Influence of Data Granularity on Smart Meter Privacy

Günther Eibl, *Member, IEEE*, and Dominik Engel, *Member, IEEE*

***Abstract*—Through smart metering in the smart grid end-user domain, load profiles are measured per household. Personal data can be inferred from these load profiles by using nonintrusive appliance load monitoring methods, which has led to privacy concerns. Privacy is expected to increase with longer intervals between measurements of load curves. This paper studies the impact of data granularity on edge detection methods, which are the common first step in nonintrusive load monitoring algorithms. It is shown that when the time interval exceeds half the on-time of an appliance, the appliance use detection rate declines. Through a one-versus-rest classification modeling, the ability to detect an appliance's use is evaluated through F-scores. Representing these F-scores visually through a heatmap yields an easily understandable way of presenting potential privacy implications in smart metering to the end-user or other decision makers.**

***Index Terms*—Data granularity, privacy, smart metering.**

I. Introduction

THERE IS a lot of public concern and discussions on the privacy impact of smart metering. However, most discussions take place without knowing the extent of personal information that can be read out of smart meter load profiles. Even more so, there is nearly a complete lack of knowledge on how the amount of personal information relates to the measured time interval, i.e., the time granularity. For example, in many countries in Europe it is planned that smart meters will deliver load data in 15 min time intervals [1]. This has sparked a (sometimes emotional) debate on privacy (see [2]–[4]). However, to our knowledge, no one has tried to assess the amount of personal information that can be extracted on 15 min time interval load profiles, or how, in general, data granularity relates to the amount and nature of extractable personal data.

Although the decrease of the time granularity can be viewed as the most straightforward and simplest privacy enhancing technology—and this method has been suggested by a number of contributions in the past (see [5]–[7])—its impact on privacy has not yet been studied systematically, apart from an initial study we published in [8]. The goal of this paper is making the first step toward a systematic evaluation by studying the impact of time granularity on determination of appliance use. The main reasoning behind this approach is that activities of persons in the house trigger appliances, which in turn sum up to the total load. The activities themselves are influenced by various aspects of personal information such as presence, sleep-wake-cycles, and personal habits

$$\text{Personal Info} \Rightarrow \text{Activities} \Rightarrow \text{Appl. Use} \Rightarrow \text{Load Curve}. \quad (1)$$

This causal chain is the reason why the knowledge of activities leads to knowledge of personal information. As a first step toward a privacy assessment this paper focuses on the detection of appliance use with a short discussion on how activities could be assessed.

Information on appliances is usually extracted from the load data by means of so-called "nonintrusive appliance load monitoring analyzes" (NIALM). There is a lot of literature on NIALM algorithms ([9]–[16], to name a few). The primary goal of these algorithms is the disaggregation of the total load into the individual appliances loads for sake of providing an energy feedback to the end-user. Seen in a different perspective, such NIALM analyzes could also be used as the first step of methods attacking personal privacy by using NIALM as the basis for the extraction of personal information. Instead of using a whole NIALM algorithm as a method for gathering private information, in this paper, a simpler method is used which only uses the first part of typical low-frequency NIALM algorithms, namely edge detection ([2], [9], [11], [12], [14], [16]).

Compared to the large amount of literature aiming at providing energy feedback to the end-user, privacy implications are only rarely treated. In [2], load data were recorded with parallel video data which were processed into activity logs. A NIALM analysis was done yielding the input for subsequent behavior-extraction routines. Extracted behaviors include, e.g., presence, sleep cycles, or meal times. The amount of information disclosure is measured by an overall number called "degree of disclosure." In [4], the load profile is divided into so-called power segments using a density-based clustering technique. These power segments are described by features such as start time, average power, and duration. It is illustrated how such power events could be used for answering several privacy-sensitive questions. In [17], it is shown that under ideal conditions and using small measurement time intervals, even the consumed TV-program can be inferred from load curves.

Manuscript received March 4, 2014; revised July 16, 2014 and September 11, 2014; accepted November 20, 2014. This work was supported in part by the Austrian Federal Ministry of Science, Research, and Economy, and in part by the Austrian National Foundation for Research, Technology, and Development. Paper No. TSG-00197-2014.

The authors are with the Salzburg University of Applied Sciences, Josef Ressel Center for User-Centric Smart Grid Privacy, Security, and Control, Puch/Salzburg A-5412, Austria (e-mail: guenther.eibl@en-trust.at).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSG.2014.2376613

TABLE I
TIME GRANULARITIES OF LOW-FREQUENCY NIALM-STUDIES

Time	1s	3s	15s	20s	1min
Paper	[4], [18], [11], [12]	[14], [15]	[2]	[13]	[16]

This paper is organized as follows. In Section II, NIALM and edge detection methods are reviewed. After the description of the experimental setup in Section III, event detection is applied on the load profiles in Section IV as a method for the extraction of personal information. After this attacking method has been developed, the decrease of time resolution is applied as a countermeasure in Section VI, where the influence of time granularity on the event detection performance is studied. By applying a classification setting, results are described by precision and recall rates which are used as inputs for a systematic privacy analysis in Section VII.

II. BACKGROUND

A. NIALM Analyzes

NIALM analyzes can be broadly divided into two kinds of methods: 1) high frequency; and 2) low frequency methods. High frequency methods look at the waveform of appliances or study transients or higher order harmonics [10]. While the high-frequency methods need a sampling in the range of kHz, the low frequency methods typically analyze load profiles which are sampled using time intervals in the order of seconds to minutes (see Table I).

This paper focuses on low-frequency methods. Particularly, the methods developed here follow the class of supervised NIALM methods [9]. Supervised methods usually consist of several blocks: edge detection, cluster analysis, and finding pairs of on-and-off clusters for the determination of the duration of an appliance. Edges are sharp increases or decreases of the load signal due to turning on or off an appliance. More generally, edges arise due to the change from one state to another state of an appliance when modeled as a finite state machines (FSM). NIALM algorithms commonly use edges instead of the absolute values for two reasons.

1) First, if absolute values were used in the presence of unknown appliances, these appliances would have to be described as a combination of other known appliances.
2) Second, there are adverse cases, where a small change in the measured power would result in a big change in the configuration of used appliances which is not plausible [9].

Although the use of edges is most common other features can be used as well such as the shape features of [4]. A typical assumption in the disaggregation processes is the switch continuity principle which states that in a small time interval only a small number of appliances is expected to change the state [9]. Often, this assumption is tightened by requiring that in a time interval at most one appliance changes its state (one-at-a-time condition).

The usual performance measures of NIALM methods are the error in the total energy assigned to a given appliance or the error in the estimated on-time. Event-based methods state the performance in terms of precision p and recall r [2], [14] or the F-score [15]. Precision p is the proportion of events classified as stemming from appliance A which is really stemming from appliance A. Recall is the proportion of all events stemming from appliance A that is also classified as stemming from appliance A. Performance is either given by the pair (p, r), or if a single performance number is needed by the F-score F

$$F = 2\frac{p \cdot r}{p + r}. \tag{2}$$

B. Event Detection Methods

In this section, event detection methods are reviewed. The main assumption is the validity of modeling appliances as FSM having different power values for different states. An edge or event $e = (t_e, \Delta P_e)$ is a transition between two such states. It is represented by the onset time t_e and a transition value ΔP_e, which is the difference in power levels of the two states. Events with increasing signal ($\Delta P > 0$) are called on-events because they typically arise from turning on an on–off-appliance. Analogously, events with $\Delta P < 0$ are called off-events.

The most straightforward edge detection method, called difference method, detects an edge, if the difference $\Delta P_i = P_{i+1} - P_i$ between consecutive power values exceeds a threshold. Each detected edge is considered to be an event $e = (t_i, \Delta P_i)$. If the transition between two levels needs several time intervals, the method divides the transition between two levels in several edges having smaller values than the transition.

Due to this drawback, the edge merging method merges subsequently occurring edges into a single event [12]. The value of the event is the sum of the individual edge values, which can be both positive and negative. The time where the event occurs is defined as the onset time, i.e., the time of the first edge contributing to the event.

While the previous two methods focus on the transition between two levels of a signal, the next method focuses on the power levels of the two transition states. The method was proposed in [9], where it is called transient passing method for edge detection. A transition is inversely defined as being not a steady subsequence. In the first step the method finds the steady subsequences of the signal. This is done using a sliding window approach where a subsequence consisting of n points is considered as steady, if the range of its values does not exceed a given threshold. As a result, the whole signal is divided into consecutive steady parts st and unsteady transitions tr. For the description of the event e arising from transition tr_i the three subsequences (st_{i-1}, tr_i, st_i) are considered. The onset-time t_e for the description of the event is the last time point of the first steady part st_{i-1}. The transition value ΔP_e is the difference between the median of the values of the second steady part st_{i+1} and the median of the values of the first steady part st_{i-1}. Taking the median value over the whole steady subsequences increases the robustness of the event value ΔP_e.

In order to account for noise, for all methods, events e with a value ΔP smaller than a specified threshold are discarded.

III. Experimental Setup

In this section, the method that extracts personal information is described, decreasing granularity as a method for preserving privacy is briefly discussed and the used dataset is introduced.

A. Assessment of Appliance Use by Edge Detection

The goal of NIALM algorithms is energy disaggregation, which means that the interest lies in partitioning the consumed power into the portions used by individual appliances. In order to accurately measure the energy used by an appliance, the on-duration T_{on} of an appliance needs to be assessed precisely.

However, from a privacy viewpoint it is not necessarily important to assess the energy used by an appliance. In a privacy attack setting, the ultimate goal is the determination of private information like habits, personal properties or special circumstances. Since this information is typically not known in common data sets (including the REDD data set used here) this paper focuses on the determination of appliances together with a simple determination of activities within a household according to the causal chain (6). Regularly occurring activities could in turn provide information about e.g., habits, but such a study is out of scope of this paper. Here, the kind of an activity is inferred from the appliances that are used. For example, the activity cooking is inferred from the use of any one of the appliances stove, oven1, oven2, and microwave (compare also Table II).

The other important information about an activity is the usage time. For the description of an activity and possible inference of habits it is important when it takes place. Here, edge detection can provide the onset of an activity by providing the starting time of the corresponding appliance.

The second information is the duration of an activity or an appliance. The information about the duration could provide further information like e.g., the kind of meal that is cooked. Since no ground truth about activities is available and especially no details are known, it was decided not to assess the exact duration of an appliance. Moreover, initial trials showed that the matching of on- and off-events is far from being straightforward and would possibly limit the validity of results. Note that the matching applied by NIALM-algorithms for obtaining the on-durations needed for the assessment of the total energy used by an appliance is typically quite complicated. In order to keep the assessment clear and simple, it was decided to avoid the matching procedure. Instead, the on-duration of an appliance is simply measured as the time until the next off-event of this appliance occurs. Thus, typical on-durations of appliances are provided for the explanation of results in Section VI-B. However, the on-durations are never used for any other use including the determination of activities. Note that for FSMs the term "duration of stay in the present state" would be a more adequate name.

Since the signals of the available REDD-dataset [19] not only contain the mains but also the signals of the individual appliances, it is not necessary to compute the whole disaggregation. Instead, the following analysis focuses on the determination of events, which can directly be done using the edge detection methods of Section II-B.

B. Decreasing Time Granularity for Privacy Enhancement

Several possibilities for decreasing the time granularity exist. Considering a single time interval, different statistics could be computed. The most straightforward statistic is the average load value which should suffice for most practical solutions such as standard billing or time-of-use billing. For pricing based on the maximum load or for control reasons, the maximum load needed during the time interval could be useful. Additionally, (uniform) sampling could be done, i.e., taking the load value at (evenly) spaced points in time.

In the experiments presented in this paper, three variants are used: taking the average and maximum load in a time interval, respectively, and uniform sampling.

C. Dataset

All experiments were done using the so-called low-frequency dataset of the publicly available REDD-dataset [19]. The dataset contains measurements of the apparent power for six different houses. Measurements are available for mains1 and mains2, for some circuits, e.g., kitchen outlets and for individual appliances.

Although the analyzes were performed for all six houses, the evaluation is shown for house 1 only. House 1 has a relatively high number of measured appliances or circuits and includes labeled measurements both for high and low power appliances. The overlap of the power values of individual appliances is rather low, so that a possible increase in the overlap due to lower time resolutions could be detected.

One of the kitchen outlets, one of the washer dryers and the electric heat appliances showed less than three events at the highest time resolution and were excluded for further analyzes. Due to its automatic working mode, in the privacy attack setting the refrigerator is more a disturbing noise appliance than a privacy relevant appliance. A mains appliance was created as the sum of mains1 and mains2 by interpolating the values of mains2 to the values mains1 at the highest time granularity.

IV. Descriptive Event Detection Results

In this section, event detection is applied to the load curves of individual appliances. First, the quality of different event detection methods is assessed (Section IV-A), then it is shown how the overlap of events of different appliances affects the precision of subsequent classification algorithms (Section IV-B).

A. Comparison of Event Detection Methods

Since the results below are based on the events found, the performance of the event detection methods is assessed first. The evaluation is mainly done visually.

Generally, transient passing and edge merging yield good and very similar results (upper panel of Fig. 1). Note that the load curve is quite complex, especially power levels are not necessarily constant. As expected, the simple difference method yields more, but disturbing events and can therefore not be recommended as is (upper panel of Fig. 1).

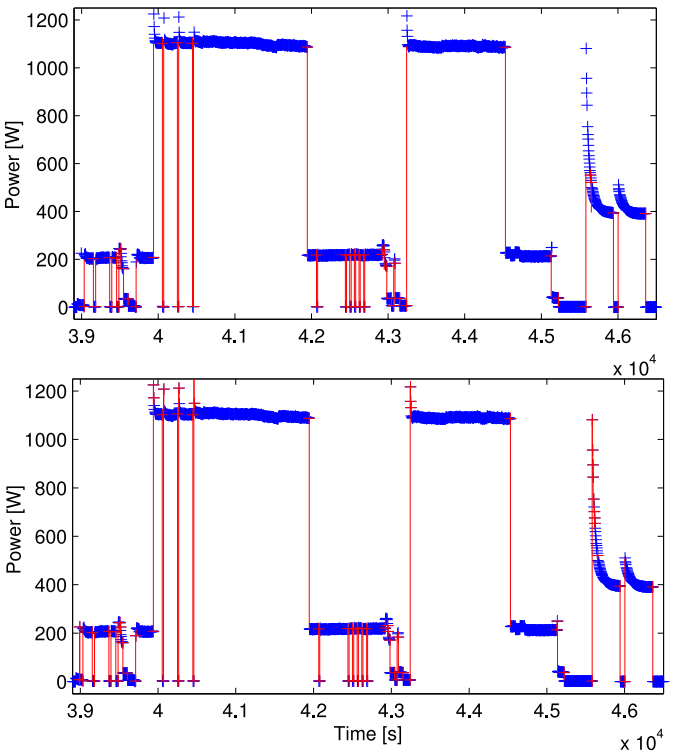

Fig. 1. Dishwasher events, marked as "+," detected using the transient passing method (upper panel) and the difference method (lower panel).

High-power devices such as heating are usually purely ohmic and consume high power values with a greater deviation of values. For low-power devices such as lighting the deviation of values is smaller. This leads to a tradeoff between noise removal and detection of events. If the noise threshold is set too low, the noise of high-power devices exceeds the threshold resulting in additional, unwanted edges. A noise threshold that is set too high in turn leads to a loss of events for low power devices. For all subsequent evaluations we used 20 W as noise threshold.

While for high time resolutions the edge merging and transient passing methods give very similar results, for lower time resolutions the transient passing method is more robust in determining the edge values. The results of the transient passing and edge merging methods turned out to be quite insensitive to the kind of statistic. For lower time resolutions the performance of the difference method is better with taking the max statistic or with sampling than with taking the average statistic. If not stated otherwise, the remaining analyzes will use the transient passing method and the average statistic.

B. Description of Events

This subsection contains a visual description of the events that occurred. Since the mains signal—which was generated by summing up the mains1 and mains2 signals—is supposed to contain the events of all appliances, the time between subsequent events is smaller than for the events of a single appliance. As a check that this property does not negatively influence the event detection of mains, the events of mains and the events of the individual appliances are compared in

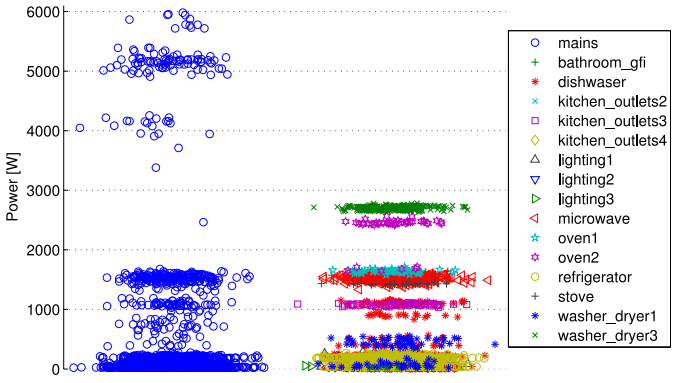

Fig. 2. Events of mains (left) compared with single appliances' events (right).

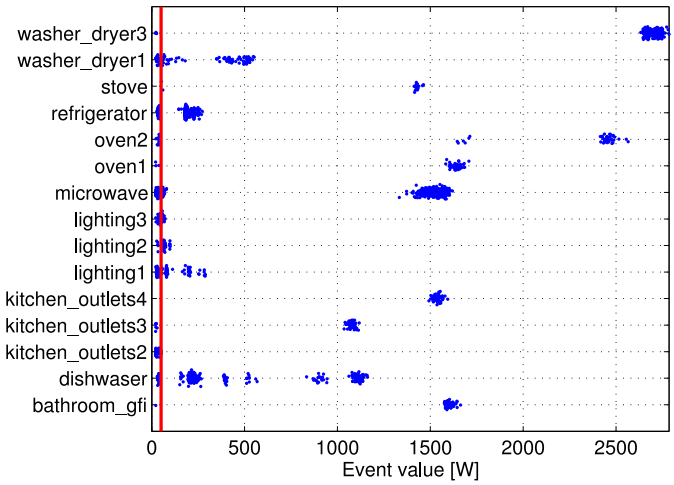

Fig. 3. Overlap of events for house 1.

Fig. 2. In fact, there is clear connection between events of the mains signal and the events of the individual appliances.

However, there are events that only occur for mains but not in any of the single appliances signals (Fig. 2). Additionally, there are events from appliances that do not occur in the mains signal (Fig. 2). This happens for all houses. To rule out that this could be an effect of bad event detection, both the absence of the appliance events and the presence of additional mains events was verified by visual inspection of the load curves. Due to this inconsistency of the mains signals and the signals of single appliances it was decided that all further analysis steps should be done with the load curves of the individual appliances only ignoring the mains signal.

Analyzing the quality of edge detection, for some high-power appliances unwanted noise events below 50 W are detected. Events below 50 W (left to the red, dashed line in Fig. 3) are considered as being hard to assign to appliances due to the high overlap of several appliances within this region. Therefore these events are discarded for further evaluation.

Even without performing a NIALM-analysis, the overlap of events stemming from different appliances can give valuable insights into the possibilities of disaggregation of the mains signal (Fig. 3). Appliances whose events have low overlap with other appliances' events, like e.g., washer_dryer3 will be easier to distinguish from them than appliances with high overlap

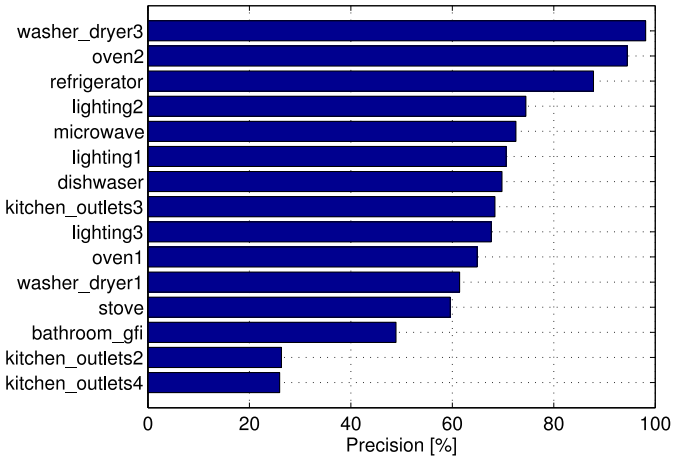

Fig. 4. Precision at highest resolution (3 s).

such as e.g., kitchen_outlets4 (compare also with Fig. 4). It should be noted that the negative events of an appliance typically have the same absolute values as the positive events, thus only the positive events are shown here.

V. Evaluation of Appliance Determination Ability

According to the causal chain (1), the first step in the determination of private information is considered, i.e., the ability to determine appliance use is evaluated

$$\text{Load Curve} \Rightarrow \text{Appliance Use.} \tag{3}$$

Note that the subsequent analysis models the detection of a given appliance. Due to the reasoning stated in Section III-A the assessment of the on-duration of appliances is not evaluated.

A. Classification Method

The chosen methodology can be identified more clearly, when the problem is stated in another form. Considering a detected event, one wishes to know, which appliance this event is stemming from. This is exactly a multiclass classification problem where the number of classes is the number of appliances. This multiclass classification problem is split into several one versus all two-class classification problems, one classification problem for each appliance. The input is the 1-D value of the event to be classified, the output is the information, if the event is stemming from this appliance or from one of the other appliances. Due to this setting, natural measures for appliance detection performance are precision and recall of classification. If a single performance value is required, the F-score (2) can be used. In contrast to a normal classification scenario where a good performance is requested, here small values are desirable with respect to privacy preservation.

It is expected that the overlap affects the precision of the classification task. Appliances with negligible overlap of their event values with event values of other appliances, such as washer_dryer3 and oven2, are expected to lead to simple classification problems with high precision. The precision of the corresponding classification problem is expected to decrease with increasing overlap.

Of course, more sophisticated analyzes could be done exploiting, e.g., the periodicity of the refrigerator or the typical duration between events of the appliance [15]. The information about the time of the day when the appliance was used could be taken into account [15], too. A dishwasher run consists of a series of events with different event values. The fact that different runs all look very similar to the time pattern of events shown in Fig. 1 could be exploited as follows. Event values of kitchen_outlets3 have similar values as one particular level of the dishwasher values (Fig. 3). Looking at the statistics of events over some past time window, if some other event values of the dishwasher do not occur, the dishwasher could be ruled out and thus kitchen_outlets3 could be distinguished from dishwasher. The same argument could be applied to washer_dryer1 and the dishwasher. However, such a detailed analysis is not the scope of this paper.

For the sake of simplicity, as classification algorithm the nearest neighbor method using three nearest neighbors is used. The resulting precision of the several two-class classification tasks for the highest time resolution is shown in Fig. 4. Precision is typically in the range between 60% and 80% with a maximum precision for washer_dryer3 of nearly 100%. By comparing Figs. 3 and 4, the negative influence of the overlap with events from other appliances on the precision is evident.

Note that here no direct NIALM analysis was done. Instead, only the event-values of the individual appliances (or circuits) are directly taken in order to analyze possible NIALM performance. The result can be used for an optimistic (in the sense of precision) estimate for the precision of a NIALM analysis, if several assumptions hold. The first assumption requires that the mains signal is the sum of the individual appliances loads plus a possible constant offset value which has no influence on events. Secondly, the noise must be of equal size both for all individual appliances and for the mains signal. Thirdly, and most importantly, the one-at-a-time condition which is a special form of the switch continuity principle [9] is assumed to be fulfilled. This condition states, that during each time interval at most one of the appliances changes its state.

B. Method Evaluating the One-at-a-Time Condition

The one-at-a-time condition is already known as a common necessary condition for some NIALM algorithms [9]. When more than one appliance change their state the edges of the aggregated signal are the sum of the individual edges. This leads to a much bigger search space of possible solutions which must be handled by the NIALM algorithm. Additionally, when more than one combination of appliances have the same aggregate edge value, ambiguities arise.

The classification method above looks at the signals of single appliances. Consequently, the one-at-a-time condition is ignored. The information about each appliance is obtained by separately applying the edge detection algorithm on the signal of each single appliance. However, in a usual setting, only the aggregate signal is given, thus hardening the disaggregation problem. The one-at-a-time condition suggests that a change

of an appliances' state can only be detected, if only this single appliance changes its state during the measurement interval. For the assessment of the one-at-a-time condition, for each event found, it is checked, if this is the case or not.

First, the edges are computed from the individual signals of all appliances at the highest time resolution available and all event times are evaluated. An event is the only event within a measurement interval, if the duration to both the previous and the next event time exceeds the measurement interval. If the smaller duration is less than the measurement interval an event can be classified as single event, otherwise an event is classified as a coincidental event. As a performance measure now the proportion of single events for each appliance and measurement interval is calculated. Also here, small values are desirable with respect to privacy preservation.

VI. Influence of Time Granularity on Appliance Classification

In this section, the influence of time granularity Δt on precision and recall of the classification method shown above is studied.

A. Influence of Time Granularity on Recall

In a normal NIALM classification setting, the recall of a given appliance is defined as the proportion of events stemming from this appliance that can be found in the aggregate signal. However, due to the unknown differences between the mains signals and the signals of the individual appliances signals (Fig. 2), it was decided not to use the aggregate signal. As a consequence, the recall cannot be evaluated directly. In order to assess a quantity similar to the recall rate, the numbers of detected events of an appliance are compared for different time resolutions. Considering the events found at the highest resolution as ground truth, the number of events found at different time granularities can be normalized by this ground truth. Since the goal of this paper is studying the changes that arise due to changes in time resolution, this normalized number of events sufficiently serves as a measure of the recall rate. This measure for the recall is too optimistic because it is assumed that the recall at the highest resolution is 100% and the events of the appliances are found from the appliances signals instead of the mains signal. This overestimated recall measure goes down to near zero with decreasing granularity (Fig. 5) which is sufficient for a decrease of the exact recall rate.

In the privacy setting, the decrease of the recall to near zero means that with the time interval exceeding an appliance-specific threshold, a device will not be detected any more. Undetectability of devices in turn increases privacy.

B. Influence of On-Duration on the Recall

Fig. 5 shows that the recall of the appliances decreases with increasing measurement interval Δt. The measurement interval Δt_{drop} where this decrease takes place differs among appliances. This appliance-dependent quantity is denoted as drop-time. This subsection shows that the property of the

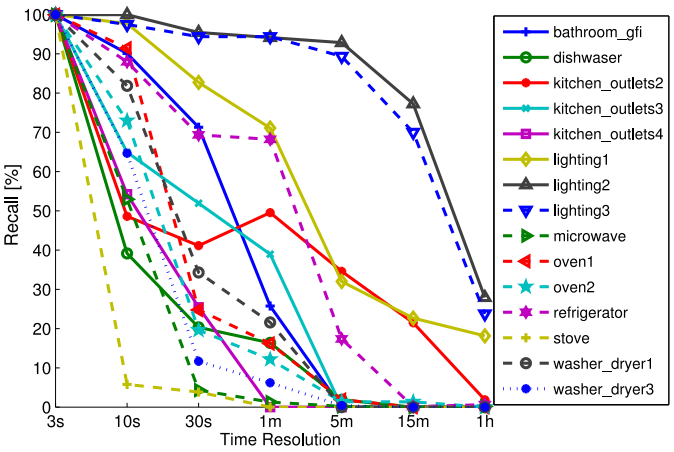

Fig. 5. Recall dependent on time granularity.

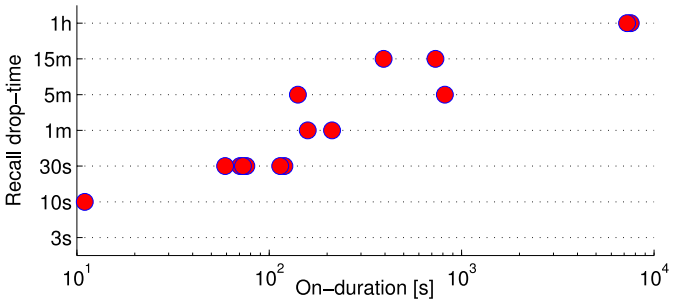

Fig. 6. Drop time Δt_{drop} and median on-durations of different appliances.

appliances by which this critical duration is influenced is the on-duration T_{on}.

For an experimental assessment of this influence, for each appliance the drop-time Δt_{drop} of the recall is assessed as the time granularity where the recall in Fig. 5 is below 30% at first time. The value 30% was chosen for making Δt_{drop} robust against false positive events. A comparison of the obtained recall drop-time Δt_{drop} and the on-duration of the appliances in Fig. 6 shows a clear increase in drop time with increasing on-duration.

The connection between the on-duration and the drop of the recall can be explained by the mechanism of the transient passing method applied to a simple on–off-appliance with fixed on-duration T_{on}. For ease of explanation sampling of values is assumed. The transient passing method detects an on-state as a steady sequence of at least n values with higher energy consumption. As in [9], in this paper, n is set to 3 which is one of the smallest possible choices for n having thus a good detection property with reasonable robustness. If the on-duration T_{on} is too small, $T_{\text{on}} < (n-1)\Delta t = 2\Delta t$, at most two subsequent values can have higher loads which is just not enough to detect the on-state. Consequently, no change from or to the on-state can be detected. Rewriting this condition, the recall rate should drop to zero, if the time interval Δt exceeds a threshold which depends on the on-duration

$$\Delta t > \Delta t_{\text{drop,ideal}}(T_{\text{on}}) = \frac{T_{\text{on}}}{2}. \tag{4}$$

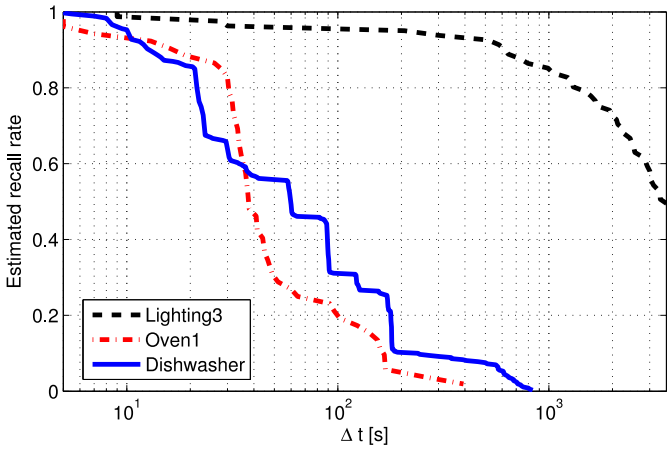

Fig. 7. Recall of 3 different appliances, estimated by the rule of thumb (5).

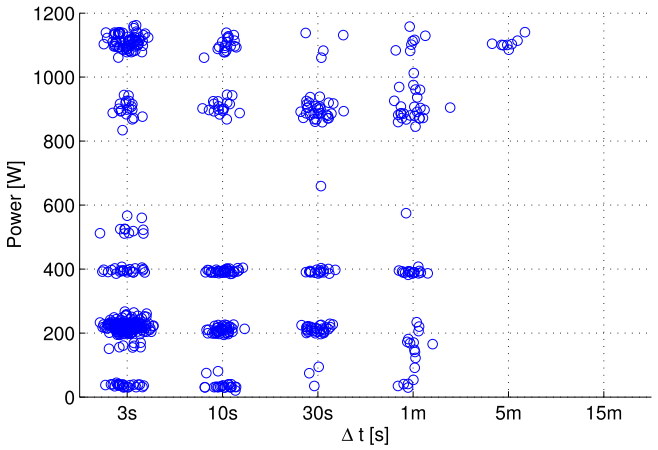

Fig. 8. Robustness of dishwasher event values when determined with transient passing edge detection.

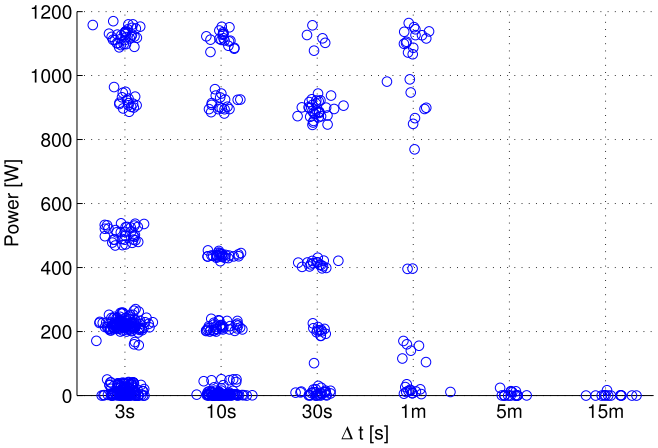

Fig. 9. Dishwasher event values determined with the edge merging method.

Using this connection, the knowledge about the on-duration of an appliance—which is often available as an initial guess without any NIALM-like analyzes—can be used for estimating the time interval Δt needed to significantly decrease the recall rate. If the time interval exceeds half of the typical on-duration of an appliance, a considerable proportion of events stemming from this appliance cannot be detected any more. Using the cumulative distribution of on-durations $F(T_{\text{on}})$, this rule can be formalized: dependent on the measurement interval Δt, an approximation for the recall rate $R(\Delta t)$ can be calculated as

$$R(\Delta t) = 1 - F(\Delta t/2). \tag{5}$$

This estimated recall rate of events is illustrated in Fig. 7 for lighting3, oven1 and the dishwasher. Despite the different choice of x-axes a strong similarity to Fig. 6 can be noticed. Due to the long on-durations, lighting3 exhibits high recall rates. The different on-durations of the dishwasher-states result in a staircase-like recall-curve.

C. Influence of Time Granularity on Precision

After studying the influence of the time resolution on the recall rate in Section VI-A, now the precision for the remaining events of the remaining appliances is investigated.

Interestingly, for increasing time interval Δt the precision for the classification of the remaining events keeps being high. This behavior is illustrated for house 1 and a time interval of 15 min. Due to the low recall, only four out of 15 appliances/circuits are still detectable. The precision of classification for these four remaining appliances is even higher than for the highest time resolution. One reason for this behavior is that a four-class classification problem is much simpler than a 15-class classification problem.

Another prerequisite for this behavior is the surprisingly robust estimation of the event values which is exemplarily shown for the dishwasher in Fig. 8. This stability property only holds for the transient passing method. For the edge merging method event values are relatively stable but show a slight decrease of event values (Fig. 9) while for the difference method event values get smeared for decreased time resolution (not shown). The amount of smearing for the difference method is most pronounced for the averaging-statistic.

VII. Understandable Privacy Analysis

This section aims at presenting the results about the influence of time granularity. As an important requirement, these results should be easily understandable and thus be suitable for unexperienced people like end-users or other decision makers. The influence of the time resolution is discussed in two parts: 1) the first part shows the influence on appliance use detection and 2) the second part shows the influence on higher-level personal information.

A. Detection of Appliance Use

An appliance can provide insights into personal information only if it can be detected and if the precision of detection is high. An appliance with these two properties will be called measurable. Measurability of an appliance itself does not necessarily imply danger for privacy, because appliances that are automatically controlled such as the refrigerator do not provide personal information even if their operational states are known. In contrast, nonmeasurability does imply privacy-safety which is the property that should be assessed here. Measurability

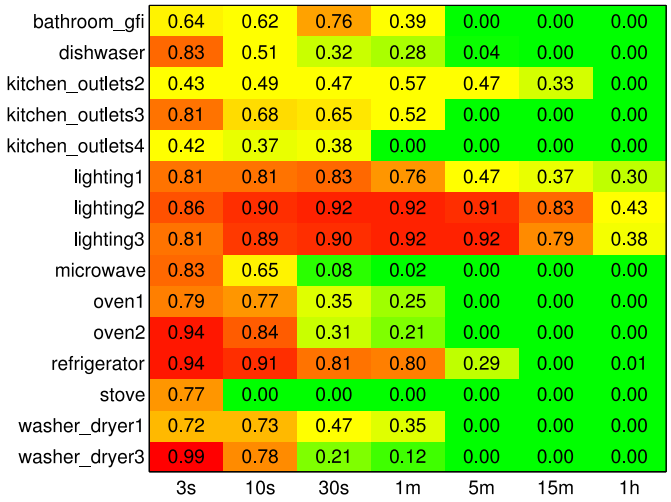

Fig. 10. F-score matrix. Small values are desirable for privacy.

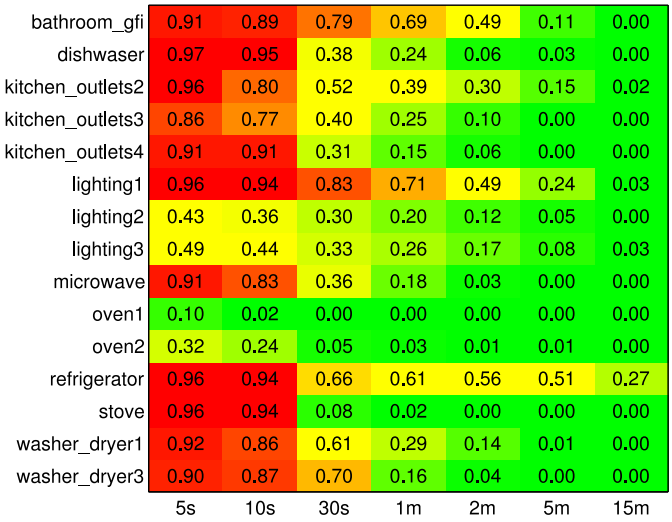

Fig. 11. Proportion of single events. Small values are desirable for privacy.

can be assessed using the commonly used F-score which is computed from recall r and precision p by (2). Arranging the F-scores for all appliances and all time resolutions the resulting matrix can be visualized by a heatmap as shown in Fig. 10. There, the privacy-harmless appliances having low F-scores are colored green, while measurable and thus potentially privacy-decreasing appliances having high F-scores are colored red or orange.

The visualized F-score matrix (Fig. 10) clearly shows that measurability decreases and consequently privacy increases with increasing time interval.

Interestingly, measurability not necessarily decreases with increasing time interval. For example the F-score of appliance bathroom_gfi is maximal at a time interval of 30 s. This behavior can be explained by the high overlap of its event values with the events values of the appliances microwave, oven1, oven2, and kitchen-outlets4 (Fig. 3). This overlap leads to a rather small precision and consequently small F-scores of bathroom_gfi at high time resolution. However, the other appliances have shorter on-durations than bathroom_gfi. The short on-duration leads to a sharp drop of their recall at a time interval of 30 s. For bathroom_gfi the drop at the 30 s interval is relatively small, the sharp drop occurs later at a time interval of 1 min (Fig. 5). Thus, since the masking events of the other appliances are not present at a 30 s interval the precision of bathroom_gfi increases from 47% at 10 s intervals to 81% at 30 s intervals. This increase in precision overcompensates the drop in recall from 90% to 71% leading to an increased F-score (from 0.62 to 0.76) and thus explaining why bathroom activities are only measurable at 30 s time intervals.

Assessing appliance use with the one-at-a-time condition method shows that the proportion of single events decreases with increasing time interval (Fig. 11) which again implies an increase of privacy.

Comparing the results of the two evaluation methods shows a similar behavior. The only big differences can be seen for lighting1 and lighting2 which look much more privacy-safe when evaluated by the one-at-a-time condition method. This increase in privacy compared to the F-score assessment can be explained by the fact, that this method considers all appliances at once instead of just a single appliance. For the chosen house lighting1 and lighting2 are strongly co-occurring, therefore the proportion of single lighting events is small already at a very fine time resolution. This dependence of appliances can not be modeled with the classification method which looks only at the event values and not at the event time.

B. Detection of Activities

Now, according to the causal chain (1), higher level privacy implications of the resulting matrices are illustrated

$$\text{Appliance Use} \Rightarrow \text{Activities, Presence/Absence.} \quad (6)$$

For ease of explanation, a privacy-threshold of 0.7 is introduced. Entries with higher values are classified as measurable, entries with lower values as unmeasurable. Thus, red or orange entries are regarded as privacy-relevant while green or yellow entries are regarded as privacy-safe.

Looking at the F-score matrix, for 1 h time intervals all appliances are privacy-safe. For a 1 min time interval only the lights are privacy-relevant (because of its automatic operation mode the refrigerator is regarded as safe in this analysis). Interestingly, increasing the time interval from 1 to 5 or 15 min only negligibly increases privacy here. Bathroom activities (bathroom_gfi) are only measurable at exactly 30 s time intervals. Cooking (stove, oven1, oven2, and microwave) and housework (washer-dryer and dishwasher) are privacy-safe for time intervals of 30 s or more. It should be noted that the kitchen outlets were not considered for this analysis due to the unclear nature of the corresponding appliances. The result of this short discussion is shown in Table II.

Considering the one-at-a-time condition evaluation method, already at a measurement interval of 2 min, all appliances are privacy-safe. As before, the increase in privacy compared to the F-score assessment can be explained by the co-occurrence of lighting1 and lighting2.

The results of Tables II and III should be seen as a first evaluation of privacy that is likely to be too optimistic. On one

TABLE II
TIME INTERVAL Δt NEEDED TO INFER DIFFERENT KINDS OF PERSONAL INFORMATION FOR HOUSE 1 USING THE F-SCORE, THRESHOLD 0.7

Information	Inferred from	Safe for
Presence/Absence	*Lighting*	$\Delta t \geq 1h$
Bathroom Activities	*bathroom-gfi*	$\Delta t \neq 30s$
Cooking	*stove, oven1, oven2, microwave*	$\Delta t \geq 30s$
Housework	*washer-dryer, dishwasher*	$\Delta t \geq 30s$

TABLE III
TIME INTERVAL Δt NEEDED TO INFER DIFFERENT KINDS OF PERSONAL INFORMATION FOR HOUSE 1 USING THE ONE-AT-A-TIME CONDITION, THRESHOLD 0.7

Information	Inferred from	Safe for
Presence/Absence	*Lighting*	$\Delta t \geq 2m$
Bathroom Activities	*bathroom-gfi*	$\Delta t \geq 1m$
Cooking	*stove, oven1, oven2, microwave*	$\Delta t \geq 30s$
Housework	*washer-dryer, dishwasher*	$\Delta t \geq 1m$

hand, this privacy analysis is based on the effect of an increased measurement interval on event detection. While fine-grained personal information is likely to be based on appliance events, it seems plausible that coarse information such as presence or absence could easily be found using other methods. Such methods could for example examine the difference in average power consumption for times where the inhabitants are present or absent. For the detection of certain activities it could be sufficient to distinguish different groups of appliances such as appliances used for cooking.

On the other hand, the choice of the value 0.7 as the privacy-threshold is quite arbitrary and mainly intended for demonstrating the privacy evaluation. Choosing this value as a threshold for the F-score, an appliance is considered measurable, if nearly each single event can be detected and distinguished from other appliances events. However, for the detection of regular personal habits it is not necessary to detect each single event, it is rather necessary to detect enough events during the recording time. Having data for long durations such as years, a lower recall rate could be considered privacy-relevant leading in turn to a lower acceptable F-score privacy-threshold. Looking at a thought experiment of an appliance used twice a day and a measurement duration of three years leads to approximately 300 events. Even one-third of these events would be enough to estimate typical usage times.

The privacy-threshold should also be chosen separately for each appliance. For example, one run of a dishwasher leads to many events. Although for a time interval of 30 s the F-score goes down to 0.32 (Fig. 10), the main big events are still detectable at this time granularity (Fig. 12, upper panel) suggesting that a lower threshold is needed for the dishwasher. Averaging over a 5 min interval, only one edge is left (Fig. 12, lower panel), using a 15 min interval, also this last event can not be detected any more suggesting for the dishwasher an F-score threshold of 0.04 or less. Despite these open issues, the usefulness of the performed evaluations for a first assessment of the impact of time granularity on personal information could be shown.

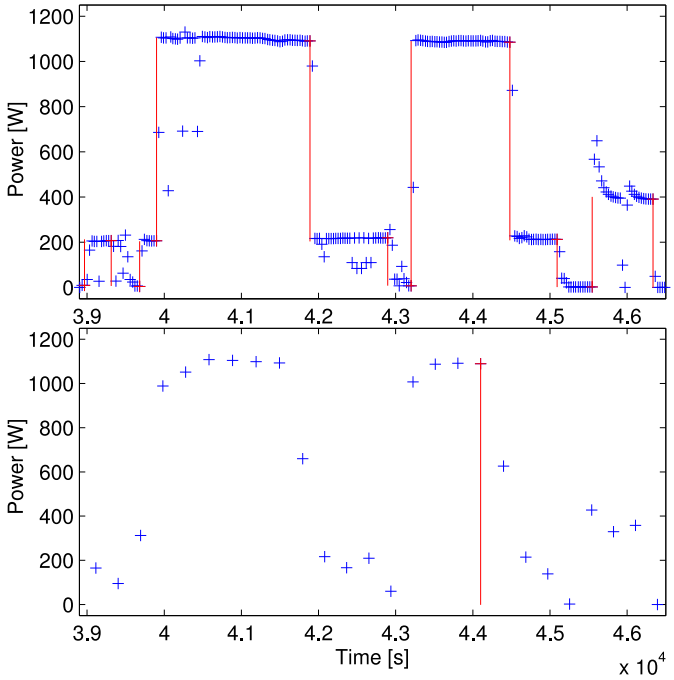

Fig. 12. Edges for dishwasher for $\Delta t = 30$ s (upper panel) and 300 s (lower panel).

VIII. CONCLUSION

Although being the simplest possible privacy enhancing technique, the impact of decreasing the time resolution on privacy analyzes of load signals obtained from smart metering to date has not been studied systematically. Since the first step in a privacy attack can consist of the assessment of appliance use which is in turn often based on edge detection methods, the influence of the time interval on edge detection methods applied on load signals is studied.

Using edge detection alone already leads to valuable insights about the disaggregation possibilities for different appliances, a full NIALM-analysis is not necessary. Appliances whose events have a small overlap with the events of the other appliances can more easily be disaggregated.

With increasing time interval, the recall, i.e., the proportion of detected edges stemming from a device decreases. This decrease is more pronounced for appliances with shorter on-durations. As a coarse rule of thumb, when the time interval exceeds half the typical on-duration of an appliance, the appliances event values cannot be reliably detected any more. For the house analyzed in detail, increasing the measurement interval to 15 min has the effect that only four out of 15 appliances/circuits remain detectable (three lighting circuits and the refrigerator). For these remaining appliances the disaggregation precision stays high, because even for high time intervals the transient passing edge detection method robustly determines edge values.

Privacy implications can be evaluated by F-score values or the proportion of single events of an appliance. Evaluating these values for different appliances and time granularities,

the resulting matrices can be visualized. This visualization represents the impact of time granularity on privacy in an easily understandable way suited for nonexperts like the users themselves or other decision makers.

For the next natural steps toward privacy evaluation datasets that include personal information or activity logs are needed enabling a more direct assessment of personal information. Such data would be the basis for finding a well-founded way of choosing privacy-thresholds, an evaluation method that combine the two methods proposed here or other methods especially designed for low measurement intervals.

REFERENCES

[1] R. Segovia and M. Sánchez, "Set of common functional requirements of the smart meter," DG INFSO and DG ENER, European Commission, Brussels, Belgium, Tech. Rep. 73, Oct. 2011. [Online]. Available: http://ec.europa.eu/energy/gas_electricity/smartgrids/doc/2011_10_smart_meter_funtionalities_report_full.pdf

[2] M. Lisovich, D. Mulligan, and S. Wicker, "Inferring personal information from demand-response systems," *IEEE Security Privacy*, vol. 8, no. 1, pp. 11–20, Jan./Feb. 2010.

[3] A. Cavoukian, J. Polonetsky, and C. Wolf, "SmartPrivacy for the smart grid: Embedding privacy into the design of electricity conservation," *Identity Inf. Soc.*, vol. 3, no. 2, pp. 275–294, 2010. [Online]. Available: http://dx.doi.org/10.1007/s12394-010-0046-y

[4] A. Molina-Markham, P. Shenoy, K. Fu, E. Cecchet, and D. Irwin, "Private memoirs of a smart meter," in *Proc. 2nd ACM Workshop Embedded Sens. Syst. Energy-Eff. Build. (BuildSys)*, New York, NY, USA, 2010, pp. 61–66. [Online]. Available: http://doi.acm.org/10.1145/1878431.1878446

[5] C. Efthymiou and G. Kalogridis, "Smart grid privacy via anonymization of smart metering data," in *Proc. 1st IEEE Int. Conf. Smart Grid Commun.*, Gaithersburg, MD, USA, Oct. 2010, pp. 238–243.

[6] D. Engel, "Wavelet-based load profile representation for smart meter privacy," in *Proc. IEEE PES Innov. Smart Grid Technol. (ISGT)*, Washington, DC, USA, Feb. 2013, pp. 1–6. [Online]. Available: http://dx.doi.org/10.1109/ISGT.2013.6497835

[7] D. Engel and G. Eibl, "Multi-resolution load curve representation with privacy-preserving aggregation," in *Proc. IEEE Innov. Smart Grid Technol. (ISGT)*, Copenhagen, Denmark, Oct. 2013, pp. 1–5.

[8] G. Eibl and D. Engel, "Influence of data granularity on nonintrusive appliance load monitoring," in *Proc. 2nd ACM Workshop Inf. Hiding Multimedia Security*, Salzburg, Austria, 2014, pp. 147–151. [Online]. Available: http://doi.acm.org/10.1145/2600918.2600920

[9] G. Hart, "Nonintrusive appliance load monitoring," *Proc. IEEE*, vol. 80, no. 12, pp. 1870–1891, Dec. 1992.

[10] M. Zeifman and K. Roth, "Nonintrusive appliance load monitoring: Review and outlook," *IEEE Trans. Consum. Electron.*, vol. 57, no. 1, pp. 76–84, Feb. 2011.

[11] D. C. Bergman *et al.*, "Distributed non-intrusive load monitoring," in *Proc. IEEE/PES Conf. Innov. Smart Grid Technol. (ISGT)*, Anaheim, CA, USA, Jan. 2011, pp. 1–8.

[12] M. Baranski and J. Voss, "Genetic algorithm for pattern detection in NIALM systems," in *Proc. IEEE Int. Conf. Syst. Man Cybern.*, The Hague, The Netherlands, 2004, pp. 3462–3468.

[13] E. Vogiatzis, G. Kalogridis, and S. Z. Denic, "Real-time and low cost energy disaggregation of coarse meter data," in *Proc. 4th IEEE PES Innov. Smart Grid Technol. Europe (ISGT Europe)*, Lyngby, Denmark, 2013, pp. 1–5.

[14] J. Z. Kolter and T. Jaakkola, "Approximate inference in additive factorial HMMs with application to energy disaggregation," *J. Mach. Learn. Res. Proc. Track*, vol. 22, pp. 1472–1482, Apr. 2012.

[15] H. Kim, M. Marwah, M. Arlitt, G. Lyon, and J. Han, "Unsupervised disaggregation of low frequency power measurements," in *Proc. 11th SIAM Int. Conf. Data Min.*, Mesa, AZ, USA, 2011, pp. 747–758.

[16] O. Parson, S. Ghosh, M. Weal, and A. Rogers, "Non-intrusive load monitoring using prior models of general appliance types," in *Proc. 26th Conf. Artif. Intell. (AAAI)*, Toronto, ON, Canada, 2012, pp. 356–362.

[17] U. Greveler, B. Justus, and D. Löhr, "Multimedia content identification through smart meter power usage profiles," in *Proc. Int. Conf. Inf. Knowl. Eng. (IKE)*, Las Vegas, NV, USA, 2012.

[18] A. Marchiori, D. Hakkarinen, Q. Han, and L. Earle, "Circuit-level load monitoring for household energy management," *Pervasive Comput.*, vol. 10, no. 1, pp. 40–48, Jan./Mar. 2011.

[19] J. Kolter and M. Johnson, "REDD: A public data set for energy disaggregation research," in *Proc. Workshop Data Min. Appl. Sustain. (SIGKDD)*, San Diego, CA, USA, 2011, pp. 1–6.

Günther Eibl (M'13) received the Ph.D. degree in mathematics and the M.Sc. degree in physics from the University of Innsbruck, Innsbruck, Austria, in 1997 and 2002, respectively.

He is a Research Associate with the Josef Ressel Center for User-Centric Smart Grid Privacy, Security, and Control, Salzburg University of Applied Sciences, Puch/Salzburg, Austria. He held research positions at the Institute of Biostatistics and the Theoretical Physics Institute, Innsbruck. His current research interests include extraction of information from data with a focus on statistical modeling, data mining, and privacy preserving technologies.

Dominik Engel (S'06–M'08) received the Ph.D. degree in computer science from the University of Salzburg, Salzburg, Austria, in 2008.

He is a Professor with the Salzburg University of Applied Sciences, Puch/Salzburg, Austria, where he heads the Josef Ressel Center for User-Centric Smart Grid Privacy, Security, and Control. He was a Researcher at the University of Bremen, Bremen, Germany, and the University of Salzburg, and the Product Manager at Sony Digital Audio Disc Corporation, Anif, Austria, where he was responsible for video content security. His current research interests include smart grid security and privacy, multimedia security, and technological methods for enhancing end-user trust.