Andreas Reinhardt, Paul Baumann, Daniel Burgstahler, Matthias Hollick, Hristo Chonov, Marc Werner, Ralf Steinmetz: *On the Accuracy of Appliance Identification Based on Distributed Load Metering Data*. Proceedings of the 2nd IFIP Conference on Sustainable Internet and ICT for Sustainability (SustainIT). October 2012

On the Accuracy of Appliance Identification Based on Distributed Load Metering Data

Andreas Reinhardt*, Paul Baumann‡, Daniel Burgstahler*, Matthias Hollick†, Hristo Chonov*, Marc Werner†, Ralf Steinmetz*

**Multimedia Communications Lab, Technische Universität Darmstadt, Germany*
{andreas.reinhardt, dburgsta, hchonov, ralf.steinmetz}@kom.tu-darmstadt.de

‡Wireless Sensor Networks Lab, Technische Universität Darmstadt, Germany
paul.baumann@wsn.tu-darmstadt.de

†Secure Mobile Networking Lab, Technische Universität Darmstadt, Germany
{matthias.hollick, marc.werner}@seemoo.tu-darmstadt.de

***Abstract*—Dynamic load management, i.e., allowing electricity utilities to remotely turn electric appliances in households on or off, represents a key element of the smart grid. Appliances should however only be disconnected from mains when no negative side effects, e.g., loss of data or thawing food, are incurred thereby. This motivates the use of appliance identification techniques, which determine the type of an attached appliance based on the continuous sampling of its power consumption. While various implementations based on different sampling resolutions have been presented in existing literature, the achievable classification accuracies have rarely been analyzed. We address this shortcoming and evaluate the accuracy of appliance identification based on the characteristic features of traces collected during the 24 hours of a day. We evaluate our algorithm using more than 1,000 traces of different electrical appliances' power consumptions. The results show that our approach can identify most of the appliances at high accuracy.**

I. Introduction

In a number of application scenarios, the identification of electric appliances plugged into a wall outlet plays a major role. Firstly, the increasing number of renewable energy sources in the power grid requires electricity utilities to be able to quickly react to changes in supply and demand. Dynamic load management, i.e., the ability to remotely control devices, is a key element in order to cope with the partially unpredictable behavior of regenerative sources. It must be ensured that certain appliance types (e.g., computers) are not turned off while they are in use in order to increase the user's acceptance of such means. Determining the appliance type prior to making it available for remote actuation is thus essential. Secondly, globally rising energy prices are of major concern to households, which strive for possibilities to preserve energy and minimize its inefficient use. Commercial units like Kill A Watt [1] help users to assess the energy consumption of their appliances. However, the collected data needs to be manually annotated with the measured device type in order to draw a fine-grained portrait of a household's energy balance. Finally, the capacity to remotely switch the mains connection of appliances is a major enabler for smart environments, in which appliances are automatically controlled based on the user's preferences [2].

In this paper, we present and evaluate an automated approach to identifying devices based on distributed power measurement and actuation units (*MAU*s). MAUs are connected in-between an appliance's power plug and the wall outlet, as visualized in Fig. 1, and thus provide the capability to analyze the power consumption and switch any connected device individually. Our concept is different from the prevailing approach of non-intrusive appliance load monitoring (*NIALM*) [3] in two major aspects. Firstly, our approach operates in a distributed fashion, while NIALM relies on a single measurement unit for a household's overall power consumption. The use of MAUs thus allows for a more fine-grained analysis of attached consumers and simultaneously reduces ambiguities when several consumers with similar energy consumption behavior are present. Also, more sensitive measurement equipment can be used in MAUs, because the load attached to a wall outlet is usually smaller than the total household load. The second difference to NIALM is the fact that MAUs are intrusive devices, i.e., appliances are connected to the mains via the MAU and must hence be unplugged during its installation.

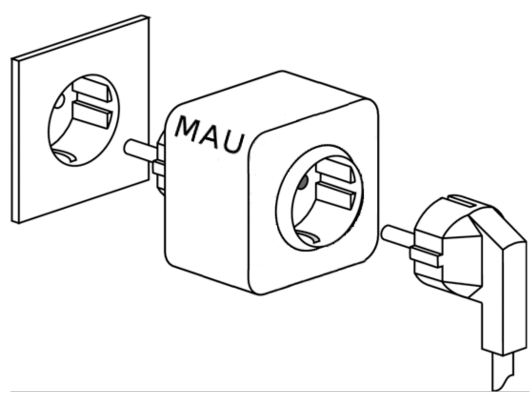


Figure 1. Schematic placement of a MAU into an appliance's mains connection

The two main contributions of this paper are the introduction of our *tracebase* repository of real-world power consumption traces [4] and the evaluation of the automated recognition of electric appliances based on the collected traces. To this end, we introduce methods to measure and represent load characteristics, and present the applied signal processing and estimation techniques for load recognition. This paper is structured as follows. First, we provide an overview of the data collection setup in Sec. II, and introduce the *tracebase* repository, where we make our collected traces publicly available to researchers. Subsequently, we present implementation details of our appliance identification framework in Sec. III, and evaluate the accuracy of our proposed algorithm in Sec. IV. We summarize related work in Sec. V and conclude this paper in Sec. VI.

II. Data Collection

For a thorough evaluation of device classification mechanisms under real-world conditions, a comprehensive collection of real-world traces is needed. Although traces that represent the power demand of complete households are available to the research community, such as the REDD dataset [5], they primarily contribute to research on load disaggregation algorithms. For an identification of appliances at high accuracy and solely based on their power consumption, we argue that data is required at the granularity level of individual devices. The first contribution of this paper is thus the presentation of our *tracebase* respository, which contains more than a thousand electrical appliance power consumption traces that have been collected in more than ten households and office spaces.

A. Setup

For the collection of the data sets from a large range of electric appliances, we have used the Plugwise system [6], a commercially available distributed smart metering platform. In contrast to research prototypes like ACme [7], the Plugwise system is available in larger quantities as well as being approved for its electrical safety. Its is based on two main components, namely the Circle, which takes the role of the MAU and is connected between an appliance and the wall outlet, and the Stick, which wirelessly interfaces the deployed Circles to a computer. Circles can be queried periodically by the stick, returning the average real power demand of the attached appliance over intervals of one and eight seconds, respectively. Although Circles lack support for measuring reactive power and phase shift, we have selected the platform due to its availability in large numbers and the convenient installation.

For the data collection, we have created a polling application, which requests the wattage readings collected by each of the deployed Circles. In order to successfully establish a communication to the data collecting host system, the "Plugwise Unleashed" protocol [8] has been used. The collected readings are stored on the data collection node in comma-separated value files; one file is created for each metered appliance every day. Each entry in the file is preceded by the date and time of its collection, followed by the two power consumption readings. An excerpt of a trace collected by a Circle attached to a refrigerator is shown in Fig. 2.

```
14/01/2012 10:48:47; 151; 156
14/01/2012 10:48:48; 147; 151
14/01/2012 10:48:49; 147; 151
14/01/2012 10:48:50; 145; 149
14/01/2012 10:48:51; 145; 147
14/01/2012 10:48:52; 145; 147
14/01/2012 10:48:53; 143; 145
14/01/2012 10:48:54; 143; 145
14/01/2012 10:48:55; 143; 143
```

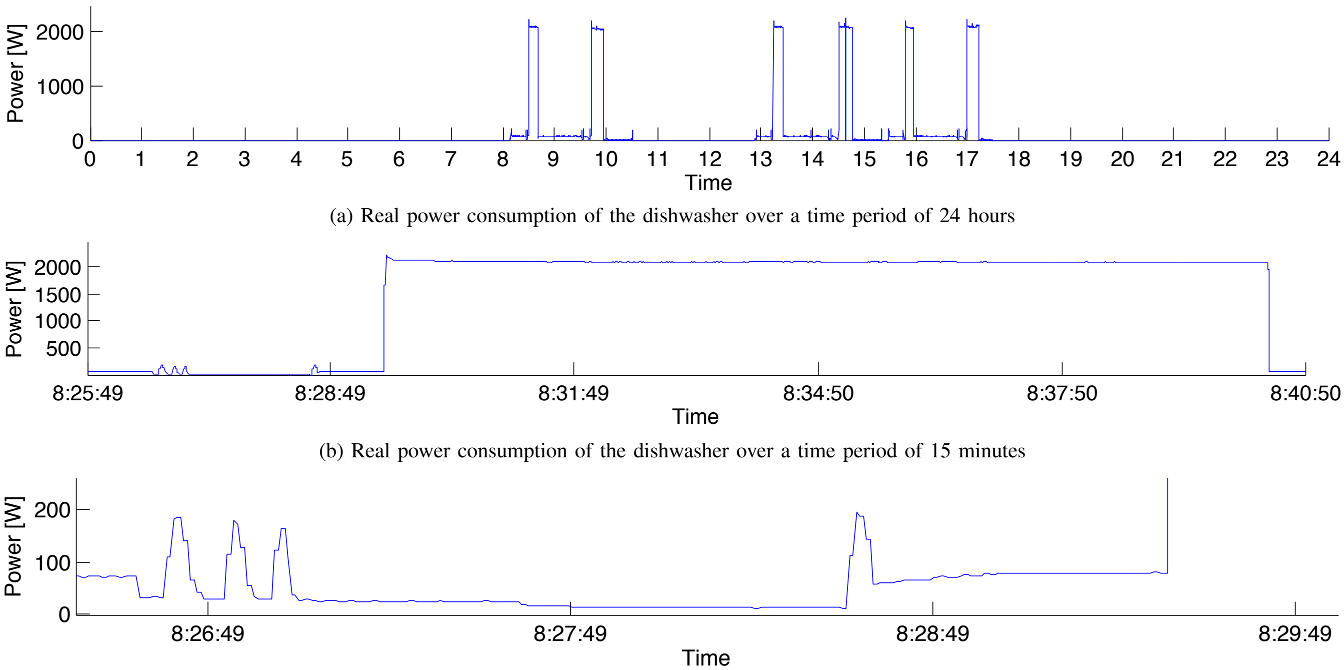
Figure 2. Excerpt of a refrigerator's trace, indicating time and date of collection as well as one- and eight-second average real power consumption

B. Trace Collection

We have attached Circles to more than a hundred devices in order to collect a large variety of fine-grained power traces. Both household and office appliances have been monitored in order to provide a broad foundation for analyses. Circles were connected to some appliances for

Table I
NUMBER OF COLLECTED TRACES FOR EACH DEVICE TYPE

Device type	# appliances	# traces
Alarm clock	1	5
Bean-to-cup coffee maker	1	44
Coffee maker	5	39
Computer monitor	14	156
Computer printer	2	16
Cooking stove	1	16
Desktop computer	9	90
Digital TV receiver	2	24
Dishwasher	3	47
DVD player	3	5
Ethernet switch	3	33
External USB hard disk drive	4	30
Freezer	1	9
HDTV Media center	3	17
HiFi stereo amplifier	3	52
Iron	1	3
Lamp	6	45
Laptop computer	6	67
Microwave oven	5	48
Playstation 3 console	2	14
Refrigerator	7	130
Subwoofer	2	28
Television set	10	94
Toaster	4	25
Tumble Dryer	2	9
Vacuum cleaner	1	1
Video projector	1	19
Washing machine	7	22
Water fountain	1	56
Water kettle	8	86
WiFi router	4	40
Total	122	1,270



(a) Real power consumption of the dishwasher over a time period of 24 hours

(b) Real power consumption of the dishwasher over a time period of 15 minutes

(c) Real power consumption of the dishwasher over a time period of 3 minutes

Figure 3. Visualization of the fine temporal granularity of the data collected in the *tracebase* repository using the example of a dishwasher

several months (e.g., the water fountain and a bean-to-cup coffee maker), whilst others were only in place for a couple of days. To the present day, we have collected a data set comprised of more than 1,000 power consumption traces from 31 different types of appliances, as listed in Table I. Further deployments are still ongoing, and collected traces are added to the *tracebase* continually. Besides indicating the monitored appliance type, the table also provides the number of instances for a given type (e.g., the four different instances of the type *toaster*) as well as the total number of traces per appliance. In order to get an impression of the fine granularity of the collected traces, we have visualized a dishwasher's power consumption trace in different degrees of detail in Fig. 3. In contrast to smart meters, which are mostly limited to data collection intervals of several minutes, the high temporal resolution of our traces enables unprecedented analyses of the power consumption.

We have deliberately decided to collect traces for each appliance individually in order to give researchers the opportunity to conduct analyses on a per-device level. More realistic measurement scenarios, such as sub-level metering, can however be easily realized by the superposition of several traces. Similarly, the collection of data at a sampling rate of one reading per second might not reflect the actual sampling rate of typical smart meters, but can be reduced easily by means of integrating filters. All collected traces are available to the research community within the scope of our *tracebase* project, which can be accessed online at `http://www.tracebase.org`.

III. Classification of Appliances

By having created a comprehensive repository of power consumption traces, the foundation of our appliance identification algorithm is established. In this section, we discuss how we extract representative features from each of the traces. These are used to train the classifiers that finally enable the classification of yet unknown traces.

A. System Operation

An architectural overview of our appliance classification system is shown in Fig. 4, and is explained as follows:

- In its first step, traces (as well as an annotation of the device type that has been monitored) are collected using the MAUs. The system has been optimized for the use of diurnal traces, i.e., traces collected over the 24 hours of a day, but can be easily adapted to other input formats. In the remainder of this paper, we utilize the traces collected in our *tracebase* project, as described in Sec. II.
- In the second step, the system extracts several characteristic features from the diurnal input traces and stores them in a feature vector. We have implemented the feature extraction stage in a modular fashion, such that the implementations which extract features (we refer to them as feature *processors*) can be modified and deactivated easily and without any interdependencies. Each resulting feature vector is annotated by the actual device class from which the trace was collected in order

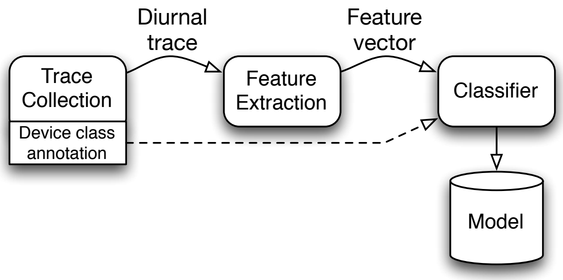


Figure 4. Architecture of our appliance identification system

to train the classifier. The output of this second stage is a list of feature vectors, one per input trace.

- Finally, the list of generated feature vectors is used to train a classifier. The underlying model is internally created by the selected classifier, and allows it to categorize new and yet unseen feature vectors into one of the previously trained device type classes.

While the first step involves both the data collection on a MAU and the storage in a database system, the remaining two steps are executed on an evaluation server with sufficient computational power to handle the high computational load incurred by the feature extraction and the memory required to maintain the resulting models. For an evaluation of the resulting classification accuracy, the classifier can either be invoked to classify a dedicated testing set of feature vectors, which are not part of its model. Alternatively, a cross-validation can be performed, in which a percentage of the available feature vectors is deliberately removed during the model construction, and instead used to assess the achievable classification accuracy. In this paper, we follow the latter approach of using cross-validation to assess the achievable classification accuracy.

B. Feature Selection

We have analyzed the collected readings with the goal to extract unique features to support a reliable device identification. The resulting feature processors extract features for the following four main classes, which are described in more detail in the following paragraphs:

- Temporal appliance behavior
- Energy and power consumption levels
- Shape of the power consumption
- Noise level characteristics and statistic features

In total, our processors extract numerous features for all of the listed classes (an analysis of the most important features is presented in Sec. IV-C), hence we confine our descriptions to the most relevant sets of features. Prior to the immediate consideration of data traces for the training, however, we have applied a preprocessing step that disregards traces in which the power draw observed throughout the entire day has been constant (e.g., if an appliance has not been used all day).

1) Time-based features: We describe the time during which a device is switched on or in use by several features. Firstly, we count the number of seconds during which the appliance's power consumption exceeds a threshold of two watts, above which we consider the appliance as *active*. This empirically determined values also suppress the noise floor introduced when using Plugwise Circles as MAUs. We consider the total duration of a device's activity over the whole day as well as in intervals of six hours each. Additional usage information is extracted for the different times of the day, i.e., morning, noon, evening and night time.

We count the number of alternations between the on and off state of the individual appliances as well as the number of complete activity cycles throughout a day. The average activity duration of these intervals, divided by the number of intervals during a day leads to a numeric feature that returns a nearly steady value for many device classes. Additionally the duration of the smallest and the longest interval as well as corresponding average values are being extracted. Beyond this, the complementary values of the interval times, the off times, are calculated as the shortest and longest time periods during which an appliance is inactive. We supplement these features by their average values and sums.

As an extension to this, the time between blocks of closely spaced on-off-intervals, so called *usage-blocks*, is considered separately with respect to shortest, longest, and average time spans. We also extract if the activity durations are increasing or decreasing throughout the day, or if the time between the usage-blocks is increasing. Furthermore, we calculate quotients of the difference of the longest and shortest values to the corresponding average value.

The time spans during which the power consumption is within a bound of 2% to the highest, the lowest, and the average power values during the whole day are also considered. We furthermore regard the proportion of time during which the device was active as compared to the time it was switched off. The day of the week is taken into account, indicating if the device has been seen active on weekdays and weekends regularly. Finally, we divide the whole day into 144 intervals of ten minutes each, and analyze during which periods the device has been active.

2) Energy and power: As one of the most characteristic features we extract the maximum power consumption during a day as well as for each hour of the day individually. We calculate the average power and energy consumption over the whole day in various ways. Firstly, we regard blocks of two and six hours duration each. Secondly, we consider the first and second half of the day, and thirdly confine our average calculation to the time during which the device was active. Moreover, the median of the power consumption is calculated for the active time. In addition, we again divide the whole day into 144 time slots and calculate the average power and energy demand for every slot.

For the energy- and power-related features, we also consider the aforementioned usage-blocks. We calculate the smallest and the largest power and energy levels as well as their averages for each of the usage-blocks. Moreover, we analyze the variances of power and energy between these blocks and consider the increase and/or decrease in power between successive intervals. We pay special attention to the longest activity interval, for which the minimum and maximum power values are extracted as separate features. Furthermore, we calculate the average power and integrate over the interval to extract the energy consumed. We determine the minimum and maximum power value during the longest activity interval and calculate the corresponding percentage in time of these values during this interval. An additional feature of boolean type is used to indicate if there is more than one interval of the maximum length.

The next regarded energy and power feature indicates whether the power consumption over the whole day is nearly constant, i.e., only a fluctuation of at most three watts around the average value. In consideration of the position of the strongest peak, we use a median filter to eliminate potential outliers caused by disturbance or erroneous sensor readings. Finally, the range between the lowest and highest observed power consumption is divided into ten equal-sized subranges, and the extent of time during which the measured power level falls within each of these ranges is determined.

3) Power consumption shape: We describe the strength of variations throughout a day by two numeric features that are incremented every time two consecutive values differ by more than 5% or 60% of the moving average value, respectively, indicating the smoothness of the power consumption's shape. Next, we count how many times the power consumption curve crosses the thresholds of 5, 10, 50, 200, 500, and 2,000 watts, respectively. Short-lived threshold exceedings, lasting for less than 20 seconds, are counted separately from longer-lasting ones.

We have determined that important and characteristic features can be extracted from the shape of the power consumption around peaks. Thus, special attention is paid to the area around the highest peak. After finding this peak, the average value of the previous ten samples is compared with the following two threshold values. The first threshold corresponds to the average power level during the on state of the device, the second one to the average value throughout whole day. These thresholds are being used to determine the steepness of the rising and falling edge of the shape. Likewise, we calculate the slopes around the smallest and highest peak as well as an average value. We do this twice, once by only taking the first 20% of the interval into account and once taking the whole interval into account. Further features are extracted by taking each activity interval into account individually and extracting the smallest, highest, and average number of peaks within it along with the relative position of its occurrence. Also, the corresponding smallest, highest, and average power levels of these peaks are collected.

We set a threshold to 90% of the highest peak value, to which we compare the average value of the ten samples following the peak. The resulting feature describes if the shape keeps on a high value or if the power demand quickly declines after an initially large power demand. The time until the difference between consecutive values to the average of their 10 previous samples is less than 2% is counted, and lastly, we calculate the slope of the shape between the peak point and the point in time that meets aforementioned condition.

4) Noise level and statistic features: To describe the noise or high fluctuations in the measured power curves, we compare the original measured curve with a low pass filtered one. We individually compare the curves for each of the activity intervals and the off intervals, and match the smallest and the highest value as well as the average value of each of the sub-traces. We further compute the autocorrelation of each diurnal trace. Extracted features are the number of points of inflection and the value of the first minimum. As a last feature extracted by the autocorrelation processor, we determine the value and the time shift of all local extrema.

A feature that allows an assessment of the previous calculated average value is the standard deviation. We do several computations on the frequencies in the collected data traces. By using the Fast Fourier transform, we determine the highest frequency in the input trace, and analyze the amount of energy within frequency ranges of 5%, 15% and 25% of the highest frequency. We also use a ten-point discrete Fourier transform to directly extract feature values that describe periodicities in the data traces of the different device classes. To consider the fact that different device classes are likely to draw different power levels during their operation, we use the concept of a histogram. We divide the range of power consumption into 18 equidistant compartments, and we analyze the class-wise average distribution of the power consumption for each of these compartments.

5) Summary: We have developed an appliance classification framework, in which feature processors can be integrated in a modular fashion. Our current implementation uses more than 10 feature processors which extract a total number of 517 characteristic features from each diurnal power consumption trace. While the framework is tailored to the simple integration of feature processors and the analysis of their classification accuracy, it can also be used to classify new and yet unseen traces. Therefore, the same set of features is extracted from the incoming trace, and the classifier (which has already trained its model using the data from the *tracebase*) is invoked to classify the trace.

Table II
CLASSIFICATION ACCURACY FOR DIFFERENT CLASSIFIERS

Algorithm	Cross-validation accuracy (25 folds)	Model construction time
Bagging	92.73%	7.38s
Bayesian Network	91.48%	1.46s
J48	91.31%	2.18s
JRip	84.21%	10.75s
LogitBoost	93.65%	31.3s
Naive Bayes	89.89%	0.13s
Random Committee	**95.5%**	**0.42s**
Random Forest	95.07%	0.43s
Random Tree	84.13%	0.1s

IV. EVALUATION

For our practical evaluation, we have implemented the presented appliance identification framework using the Java programming language. The framework caters for the extraction of the features (cf. Sec. III-B) from the collected traces, the creation of a machine learning model, and the classification of traces according to the established model. For the construction of the classification model, we have used the Weka data mining toolkit [9].

A. Classifier Selection

In the first step of our evaluation, we have assessed the classification accuracy of different classifiers. To this end, we have used 1,197 of the power traces collected in the *tracebase* and extracted all 517 features, the most important ones of which have been described above, for each of the traces. Traces without any activity, i.e., a continuous power consumption below two watts, were excluded from our analysis, such that not all samples present in the *tracebase* were used. On average, the extraction of all 517 features from any diurnal input trace required 4.97 seconds on an Intel Core2Duo E7400, clocked at 2.8GHz. A measurable fraction of this duration can however be attributed to the file read operations on the hard disk drive. In terms of its memory consumption, the feature extraction exposed a peak memory demand of slightly more than 800MB of RAM.

Nine classifiers were trained with the resulting feature set and analyzed with respect to their classification accuracy. The resulting values for the classification accuracies for 25-fold cross validations are presented in Table II, along with the time required to construct the corresponding models. A general observation from the table is that all considered algorithms achieve accuracies in excess of 80%, which can be taken as an indication that the extracted features are well suited to describe individual appliance types and distinguish them from others. The best results were observed for the Random Committee algorithm (which internally relies on ten Random Trees), catering for an average classification accuracy of 95.5%. The Random Committee is also among the fastest classifiers to construct its model; it is only outperformed in terms of speed by the Naive Bayes and Random Tree approaches, which however have significantly lower classification accuracy values. As a result, we have confined our analyses to the use of the Random Committee algorithm in our further evaluations.

B. Detailed Classification Results

In a subsequent evaluation, we have regarded the overall classification result of the Random Committee algorithm in more detail. Table III thus shows the resulting values for true and false positives as well as precision and recall for each device class.

Only a single device type (the CRT monitor, entry Y) has a comparably poor classification result with a true positive ratio of only 11.8%. Across all other device classes, our implementation however reaches at least 80% classification accuracy, and more than half of the device classes are always classified correctly, i.e., they have a true positive ratio of 1.0. The observed precision and recall values are well inline with the observations for the true positive rate. Confusion, indicated by the false positive classification rate, is below 0.02 across all 33 device classes, and highest for the CRT and LCD monitor classes (types Y and 1). More than half of the devices can be unambiguously classified, with no confusion between their device classes.

Table III
CLASSIFICATION ACCURACY FOR EACH APPLIANCE WITH TRUE POSITIVE (TP), FALSE POSITIVE (FP), PRECISION, AND RECALL VALUES (25-FOLD CROSS-VALIDATION, RANDOM COMMITTEE)

Appliance	TP	FP	Precision	Recall
Playstation 3 (A)	0.917	0.002	0.846	0.917
Desktop PC (B)	0.971	0.001	0.986	0.971
Laptop computer (C)	1.0	0	1.0	1.0
USB harddrive (D)	1.0	0	1.0	1.0
Toaster (E)	0.952	0	1.0	0.952
HiFi amplifier (F)	0.984	0.001	0.984	0.984
LCD TV set (G)	0.956	0.003	0.956	0.956
Coffeemaker (H)	0.976	0.003	0.93	0.976
Dishwasher (I)	1.0	0.002	0.955	1.0
Bean-to-cup (J)	1.0	0	1.0	1.0
Ethernet switch (K)	1.0	0	1.0	1.0
Video projector (L)	0.963	0	1.0	0.963
Subwoofer (M)	1.0	0.001	0.971	1.0
Media center (N)	1.0	0	1.0	1.0
Alarm clock (O)	0.833	0	1.0	0.833
DVD player (P)	1.0	0.001	0.941	1.0
Freezer (Q)	1.0	0	1.0	1.0
Tumble drier (R)	0.923	0	1.0	0.923
WiFi router (S)	1.0	0.001	0.976	1.0
Microwave oven (T)	0.905	0.003	0.927	0.905
Cooking stove (U)	1.0	0	1.0	1.0
Water fountain (V)	1.0	0	1.0	1.0
Washing machine (W)	0.957	0	1.0	0.957
Iron (X)	1.0	0	1.0	1.0
CRT monitor (Y)	0.118	0.01	0.143	0.118
CRT TV set (Z)	0.958	0.002	0.92	0.958
LCD monitor (1)	0.885	0.016	0.864	0.885
Refrigerator (2)	1.0	0.001	0.992	1.0
Computer printer (3)	1.0	0	1.0	1.0
Water kettle (4)	1.0	0	1.0	1.0
TV receiver (5)	1.0	0	1.0	1.0
Vacuum cleaner (6)	0.941	0.001	0.941	0.941
Lamp (7)	0.854	0.003	0.921	0.854
Weighted average	0.955	0.003	0.953	0.955

Table IV
CONFUSION MATRIX FOR EACH APPLIANCE (25-FOLD CROSS-VALIDATION, RANDOM COMMITTEE ALGORITHM)

	A	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*K*	*L*	*M*	*N*	*O*	*P*	*Q*	*R*	*S*	*T*	*U*	*V*	*W*	*X*	*Y*	*Z*	*1*	*2*	*3*	*4*	*5*	*6*	*7*
A	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
B	0	68	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
C	0	0	50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D	0	0	0	29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	0	0	0	20	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F	0	0	0	0	0	60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
G	0	0	0	0	0	0	65	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	1
H	0	0	0	0	0	0	0	40	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
I	0	0	0	0	0	0	0	0	42	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
J	0	0	0	0	0	0	0	0	0	43	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
K	0	0	0	0	0	0	0	0	0	0	31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	0	1	0	0	0	0	0	0	0	0	0	26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M	0	0	0	0	0	0	0	0	0	0	0	0	34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
N	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
O	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	41	0	0	0	0	0	0	0	0	0	0	0	0	0	0
T	0	0	0	0	0	0	0	2	1	0	0	0	0	0	0	0	0	0	0	38	0	0	0	0	0	0	0	0	0	0	0	1	0
U	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	16	0	0	0	0	0	0	0	0	0	0	0	0
V	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	56	0	0	0	0	0	0	0	0	0	0	0
W	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	22	0	0	0	0	0	0	0	0	0	0
X	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0
Y	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	13	0	0	0	0	0	1
Z	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	23	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	12	0	108	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	126	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	70	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	16	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	16	0
7	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	2	1	0	0	0	0	35

We investigate the occurring confusion in more detail by considering the confusion matrix in Table IV, which shows the numbers of correctly and incorrectly identified traces for each device class. From its visualization, it becomes clear that the matrix is almost diagonal, and that confusion – as determined previously – mainly exists between the CRT and TFT monitors (types Y and *1*). Due to the small number of traces available for some of the appliance types, such as the iron (type X) and the HDTV media center (type N), this analysis cannot be considered fully representative for the given classes, and we will conduct further analyses to improve the empirical soundness. In summary, however, the results achieved by the application of our feature extraction and appliance classification approach show that on average 95.5% of all appliances are correctly identified based on their diurnal power consumption traces.

C. Feature Selection

In order to determine the relevance of the extracted features, we have analyzed their information gain using the Weka toolkit, and list the fifteen most important features in Table V. It becomes clear from the table that average and maximum power consumptions during different periods as well as the durations of these periods comprise more than half of all list entries. After having clustered appliances by their typical power consumption, more specific features begin to gain importance. The energy consumption in different time intervals (cf. ranks 14 and 15) or steps in the power consumption (cf. ranks 10 and 11), hence also appear among the list of features with highest information gain. Only twelve features were attributed no information gain due to the fact that none of the 33 appliance classes in the *tracebase* has, e.g., power consumptions between 3,400 and 3,500 watts.

Table V
LIST OF THE 15 MOST RELEVANT FEATURES ACCORDING TO THEIR INFORMATION GAIN

Rank	Feature description
1	maximum power level in last daily activity phase
2	maximum power level during the complete day
3	average power level during all activity phases
4	average power level for complete day
5	average energy demand per activity phase
6	highest power level during activity phases
7	lowest power level during activity phases
8	average power level using current activity phase
9	median duration of activity phases
10	highest encountered negative power step
11	highest encountered positive power step
12	average value of all peak levels during the complete day
13	magnitude of DC offset of discrete Fourier transform
14	energy consumption between 9.00pm and 9.10pm
15	energy consumption between 9.30pm and 9.40pm

V. Related Work

The classification of appliances or the detection of their presence is generally based on the successful extraction of distinctive features from a power consumption trace. Depending on the location of the sensor, different approaches have been investigated, which we discuss as follows.

Centralized Sensing: Pioneering work for appliance identification was based on single point sensing at the meter level and presented by Hart in [3]. In his NIALM approach, one measuring device is used per household in order to identify when appliances are switched on or off. The identity of an appliance is inferred by matching the encountered change in the household's overall power draw against a database of known signatures. A better classification accuracy is achieved by collecting current and voltage samples at higher sampling rates [10]. Therewith, it is possible to detect previously undetectable changes that correspond to appliance state changes. The method is specifically suited for appliances characterized by significant inrush spikes in power draw during their activation, such as motors. The analysis of the harmonics of the transient power consumption (i.e., during device startup or shutdown) represents another extension to the traditional NIALM method. By isolating each harmonic and analyzing the spectral envelope over a fixed duration during the device startup, distinct features in the transient power waveforms can be found [11], [12]. In addition to the analysis of current transients, the harmonic components of the steady-state current also allow for a classification of the attached appliance [13]. In this case, however, the detection system needs to be trained for all possible combinations of active appliances in order to achieve a high classification accuracy. Finally, voltage signatures have also recently been used to identify consumers based on the presence of noise in the building's electrical installation [14] or magnetic fields emitted by appliances [15], and combinations are aforementioned approaches also exist [16]. A major downside of centralized sensing approaches is their incapability of disaggregating similar devices spatially, e.g., discriminating between lamps of the same type mounted in separate rooms. Similarly, the fact that loads below 150 watts cannot be clearly distinguished [17] represents a major drawback of centralized power sensing approaches.

Circuit-Level Sensing: By measuring power usage at the circuit level, the inability of single point sensing based systems to monitor very small power consuming devices can be tackled [18]. Commonly, there are fewer devices on each circuit, hence a lower occurrence of indistinguishable devices can be expected. Furthermore, high-power devices, e.g., stoves or air conditioning units, are generally connected to dedicated circuits, and thus more sensitive sensors can be employed for monitoring lower-power devices. Apart from the differences in their installation location, the methods outlined in the above paragraph on centralized sensing can also be directly applied at circuit level.

Distributed Direct Sensing: Besides monitoring the activation of appliances at distribution board level, sensor units have also been attached to consumers individually in related research. Despite the higher installation efforts, these distributed sensors bear the potential to both identify and control the plugged-in appliance. The ACme platform [7] and the Plogg [19] are examples of such direct distributed sensing platforms, which monitor the power consumption of the plugged-in appliance. Consumption is reported wirelessly to a base station at an interval of one averaged power reading per minute. Neither the platform nor the overall system however support any automatic appliance classification. Furthermore, an intelligent outlet is presented in [20], which is also based on distributed direct sensing and utilizes a wired communication network to relay the readings to a server.

Our approach of connecting MAUs between the wall outlets and the appliances to identify and control clearly falls into the category of distributed direct sensing. In contrast to existing approaches, however, our implementation permits the extraction of features from sensor data on a local scale (i.e., on the MAU) and to relay only this pre-processed data to the server for classification. Thus, our solution enriches distributed solutions by feature extraction methods well established in the realm of NIALM.

VI. Summary and Conclusions

In this paper, we have presented the *tracebase* repository, hosting a continually growing collection of diurnal power consumption traces collected from individual electrical appliances. Subsequently, we have summarized the considered feature extraction modules, which extract relevant information from each of the traces in order to prepare them for training the classifier. In total, 517 distinct features are being extracted from the data in order to create a model of their behavior. We have evaluated the classification accuracies of different classifiers for 1,197 instances of data and determined the highest accuracy level of 95.5% for the Random Committee classifier. This very high classification accuracy strongly contributes to the applicability of our approach in the envisioned application areas of smart environments, the determination how a building's energy demand is composed, or dynamic load management in the smart grid.

In order to improve our classification accuracy further, we plan to assess the performance when the available features are further enriched by microscopic properties like the frequency spectra of the observed current consumptions. Similarly, instead of connecting MAUs to each appliance individually, we plan to investigate the applicability of the presented approaches on composite power measurements collected from several appliances at the same time.

Acknowledgment

This research has been supported by the LOEWE research initiative of the State of Hesse, Germany, within the Priority Program Cocoon.

References

[1] P3 International Corporation, "Kill A Watt," Available online: http://www.p3international.com/products/special/P4400/P4400-CE.html, 2007.

[2] F. Mattern, T. Staake, and M. Weiss, "ICT for Green – How Computers Can Help Us to Conserve Energy," in *Proceedings of the 1st International Conference on Energy-efficient Computing and Networking (e-Energy)*, 2010.

[3] G. W. Hart, "Residential Energy Monitoring and Computerized Surveillance via Utility Power Flows," *IEEE Technology and Society Magazine*, vol. 8, no. 2, 1989.

[4] "Tracebase: An Online Repository of High-Resolution Appliance Power Traces," Multimedia Communications Lab and Secure Mobile Networking Lab, Technische Universität Darmstadt. Available online: http://www.tracebase.org, 2012.

[5] J. Z. Kolter and M. J. Johnson, "REDD: A Public Data Set for Energy Disaggregation Research," in *Proceedings of the SustKDD Workshop on Data Mining Applications in Sustainability*, 2011.

[6] Plugwise B.V., "Plugwise Circle," Available online: http://www.plugwise.com/idplugtype-f/circle, 2010.

[7] X. Jiang, S. Dawson-Haggerty, P. Dutta, and D. Culler, "Design and Implementation of a High-Fidelity AC Metering Network," in *Proceedings of the 8th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN), Track on Sensor Platforms, Tools, and Design Methods (SPOTS)*, 2009.

[8] M. Damen, "Plugwise Unleashed," Available online: http://www.maartendamen.com/2010/08/plugwise-unleashed-document-released, 2010.

[9] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA Data Mining Software: An Update," *SIGKDD Exploration Newsletter*, vol. 11, no. 1, 2009.

[10] A. Albicki and A. Cole, "Algorithm for Non-intrusive Identification of Residential Appliances," in *Proceedings of the IEEE International Symposium on Circuits and Systems (ISCAS)*, 1998.

[11] C. Laughman, K. Lee, R. Cox, S. Shaw, S. Leeb, L. Norford, and P. Armstrong, "Power Signature Analysis," *IEEE Power and Energy Magazine*, vol. 1, no. 2, 2003.

[12] M. Berges, E. Goldman, H. S. Matthews, and L. Soibelman, "Enhancing Electricity Audits in Residential Buildings with Nonintrusive Load Monitoring," *Journal of Industrial Ecology*, vol. 5, no. 14, 2008.

[13] D. Srinivasan, W. Ng, and A. Liew, "Neural-network-based Signature Recognition for Harmonic Source Identification," *IEEE Transactions on Power Delivery*, vol. 21, no. 1, 2006.

[14] 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 *Proceedings of the 9th International Conference on Ubiquitous Computing (UbiComp)*, 2007.

[15] Y. Kim, T. Schmid, Z. Charbiwala, and M. B. Srivastava, "ViridiScope: Design and Implementation of a Fine Grained Power Monitoring System for Homes," in *Proceedings of the 11th International Conference on Ubiquitous Computing (UbiComp)*, 2009, pp. 245–254.

[16] J. Liang, S. K. K. Ng, G. Kendall, and J. W. M. Cheng, "Load Signature Study — Part I: Basic Concept, Structure, and Methodology," *IEEE Transactions on Power Delivery*, vol. 25, no. 2, 2010.

[17] A. Marchiori, D. Hakkarinen, Q. Han, and L. Earle, "Circuit-Level Load Monitoring for Household Energy Management," *IEEE Pervasive Computing*, vol. 10, no. 1, 2010.

[18] A. Marchiori and Q. Han, "Using Circuit-Level Power Measurements in Household Energy Management Systems," in *Proceedings of the 1st ACM Workshop on Embedded Sensing Systems For Energy-Efficiency In Buildings (BuildSys)*, 2009.

[19] Energy Optimizers Limited, "Plogg Technical Specification," Available online: http://www.plogginternational.com/docs/TechnicalSpecificationV02.pdf, 2010.

[20] T. Saitoh and Y. Aota, "Current Sensor-based Non-intrusive Appliance Recognition for Intelligent Outlet," in *Proceedings of the 23rd International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC)*, 2008.