is used to open the office door, we receive a notification record that contains the key owner's identity. In the home, similar presence information can be obtained by monitoring the home network for the MAC addresses of personal devices like smartphones. Infrared motion sensors can also be used as a simpler (and anonymous) way of notifying the arrival of users in the home [7].

Feature Extraction

We use energy monitor data and key fob logs to extract features that characterize the interactions between users and appliances. From the energy monitor data we extract features for the appliance type and duration of use and from the entry logs we extract the user presence features.

Appliance Usage Features. Energy monitor data is used to differentiate between the different power states of the appliance to determine how the appliance is being used and the duration of use. We develop custom threshold based event detectors for the three appliances by performing controlled experiments and measuring the power consumption for the different power states of the appliance. Using appliance-specific criteria, we aggregate the time series of power states into usage events. Each event is represented by its starting time of day and its total duration. We quantize the time of day feature into 10-minute intervals (e.g., "8:14 AM" is converted to "8:10 AM"), thus leading to a binary vector of length 144. Similarly, we quantize the duration values into a binary vector of 10 equal-sized intervals between the smallest and the highest observed durations in our experiments.

User Presence Features. In addition to the above features, we augment each usage event with six additional features for every user. These six binary features indicate if the user has entered the office 1, 5, 15, 30, 60 minutes ago, or at any time so far in the day. Note that the features we extract from key fob logs can only provide approximate information about a user's presence for two main reasons. First, since we do not have exit logs we cannot reliably detect the absence of a user. Second, if multiple users enter the office together (which happens often due to shared commutes), we can only update the presence features of a single user (i.e., the key's owner).

SVM Training and Inference

We evaluate the effectiveness of our user identification approach using linear SVMs and use the Weka implementation [1] of Sequential Minimal Optimization (SMO) [6] to train the SVMs. In order to provide ground truth of user identities, we manually analyze the video feed of a surveillance camera installed in the kitchen area. In the home, users could simply label the events manually during the training phase instead of using intrusive sensors. We experimented with different sizes of training sets and found that user patterns are fairly stable and identification accuracy does not increase significantly with more than two weeks of training data. Figure 1 shows the two weeks of training data marked in grey and we leave the remaining five weeks as test data.

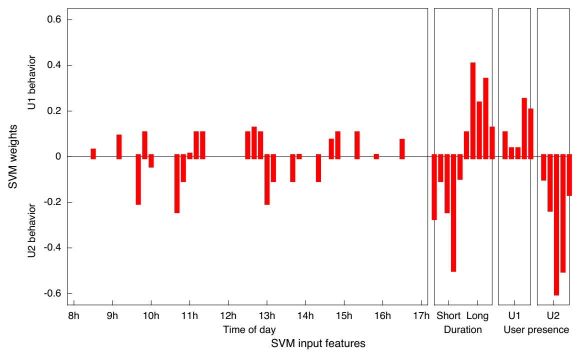

Figure 2: Experimental data for the coffee machine showing the SVM weights trained for two representative users in the dataset

We learn a binary SVM classifier for every pair of users and apply the commonly used pairwise coupling [3] method to aggregate their predictions into a single user identity. Figure 2 shows an example of a trained SVM model for the coffee machine represented as a weight for each feature. A high positive or negative weight represents the importance of the feature to differentiate between the two users. The figure show that U1 tends to use the coffee machine throughout the day whereas U2 primarily uses the machine in the morning and after lunch. The strong negative weights for the presence features for U2 also shows that U2 primarily uses the machine after entering the lab. Additionally, U2 makes longer coffees compared to U1 as shown by the weights of duration features.

Experimental Results

In this section we focus on the impact of two system parameters on identification accuracy: the maximum number of users that the classifier learns to identify, and the types of input features used.

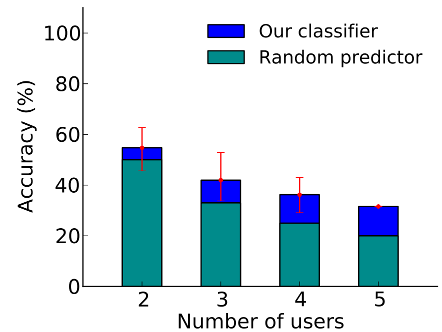

Figure 3: Usage features

Impact of Number of Users. Figure 3 shows how identification accuracy degrades with the number of users in the case where SVMs are trained only with the time of day and duration of a user's interaction with the appliance. Each bar denotes the average accuracy (y-axis) across all subsets of users of a given size (x-axis) and intervals show the minimum and maximum values. Some subsets of users are easier (or harder) to identify than average due to very distinct (or resp. very similar) usage patterns. As a baseline, we show the expected accuracy of a naive predictor which randomly guesses the user's identity[2]. The plot shows that although the SVM's accuracy degrades with additional users, it does so slower than the random predictor's. In fact, the relative advantage of our classifier over a random predictor increases with the number of users, 9-58% in the range shown in the plot.

Impact of Presence Features. Figure 4 shows the improvement in accuracy due to the presence features.

[2]We also experimented predicting the most frequent user identity in the training set, but the average results were similar to those of a random predictor.

Across all group sizes, the introduction of presence features improves accuracy by 15-20% over the classifier based only on usage features.

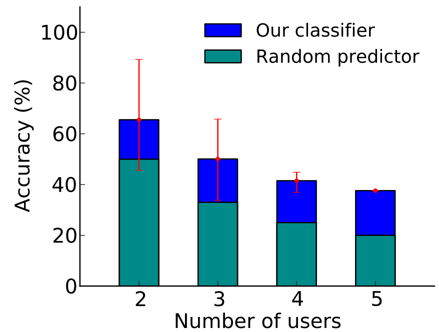

Figure 4: Usage and presence features

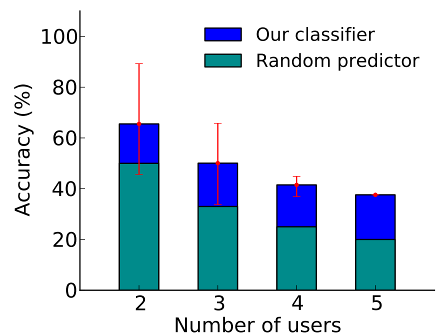

Figure 5: Benefit of presence features

To explain how presence features improve identification, Figure 5 shows relative accuracy improvements aggregated by the time of day. Each point in the plot is the relative accuracy improvement (y-axis) measured on test events for that hour of the day (x-axis). The plot shows that the improvement is very high in the early morning (peaking at 9:00), when users arrive in the office and their presence is recorded in our features. As the morning progresses, the information from presence features becomes stale, as we are not able to tell if any users have left the office. Furthermore, the presence information does not help identify users that use the coffee machine throughout the day. Note in Figure 5 that users re-enter the office after lunch (at 14:00) and the presence features improve accuracy again, but much less than at 9:00. This happens because users tend to leave for lunch and return in groups, and the only presence features that get updated are those of whichever user opens the door. Our results show that although presence features provide useful contextual information, the inherent noise in how they are acquired limits their benefit in practice. This suggests that finer-grained contextual information can lead to a more accurate identification system.

Conclusion

We present a seamless and non-intrusive approach for user identification in the home based on learning the patterns in people's routine interactions with individual electrical appliances. We evaluate the accuracy based on a real-world setup consisting of five test users that regularly used three kitchen appliances over a period of eight weeks in an office kitchen area. Our results show that while it is hard to identify users solely based on appliance usage features, additional hints about user presence can improve the average identification accuracy by 15%-20%.

The moderate accuracy of our approach can be further improved and as part of future work we plan on addressing the following two problems. First, our dataset was based on a deployment in an office setting that leads to more noisy data compared to a regular home environment. Additionally, our presence features are noisy as they do not accurately capture entry and exit events. We will address this by carrying out deployments in real homes where user schedules are well defined and presence information is accurate. And second, we will extend our approach to model user patterns across multiple appliances instead of limiting it to individual appliances.

References

[1] University of Waikato - Weka. `http://www.cs.waikato.nz/weka`.

[2] Watts Up? Power Meters. `http://www.wattsup.com/`.

[3] T. Hastie and R. Tibshirani. Classification by pairwise coupling. In *Proceedings of NIPS*, pages 507–513, 1997.

[4] Y. Kim, T. Schmid, Z. M. Charbiwala, and M. B. Srivastava. ViridiScope: Design and Implementation of a Fine Grained Power Monitoring System for Homes. In *Ubicomp*, pages 245–254, 2009.

[5] S. N. Patel, T. Robertson, J. A. Kientz, M. S. Reynolds, and G. D. Abowd. At the flick of a switch: detecting and classifying unique electrical events on the residential power line. In *Ubicomp*, pages 271–288, 2007.

[6] J. C. Platt. Using analytic QP and sparseness to speed training of Support Vector Machines. In *Proceedings of NIPS*, pages 557–563, 1998.

[7] D. H. Wilson and C. Atkeson. Simultaneous Tracking and Activity Recognition (STAR) using many anonymous, binary sensors. In *IEEE Pervasive*, pages 62–79, 2005.