# Detecting Non-Technical Energy Losses through Structural Periodic Patterns in AMI data

Viktor Botev, Magnus Almgren, Vincenzo Gulisano, Olaf Landsiedel, Marina Papatriantafilou and Joris van Rooij
*Chalmers University of Technology, Sweden* {*botev,almgren,vinmas,olafl,ptrianta,jorisv*}*@chalmers.se*
*Göteborg Energi, Gothenburg, Sweden joris.vanrooij@goteborgenergi.se*

***Abstract*—The introduction of Advanced Metering Infrastructures in electricity networks brings new means of dealing with issues influencing financial margins and system-safety problems, thanks to the information reported continuously by smart meters. Such an issue is the detection of Non-Technical Losses (NTLs) in electric power grids. We introduce a data-driven method, called *Structure&Detect*, to identify possible sources of NTLs; the method is based on spectral analysis of structural periodic patterns in consumption traces, that allows for scalable processing, using features in the frequency domain. *Structure&Detect* uses only on consumption traces, with no need for exogenous data about customers (e.g., trust or credit history) or explicit information from domain experts. As such, it complies better with privacy concerns that may be present when processing data from different sources. Using real-world consumption traces, we show that it provides high accuracy and detection rates comparable to methods that require additional, customer-specific information. Moreover, *Structure&Detect* can also be used orthogonally due to its high detection rate, as a filter, providing a narrowed-down input set to methods requiring different treatment (e.g. additional data or on-site inspection) and thus make the search for NTLs more scalable. *Structure&Detect* also enables processing each meter trace on-the-fly, as well as in a parallel and distributed fashion. These properties make *Structure&Detect* suitable for online analysis that can address common big data challenges such as the need for scalable, distributed and parallel analysis close to IoT edge devices, such as smart meters.**

***Keywords*-Non-Technical Losses, NTL, Power-Grid, Data-Driven, Discrete Fourier Transform, DFT**

## I. Introduction

Power grids suffer technical and non-technical losses of energy [24]. Technical losses stem from natural and physical reasons, mainly due to power dissipation in electrical components of the grid, for example, in distribution and transmission lines, transformers, or power-measurement equipment. In contrast, non-technical losses (NTLs) are due to events external to the power-system, such as broken measurement equipment, fraud, errors in accounting and record-keeping, installation problems that might imply safety risks [5], [20], [24]. The sum of both technical and non-technical losses is up to 50% in developing countries while developed countries show 7 to 10% losses [4], [8], [25]. Of these losses, studies attribute 25% to 50% to NTLs [25], [28]. In this paper, we introduce a data-driven approach to detect possible sources of the latter, aiming to help grid operators limit NTLs, as these may have a significant impact on the financial margin and the system safety guarantees.

Finding the source of non-technical losses is challenging. For billing purposes, the operator of a power-grid deploys electricity meters for each customer. In addition, operators commonly deploy a meter in each substation. A substation is the local distribution point of the power grid, and commonly serves in the order of 100 households for residential areas. The sum of the consumption measured separately for each household in this set, plus any technical and non-technical losses, matches the consumption as measured in the substation. Detecting the meters that incur in NTLs (i) by comparing their cumulative consumption with the consumption reported by the substation they are connected to or (ii) by manually inspecting each meter trace does not scale [5], as explained below: In the former, NTLs can be confused with technical ones, which are proportional to the number of connected meters and change over time depending on factors such as temperature, humidity, load on the grid, type of cables and age of the equipment. In the latter case, merely traces with large NTLs (thus raising suspicion because of the low consumption) can be detected. The scalability problem is exacerbated by the need for manual on-site inspection from the utility to confirm the source of a NTL (e.g., a broken meter or a fraudulent user).

While detection of NTLs in electricity networks is not a new problem (cf. [27] and references therein), we argue that the shift to Advanced Metering Infrastructures (AMI), providing operators with continuous, fine-grained (hourly or even more frequently) consumption data, enables new scalable approaches to detect NTLs. Recent approaches (see Section III) are based on statistical or applied machine learning and sometimes need hard-to-get external information, such as the level of trustworthiness for specific customers [15], [23], [24]. Through the related literature (both the aforementioned articles and other related ones, all discussed in section III), it is easy to observe that the problem is complex, the adversarial behavior is complicated to model, given that it can vary a lot, and hence the accuracy of the methods and their assumptions differ.

In this paper, we introduce a new scalable approach to detect NTLs. It is based on identifying changes in the structural periodic patterns in consumption traces, without the

978-1-4673-9005-7/16/$31.00 ©2016 IEEE

use of complementary information. Moreover, the method has additional benefits; in particular, we make the following contributions for the NTL detection problem:

Structure&Detect *method:* We introduce this method, building on signal-processing algorithms [21], [30], that can measure similarities of time series and can capture correlations between them. In other words, unlike earlier work (cf section III), *Structure&Detect* (1) analyses the fine-grained time-series data in the frequency spectrum allowing for simple processing steps, (2) checks for similarities of the frequency-spectrum-vectors, (3) enables classification and anomaly detection, while preventing transient changes from influencing the outcome, through noise filtering. *Structure&Detect* is data-driven, based on the consumption traces and their frequency patterns, and does not require a costly training phase. Moreover, it can process each meter trace on-the-fly (i.e., producing new alerts as time passes) and the whole set in parallel and/or distributed fashion, while maintaining a small set of additional features. These properties make *Structure&Detect* suitable for online analysis that can address common big data challenges such as the need for distributed and parallel analysis for detection of critical situations as in [6] and furthermore, close to IoT edge devices such as smart meters [9], [10].

*Systematic experimental study:* Our experimental study is based on a large data-set consisting of real-world anonymized consumption traces of two years each. We show that *Structure&Detect* can detect traces with NTLs while relying on consumption data only. We study groups of traces measuring consumption of different types, such as apartments, houses, and Small-Medium Enterprises (SMEs), that generate traces with different characteristic and structural patterns. We show that the *Structure&Detect* method achieves high accuracy and detection ratio, implying that the actual NTLs are to be found with certainty among the traces that raise an alert. This is particularly beneficial for the utility, and make the search for NTLs more scalable, as it can narrow down the potential more elaborate search that might be needed, to limited sets of traces with NTLs: i.e. *Structure&Detect* can be used as a first filter, relying on consumption data only, as it does not require any further information such as the trustworthiness of a customer, his/her credit history etc., or expert information, as required by other approaches (unlike e.g. [15], [23], [24]; cf also Section III); the latter, combined with on-site inspection, can instead be applied in the limited sets as complement.

The remainder of this paper is structured as follows: We introduce the system model in Section II. Section III discusses related work and highlights the differences of our approach. We overview our approach and its algorithmic concepts in Section IV. Sections V and VI detail the approach and its individual steps. Section VII evaluates our work based on real-world consumption data and simulations. We conclude and discuss future work in Section VIII.

## II. System model

In an AMI, the consumption data flows from the consumers towards the utility (i.e. electricity company). When it reaches the utility, it is recorded and stored in a database. Previously, such data was collected primarily for billing purposes. However, with the fine-grained data collected in a modern AMI, data now can be also used for purposes such as maintenance, grid operation, or to detect anomalies based on non-technical losses (the focus of this paper).

In the problem we study there are implied notions of the *real energy*, the *reported energy* and any *missing reported data*. The *real energy* is the energy physically consumed by the customer and includes potential losses. The *reported energy* is the consumed energy as reported by the smart meter. In some cases, meter traces can be incomplete due to technical problems, such as databases or transmission errors. This can affect our temporal analysis, so we model this as *missing reported data*.

### A. Non-Technical Losses (NTL) and their detection

We define Non-Technical Losses (NTL) as all losses that do not stem from the underlying transmission and distribution infrastructure. Given a set of *traces*, our goal is to distinguish the ones that incur in NTLs (*suspicious*, i.e. those with possibly *missing reported data*) from the ones that do not (*normal*, i.e those where the *reported energy* is the *real energy* consumed). As discussed in Section I the main reasons for NTL are malfunction of the measuring equipment or fraudulent activities. As examples of the latter, we have tampering of the meter (changing its behavior) or new (illegal and not measured) connections to the power lines [27]. The intention of the proposed method is to be able to detect NTL by using only the reported energy consumption from smart meters, where our focus is on NTL that would trigger a change in the regularity pattern of the consumer. Such potential cases could be because of tampering with the meter, switching the meter on and off to hide consumption, or connecting some appliances to the grid before the measurement of the meter, where each of these cases have their specific patterns.

### B. Discrete Fourier Transform (DFT) for NTL detection

The backbone of the proposed method is the study of regularity patterns using DFT. DFT transforms a signal, which in the case of NTL is the energy consumption trace reported by the meter, from the time domain to the frequency domain. The original DFT produces a frequency diagram for the studied period. In this paper, we use the equivalent *periodogram*, which instead of frequencies, uses the corresponding period of the event (period = 1/frequency), i.e. roughly speaking, it describes the original trace as a sum of periodic signals, each having a period $P$ and magnitude $M$ that represents the maximum amplitude of the corresponding periodic signal that contributes to the trace.

### C. Evaluation Metrics

In our evaluation, we use the following metrics: the detection rate *dt* (the fraction of consumption traces, or simply traces in the remainder, correctly classified as incurring in NTLs over all the ones incurring in NTLs), the hit rate *hit* (the fraction of traces correctly classified as incurring in NTLs over the overall ones classified as incurring in NTLs) and the accuracy *acc* (the fraction of all traces correctly classified as incurring or not in NTLs over all the traces). Having $M$ denote the overall number of traces and given $TP$ (True Positive - the number of correctly classified suspicious traces), $TN$ (True Negative - the number of correctly classified normal traces), $FP$ (False Positive - the number of incorrectly classified normal traces) and $FN$ (False Negative - the number of classified suspicious traces), these metrics are defined as:[1]

$$acc = \frac{TP + TN}{M},\ hit = \frac{TP}{TP + FP},\ dt = \frac{TP}{TP + FN}$$

## III. Related work

The detection of non-technical losses has received attention for more than a decade, including both mathematical approaches and field validation [1] [27]. In recent years, the capabilities of AMIs and fine-grained data recording reduce measurement errors and make new approaches viable. There are several classes of approaches to NTL detection. For this discussion, we consider them in three groups - (a) systems which require the support of human experts, (b) systems based on load profiles which require limited human support and (c) techniques without the need for human intervention, mainly based on consumption data only.

Regarding systems requiring human intervention, a traditional AI rule-based system for NTL detection is presented in the work by Leon et al. [13]. It relies on expert knowledge translated into rules. The problem of structuring knowledge into rules has the following challenges: (1) accurately describing possible property changes in rules in the domain of NTL losses and (2) maintenance of large rule databases, given that domain knowledge changes over time and the adversarial behavior may adapt to the system [14]. In contrast, our work does not rely on human interaction to maintain large rule sets. Instead, we merely rely on fine-grained consumption data and achieve an accuracy that is similar or better than the expert systems.

In systems that are based on load profiles, further data is required apart from the consumption traces. This additional data commonly includes location, voltage levels, type of day, type of customer, etc. [12], [19]. In contrast, we show in this paper an approach that does not require any further meta-data beyond a customer's consumption trace, hence complying better with privacy concerns that may be present when processing data from different sources [11], [29].

Systems that require no human interaction include methods which rely on Support Vector Machines (SVM) [16], Artificial Neural Networks (ANN) [15], [23], Self-Organizing Maps (SOM) [2] and Optimum Path Forest search (OPF) [24]. All of these involve a phase of learning and tuning of several parameters, while our method relies merely on online data and estimation of just one threshold parameter (cf. Section VI). Some of the methods [2], [15], [24] focus mainly on industrial customers, where additional data is available, such as installed power, contracted demand, voltage levels, etc. Although some of the accuracy results are similar to the ones achieved by our proposed DFT algorithm, we argue that our advantage is that we do not require additional data beyond the meter trace.

Of particular interest is the SVM-based method [16]–[18] because it reports the best results and it accommodates not only industrial but also individual customers. However, we see three key disadvantages by the approach: (1) The method uses features in the time domain such as average consumption per month, that allows only detection of abrupt changes of the consumption. In contrast, our DFT-based method can detect even small fluctuations. (2) The method uses a trust parameter for each meter that is defined on external data, such as paying bills on time and previous odd activity. We argue that such a trust factor might be subjective and expensive to get or even non-existing in some utilities. (3) Their system is further improved by an additional decision system on top by using human knowledge, and finally by adding a fuzzy inference system on top of the previous two. Both systems seem to help to reduce the false positives but the articles do not report improvements for accuracy and false negatives, as they mainly focus on the hit rate. In our work we present transparently the trade-off between the hit rate and the detection rate for our approach. Moreover, we show how these metrics are affected by tuning a single parameter in *Structure&Detect*.

The work by Mashima et al. [14] presents a novel way of evaluating a family of theft detectors, where they consider the cost of the worst-case undetected attack (based on the deployed classifier). In their evaluation, they compare their own classifier based on ARMA-GLR with EWMA and CUSUM to detect changes in the mean of the electricity consumption. Their results are clearly better than the other evaluated algorithms, but the false alarms on energy losses below 15,000 kWh are very high and might not justify the cost for investigations. On the contrary the simplicity of *Structure&Detect* allows the method to detect losses much lower than 15,000 kWh without significant changes of false positive rate.

[1] We use the same definition as found in previous work for NTL detection [17] even though it somewhat contrasts with the usage found in other sources [26].

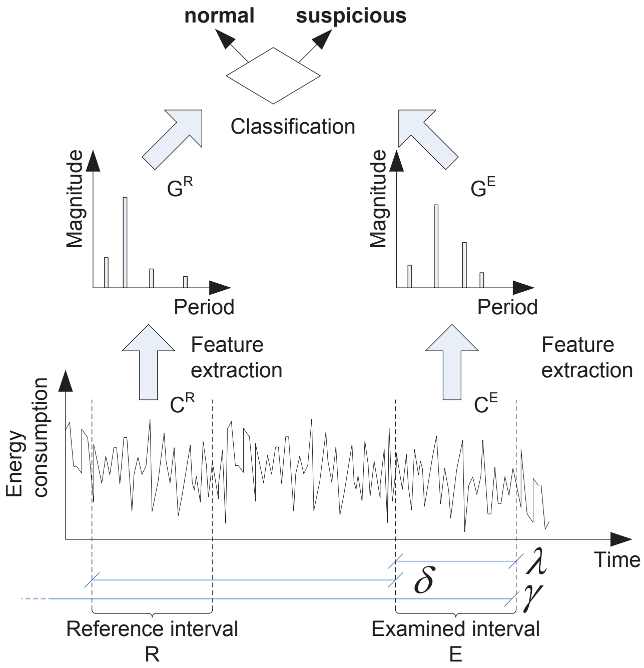


Figure 1: Overview of *Structure&Detect*'s NTLs detection method

## IV. Overview

In this section, we introduce and overview the *Structure&Detect* method. As discussed in Section II-A, given a set of traces, our goal is to distinguish the ones that incur in NTLs (*suspicious*) from the ones that do not (*normal*). For simplicity in the presentation, we focus our discussion on a single trace. Our proposed technique, nevertheless, performs the discussed analysis for all traces in parallel.

As electricity consumption is associated with human schedules (e.g., work, leisure, etc.) there is a significant amount of regularity in it. As discussed in [30] for other data domains too, this underlying regularity can be used to study the trace in the frequency rather than in the time domain (cf. Section II). When NTLs appear in a trace, such regularities change. Building and expanding the work in [30], we classify the trace as normal or suspicious over a certain interval of time by comparing its regularity over different intervals of time, as normal traces have clear regularity patterns (the latter is also validated in Section VII with a large set of real-life traces).

### A. Detection procedure

In the following we present the three steps defined by *Structure&Detect* (also shown in Figure 1), to classify a trace as normal or suspicious. Table I includes all the parameters and symbols used in the remainder.

1) **Parameter selection** *Structure&Detect* requires the user to define parameters $\lambda$, $\delta$, $\gamma$ and $\omega$. $\lambda$ represents the length of a time interval based on which a decision about the trace (normal or suspicious) is taken while $\delta$ represents the time distance between reference and examined intervals R and E. As an example, such length and distance could be 1 month and 1 year, respectively. $\omega$ represents the percentage of consumption values that should be observed during each interval (i.e., that are not missing reported data) in order to process the trace. Finally, $\gamma$ represents the interval of time during which the regularity distance is studied to define the threshold $\theta$ later used in the classification.
2) **Feature extraction** During this processing step, the reference and examined traces $C^R$ and $C^E$ are converted from the time to the frequency domain, resulting in periodograms $G^R$ and $G^E$. Subsequently, the most *significant periodicities* (i.e., those who majorly characterize the trace during intervals R and E) are extracted from $G^R$ and $G^E$. We refer to these significant periodicities as $SP^R$ and $SP^E$, and explain the feature extraction method in detail in Section V.
3) **Classification** The significant periodicities in $SP^R$ and $SP^E$ are used to compare intervals R and E and decide whether the trace is normal or suspicious during interval E based on its behavior during interval R, as explained in detail in Section VI.

### B. Continuous monitoring

Once the reference and examined interval length $\lambda$ and their distance $\delta$ are chosen, *Structure&Detect* can produce an updated classification of the trace every $\lambda$ time units (e.g., if $\lambda$ is set to one month, *Structure&Detect* can produce an updated classification once every month). It should be noted, though, that conditions such as (i) the previous classification of the trace as suspicious during R and (ii) the occurrence of a certain percentage of missing reported data in $C^R$ or $C^E$ might affect the outcome of *Structure&Detect*'s classification. The first condition does not constitute an issue from our goals' perspective, since the trace would have been already reported as suspicious to the Utility, offloading the responsibility for on-site inspection to it. To prevent the second condition from affecting *Structure&Detect*'s classification, an alert is raised and the trace is reported as suspicious if more than $\omega$% of missing reported data values are observed for an incoming interval E.

## V. Feature extraction

We explain in this section how *Structure&Detect*'s feature extraction component transforms a trace $C^i$ into its set of significant periodicities $SP^i$ and magnitudes $A^i$. We describe each processing step separately and summarize with an example.

As discussed in [30], there exist two ways to find the significant periodicities of a trace $C^i$. One is to rely on its periodogram, which gives clear indications of the periodicities showing high magnitude. The second is to rely on the circular AutoCorrelation Function (ACF), which

Table I: *Structure&Detect*'s parameters reference table.

| | |
|---|---|
| $\lambda$ | Length of the monitoring interval. |
| $\delta$ | Distance between the starting point of two intervals to be compared. |
| $\gamma$ | Interval during which expected distance between normal traces regularity is learned. |
| $\omega$ | Minimum percentage of missing reported data for which a trace over a monitoring interval is classified as suspicious (i.e classified without the analysis). |
| R | Reference interval. |
| E | Examined interval. |
| h | Period with which energy consumption is measured by the meter. |
| C$^i$ | Trace (time domain) over interval $i$, a list of $[(t_1, c^i_1), \ldots, (t_m, c^i_m)]$ where $c^i_x$ represents the consumption during time interval $[t_x, t_{x-1})$, whose length is h. |
| G$^i$ | Periodogram (frequency domain) over interval $i$, a list $[(P_1, M^i_1), \ldots, (P_n, M^i_n)]$ containing periodicities $P^i_x$ and magnitudes $M^i_x$. |
| $p^i$ | Noise level used by the noise filtering step of the feature extraction process. |
| SP$^i$ | Set of periodicities $(P^i_1, \ldots, P^i_n)$ that are significant given periodogram G$^i$. |
| A$^\alpha_i$ | Vector based on the set of magnitudes $\{M^i_1, \ldots, M^i_o\}$ (in increasing order of the corresponding periodicities) observed for the set of significant periodicities $\alpha$ during interval $i$ (or simply A$^i$ if $\alpha$ is the set of significant periodicities of interval $i$). |
| $\tilde{\text{A}}^\alpha_i$ | Vector based on the set of normalized magnitudes $(M^i_1, \ldots, M^i_o)$ (in increasing order of the corresponding periodicities) observed for the set of significant periodicities $\alpha$ during interval $i$. |
| $\theta$ | Distance threshold to classify a trace over R and E as normal or suspicious. |

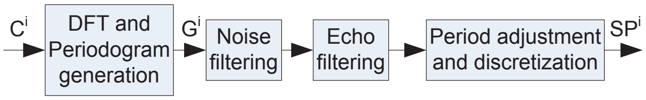

C$^i$
DFT and Periodogram generation
G$^i$
Noise filtering
Echo filtering
Period adjustment and discretization
SP$^i$

Figure 2: Main steps performed by *Structure&Detect* to extract the significant periodicities SP$^i$ of a trace C$^i$ over period $i$.

examines the similarity (and thus the potential periodicity) of a sequence to its previous values for different candidate periods. The former provides good indication of significant periodicities for small and medium periods, but loses accuracy for larger ones. The latter provides accurate estimation of all periodicities, but can possibly mask some of the significant ones (giving them a magnitude close to those of non-significant periodicites). Our feature extraction method, inspired by [30], computes SP$^i$ by:

1) distinguishing the periodicities whose magnitude can be classified as noise from the ones that contribute to G$^i$ the most,
2) distinguishing among such periodicities the dominant ones from the ones that are fractions of the former, and
3) accurately estimating large periodicities and discretizing them to multiples of h.

We describe in the following these three main steps, also shown in Figure 2.

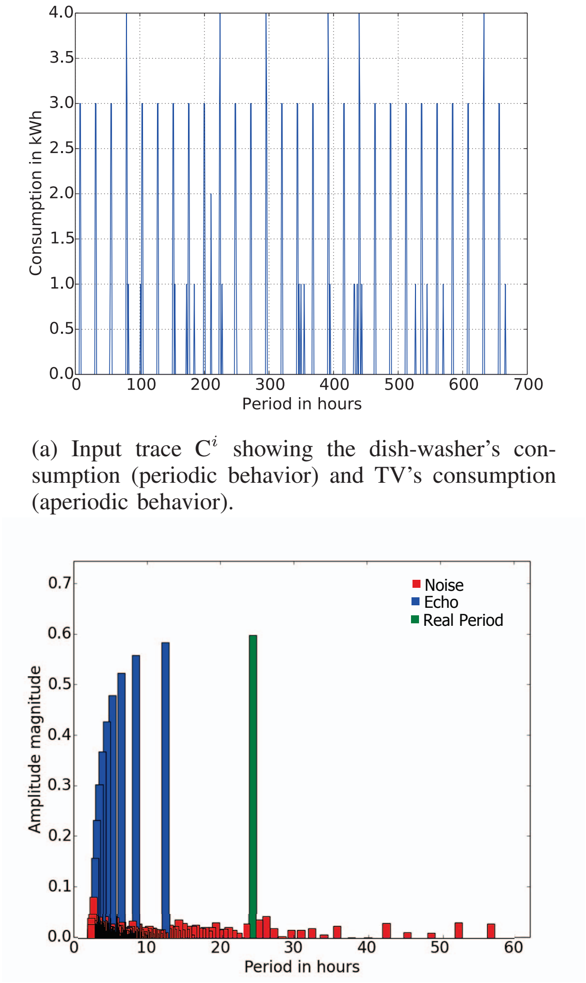
*Noise filtering:* The noise filtering step removes from the periodogram components that do not represent periodic patterns but rather echoes of periodic behavior or non-periodic, transient behavior. After this, the periodogram is thus left with the periodicities that have a magnitude higher than the *noise level* - $p_i$ and are potential candidates for SP$^i$.

As proposed in [30], $p_i$ can be chosen as the highest magnitude among the ones of the periodicities that do not take part in the periodic behavior of the trace under analysis (those periods are, roughly speaking, by-products of the processing, since the trace that is analyzed is bounded in length). One way to compute such a value is to run a permutation of C$^i$ and run DFT on it. The assumption is that since the first order statistics will be the same, the existing structural patterns will be missing in (destroyed by) the permutation, which means that the maximum magnitude found on the periodogram of the permutation will represent a possible candidate for $p_t$. In order to chose $p^i$ with high confidence, a number of permutations are analyzed and $p^i$ is set to the maximum observed an all permutations.

*Echo filtering:* After G$^i$ has been cleaned by removing noisy periodicities, the echo filtering step keeps only those who are also indicated as significant by the ACF. More concretely, we apply the ACF on the candidate Fourier coefficients corresponding to the chosen periods and check whether there is strong correlation or not. If the selected period resides on a hill (i.e., close to a strong correlation point) rather than a valley (i.e., distant from a high correlation point) it means this period definitely exists in the trace's behavior, otherwise it is filtered out.

*Period adjustment and discretization:* The periodogram gives a good approximation of the magnitude of each periodicity, but does not accurately estimate its corresponding period when the latter is large. For this reason, once a certain periodicity is classified as significant by the noise and echo filtering steps, we assume such periodicity to be the one reported by the ACF. Since h is the granularity with which energy consumption is measured in C$^i$, we also round each period to the closest multiple of h (e.g., if h is set to one hour, significant periodicities are rounded to hours too).

*Feature extraction - Example:* Figure 3 presents a sample example of how a trace C$^i$ is converted by the feature extraction method into its respective SP$^i$. Figure 3a shows the consumption of a dish-washer that runs every morning for two hours and of a TV that is randomly turned on from time to time. The TV simulates an intermittent behavior that adds noise to the regular pattern generated by the dish-washer. Figure 3b shows its resulting periodogram, with different colors for the periodicities included in SP$^i$ (green) and those discarded by the noise filtering (red) and the echo filtering (blue).



(a) Input trace C$^i$ showing the dish-washer's consumption (periodic behavior) and TV's consumption (aperiodic behavior).

(b) Corresponding periodogram with SP$^i$ periodicities (green) and periodicities discarded by noise the echo filtering (red and blue, resp.).

Figure 3: Sample execution of the feature extraction method.

## VI. Classification

We present in this section how the significant periodicities SP$^R$ and SP$^E$ are compared in order for *Structure&Detect* to classify a trace as normal or suspicious. As basis for the comparison, we use the magnitudes observed for the set of significant periodicities during the intervals $E$ and $R$ (i.e. we consider the union of those periodicities, SP$^R \cup$ SP$^E$, hence the vectors $A_R^{\text{SP}^R \cup \text{SP}^E}$, $A_E^{\text{SP}^R \cup \text{SP}^E}$). As we discuss, beside the definition of a distance metric and the tuning of an appropriate common value for the latter, the classification also requires the magnitudes of the periodicities in $E$ and $R$ to be normalized into $\tilde{A}_R^{\text{SP}^R \cup \text{SP}^E}$ and $\tilde{A}_E^{\text{SP}^R \cup \text{SP}^E}$, in order for them to be compared.

*Normalization: Structure&Detect*'s quantifies the regularity of two intervals R and E through their significant periodicities SP$^R$ and SP$^E$. Besides the periods, these components have magnitudes that depend on the amount of energy consumed and the standard deviation of the original trace. To provide an example, suppose a monitoring interval of length $\lambda = 1$ month and distance between the two intervals R and E of $\delta = 1$ year. For simplicity, suppose that the month of interest is January, and let the two winters of R and E be different the one during R - colder and the one during E - warmer. Suppose that in the two sets SP$^R$ and SP$^E$ nothing has changed in the user behavior, except the fact that the energy consumption for heating is lower during E. However, because of the differences in the the amount of energy consumed, the standard deviation of the signals will be different and similarly the magnitudes of the periodocities in the DFT will be different, hence the comparison needs to be made through scaled magnitudes.

*Structure&Detect* approaches this problem by the normalization of each magnitudeof a DFT component of a trace, through dividing by the standard of the trace. To construct $\tilde{A}_R^{\text{SP}^R \cup \text{SP}^E}$ we divide each magnitude in$A_R^{\text{SP}^R \cup \text{SP}^E}$ by $\sigma_R$ (similarly by $\sigma_E$ for the corresponding vector of $E$) . In reality this means $\tilde{\text{A}}_E^a$ contains the magnitudes as a relative measure to how close each of the periodicities is to approximating the original trace only by itself.

*Distance metric:* The distance metric used by *Structure&Detect* is chosen to allow for the monitoring of both the evolution of the significant features observed during both R and E and the monitoring of significant periodicities that are observed exclusively in R or E. For this reason, we compare the normalized magnitudes for SP$^R \cup$ SP$^E$, the union of the significant periodicities observed in R and E. Similarly to [30], we define the distance as the Euclidean distance between the corresponding vectors and we use that as similarity index between the two periodograms.

$$d(\tilde{A}_R^{\text{SP}^R \cup \text{SP}^E}, \tilde{A}_E^{\text{SP}^R \cup \text{SP}^E})$$

*Threshold distance $\theta$:* To be able to decide if a given trace over the interval $E$ is normal or suspicious, it is required to define the distance bound around the norm, that classifies everything within that bound normal and everything beyond that bound suspicious. To define this distance threshold $\theta$ we make a study over a set of meter traces that are known to be normal for an interval of at least $\lambda + \delta$. To initiate the study, the described algorithm is executed over the data in $R$. The threshold $\theta$ is defined as the biggest number that creates a false positive rate for $R$ lower than a user-defined threshold (e.g. 1%).

## VII. Evaluation

We now turn to the evaluation of *Structure&Detect*. We first introduce the experimental setup, including the the data used in the experiments and the NTLs' simulation process. We then proceed with discussing the efficiency of *Structure&Detect* for the different experiments.

### A. Data used in the evaluation

In our evaluation, we use anonymized real-world consumption traces taken from a real AMI deployment. After a pre-filtering step to find and remove traces missing large chunks of data for the full analysis period, we extract the hourly consumption traces for two years (2013 and 2014). To be able to investigate and compare the differences in consumption patterns (and thus our detection ability) among customers in apartments, residential houses, and Small and Medium Enterprises (SME), we chose 50 representative meters from each category, for a total of 150 meters.

As these traces come from a real-world deployment, the data contains a small set of potentially incorrect values (e.g., suspiciously high consumption values) and misses energy consumption readings over certain intervals of time (distinct for each household). However, in our evaluation we use the data as it is (i.e., without sanitizing it) in order for results to resemble the ones observed in real-world applications (where pre-processing or sanitization schemes might not be in place).

These captured traces represent the normal traces in our evaluation.[2] We use representative types of behavior (as described by the Utility's experts [5]) as well as a smaller set of traces from known NTLs as a basis to create a framework to simulate NTLs.

Based on the normal traces, we then run 192 NTLs generation simulations based on the following configurable properties: *number of weeks* and *number of days* per week during which NTL occur, number of NTL generating meters and number of repetitions (to capture statistical variations). Specifically, we vary the parameters as follows.

- number of weeks: 7, 25, 40 and 46.
- number of days per week: 2, 4 and 6
- number of traces affected by NTLs: 1, 2, 3 and 4 (simulating 2, 4, 6 and 8% of traces affected by NTLs for each set of 50 meters where the percentage choices were based on [28] about the region where the data was collected)
- number of permutations: 100 were conducted, in accordance to [30].
- number of repetitions: 4.

Parameters $\lambda$, $\delta$, $\gamma$ and $\omega$ (cf. Section IV) are set to 1 month, 12 months, 24 months and 50% respectively. By setting the parameters as presented, each monthly trace for 2014, possibly modified based on the NTL simulation, is classified as normal or suspicious based on its reference behavior during 2013.

### B. Results

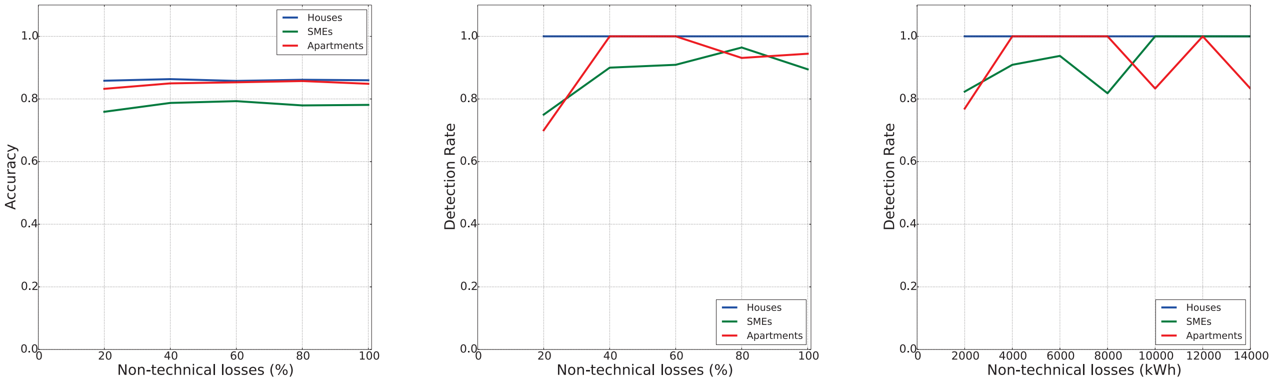
We use the metrics discussed in Section II to present the results after *Structure&Detect* has been run on the evaluation dataset. In the figures below, we summarize the following from our evaluation: overall accuracy, the detection rate, the hit rate, and finally the ROC curves.

Starting with Figure 4a, we show the accuracy of the algorithm, divided in the three categories of investigated premises. We include this graph for two purposes. First, even though somewhat limited when used to show the classification results for unbalanced classes, it is almost always included in evaluations and thus allows for comparisons across methods. Second, comparing the result across the categories we find that overall the method is quite successful but it achieves the most stable performance on houses (see also below in regards to the discussion for ROC curves). The quantity of energy not reported because of NTLs is shown in percentage of the total of the premise, meaning that for an apartment 20% NTL is much less energy than 20% for a house. Despite this, the algorithm shows good accuracy even for such small amounts, where comparative approaches in literature need higher percentage of NTLs to detect suspicious traces (see Section VII-C).

We then present the detection rate in Figure 4b. Remember that the usage of the proposed method is to prioritize investigations for the Utility based on the traces marked by *Structure&Detect* as being suspicious (sending out a repair crew to investigate the meter, etc.). For that reason, it is important to have as high detection rate as possible (i.e., all NTLs are included in this set). As seen from the figure, *Structure&Detect* is successful in all cases. Figure 4c displays the same properties as Figure 4b, but having the x-axis labeled with absolute energy. Looking then at the absolute quantity of NTLs, we see that for the largest losses our detection rate goes up but with some variability.

We also show the hit rate in Figure 5a. This metric shows how many of the traces classified as suspicious by *Structure&Detect* are true positives (i.e., traces affected by NTLs). As we assume each trace marked as suspicious will be further investigated by an expert, this metric is in principle less important than the detection rate discussed above. However, a lower rate implies a higher cost (more time spent by experts to investigate the meters) so the hit rate should be as high as possible. We achieve good results in comparison to related literature with the house category having the best result (similarly to Figure 4a and Figure 5c), where we have almost a found NTL per two

[2] Even though we cannot be 100% certain the set does not contain some traces where NTLs have occurred, the selection of meters was done together with the Utility to exclude such cases.



(a) **Accuracy** based on NTL in percentage of consumed energy
(b) **Detection rate** based on NTL in percentage of consumed energy
(c) **Detection rate** based on NTL in kWh

Figure 4: Accuracy and detection rate for apartments, houses and SMEs, respectively.

meters investigated for houses, where the behavior from 20% NTLs and onward is stable with 40% hit rate.

In our evaluation, we also studied how the performance of *Structure&Detect* changed, based on the properties of the NTLs. In one set of experiments we varied the number of days NTLs took place (5b). Even though there is some variation based on the number of days the NTLs took place, the curve is quite stable meaning that *Structure&Detect* can detect short transient NTLs as well as longer, more systematic cases, of NTLs. Again, the method demonstrates the strongest results for the house category.

Finally, we show the ROC curves for *Structure&Detect* in Figure 5c. The ROC curves show the trade-off between the true positive rate and the false positive rate, and is thus an important part of the algorithmic tuning that can be used to decide how parameters should be set (in our case $\theta$). Even though the true positive rate might appear small for a low false positive rate, the detected NTLs might still include the largest ones that would be the most important to discover. It can be seen from the graph that the ROC curve for the houses mostly dominates the other two, reinforcing the results from the other figures above that the method currently works best for the house category.

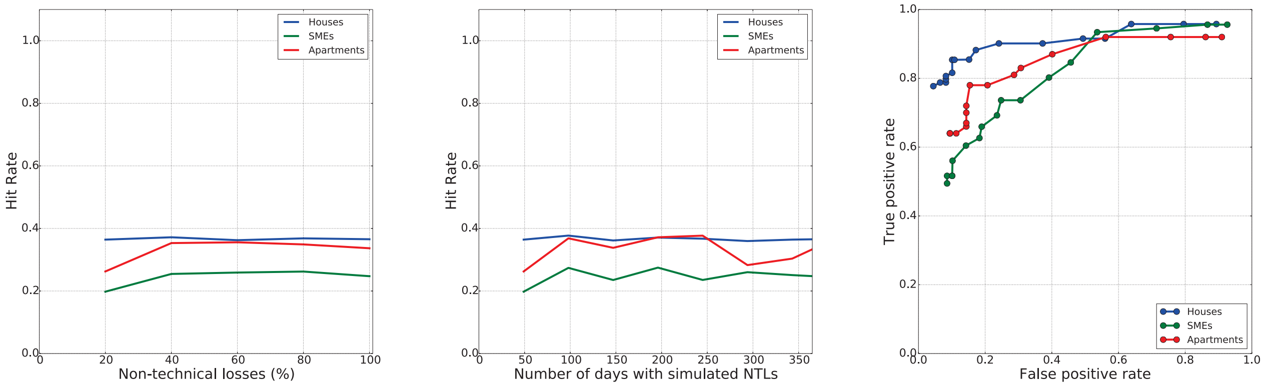
### C. Comparison to Related Work

The detection of NTLs is complex and there is no silver bullet as can be seen from the range of techniques discussed in Section III. Our design goals have thus been twofold: the method should be easy to deploy and computationally efficient, and it should complement other approaches. This influenced, for example, the choice of input data to the method, where we strive of using only the information found within the AMI (i.e. consumption traces) and not any other types of information. Even though this could influence the performance of the algorithm, it makes it easier to deploy in a real-world setting. We simply use the data available within the AMI without trying to connect sensitive information across one or more companies.[3] Secondly, as we see our proposed method as complementary to other approaches it is very important to have a high detection rate to ensure that all possible NTLs are included in the output from *Structure&Detect*, like a *prefiltered set* that can then be handed to either experts or other, more computationally expensive methods that cannot handle large-scale data.

Comparing our result with related literature, we see the expected trade-off between the detection rate and the hit rate [13]–[15]. That is, even though some methods in literature manage to achieve a higher hit rate than we do they then consequently suffer in their detection rate. This, in turn, means that such methods cannot be used a prefilters, as important cases of NTL might not be included in the output of the method.

The results reported by the algorithms in III show 65 - 95 % accuracy. In particular the SVM-based method without expert information achieves accuracy of 78% [16], with expert information 82% [17] and with an additional Fuzzy Engine - 86% [18], while the average accuracy reported by *Structure&Detect* is 83%. An important fact is that the SVM-based method detects abrupt changes, while *Structure&Detect* detects changes of just 20% with almost as good accuracy as the bigger changes (see Fig. 4a). Other methods report higher accuracy [14], [15], [20], [24], but they focus mainly on industrial customers with much higher consumption. Regarding to hit rate, the related work reports values between 20 -72% while the algorithms without human intervention report values between 26 and 50%. *Structure&Detect* is superior to many of the algorithms, with

[3] One company might own the infrastructure and handle the distribution of energy, while another is having a contract with the customer to sell energy and handle the billing.



(a) **Hit rate** based on NTL in percentage of consumed energy for apartments, houses, and SMEs respectively

(b) **Hit rate** based on number of days when the losses occur

(c) ROC curves for the method, showing the tradeoff between the true positive rate with the false positive rate

Figure 5: Hit rate and ROC curves for *Structure&Detect*

an average of 38% shown on Figure 5a, with an additional important advantage on the detection rate as per Figure 4b and Figure 4c; i.e. the latter is between 70 - 100% with good results even for low losses and low NTLs percentage. The detection rate reported by [7], [14], [15] is lower than 62%. For works where the detection rate is not reported, we expect that it is lower, because as presented in Section II-C the detection rate is inversely proportional to the false negatives, which means that algorithms with higher false negatives have lower detection rate. Since *Structure&Detect* has similar accuracy but lower hit rate, that implies higher false negatives for those algorithms and lower detection rate respectively.

In a nutshell *Structure&Detect* shows improved results also in the cases of small NTLs - lower than 40%, similar accuracy and hit rate to the top algorithms in the field without using any human intervention or additional data than the consumption trace. And finally it shows better detection rate than most of the algorithms for NTL detection which make its output useful as a *prefiltered set* for other more complicated algorithms or for manual inspections.

## VIII. Conclusion and Future Work

In this paper, we introduce *Structure&Detect*, a data-driven method to identify sources of non-technical losses (NTL) in the electrical power grid. *Structure&Detect* bases on the analysis of structural periodic patterns in consumption traces, through spectral analysis. To detect NTLs with *Structure&Detect*, we merely utilize consumption data with no need for exogenous data about customers (e.g., trust or credit history) or explicit information from domain experts.

Using real-world consumption traces, we show in our evaluation that *Structure&Detect* provides high accuracy and detection rates, comparable to methods that require additional, customer-specific information. The spectral-analysis basis implies *Structure&Detect* can be used with various granularities, enabling different monitoring possibilities. Futhermore, the traces can be processed as streams, based on incremental Fourier computation [22], allowing for continuous and distributed processing, utilizing efficient synchronization methods [3]. The method also admits parallel processing, both for parallelizing the processing of different traces and for the transformation itself. These are important for scalability, as well as deployment in contemporary infrastructures. Interesting research directions here involve also the investigation of the capabilities in resource-constraint processing enviroments.

*Acknowledgment:* The research leading to these results has been partially supported by the collaboration framework of Chalmers Energy Area of Advance project SN09: EXAMINE", by the Swedish Civil Contingencies Agency (MSB) through the project RICS", by the Swedish Foundation for Strategic Research (framework project FiC), by the Wallenberg Autonomous Systems Program (WASP) Swedish Energy Agency under the program Energy, IT and Design.

## References

[1] C. Bandim, J. Alves, J.E.R., J. Pinto, A.V., F. Souza, M. Loureiro, C. Magalhaes, and F. Galvez-Durand. Identification of energy theft and tampered meters using a central observer meter: a mathematical approach. In *Transmission and Distribution Conference and Exposition, 2003 IEEE PES*, volume 1, pages 163–168 Vol.1, 2003.

[2] J. E. Cabral, J. O. P. Pinto, and A. M. A. C. Pinto. Fraud detection system for high and low voltage electricity consumers based on data mining. In *2009 IEEE Power Energy Society General Meeting*, pages 1–5, 2009.

[3] D. Cederman, V. Gulisano, Y. Nikolakopoulos, M. Papatriantafilou, and P. Tsigas. Concurrent data structures for efficient streaming aggregation. In *Proceedings of the 26th ACM*

*symposium on Parallelism in algorithms and architectures*, pages 76–78. ACM, 2014.

[4] S. S. S. R. Depuru, L. Wang, and V. Devabhaktuni. Electricity theft: Overview, issues, prevention and a smart meter based approach to control theft. *Energy Policy*, 39(2):1007–1015, 2011.

[5] C. Fridlund, D. Hallgren, J. van Rooij, D. Svensson, and J. Svensson. Non-technical losses in electrical power systems, personal communication, 2015. Gothenburg Energy.

[6] Z. Fu, M. Almgren, O. Landsiedel, and M. Papatriantafilou. Online temporal-spatial analysis for detection of critical events in cyber-physical systems. In *Big Data (Big Data), 2014 IEEE International Conference on*, pages 129–134. IEEE, 2014.

[7] I. M. Flix Biscarri. A mining framework to detect non-technical losses in power utilities. In *ICEIS 2009 - Proceedings of the 11th International Conference on Enterprise Information Systems, Volume AIDSS, Milan, Italy, May 6-10, 2009*, pages 96–101, 2009.

[8] D. Gastaldello, A. N. Sousa, H. Amaral, Z. Vale, and F. Fernandes. Status of non-technical losses of electricity in brazil. In *ISEP GECAD Comunicaes em eventos cientficos*. ELECON, 2013.

[9] V. Gulisano, M. Almgren, and M. Papatriantafilou. Metis: a two-tier intrusion detection system for advanced metering infrastructures. In *International Conference on Security and Privacy in Communication Systems*, pages 51–68. Springer, 2014.

[10] V. Gulisano, M. Almgren, and M. Papatriantafilou. Online and scalable data validation in advanced metering infrastructures. In *IEEE PES Innovative Smart Grid Technologies, Europe*, pages 1–6. IEEE, 2014.

[11] V. Gulisano, V. Tudor, M. Almgren, and M. Papatriantafilou. Bes: Differentially private and distributed event aggregation in advanced metering infrastructures. In *Proceedings of the 2nd ACM International Workshop on Cyber-Physical System Security*, pages 59–69. ACM, 2016.

[12] S.-C. Huang, Y.-L. Lo, and C.-N. Lu. Non-technical loss detection using state estimation and analysis of variance. *IEEE Transactions on Power Systems*, 28(3):2959–2966, 2013.

[13] C. Leon, F. Biscarri, I. Monedero, J. I. Guerrero, J. Biscarri, and R. Milln. Integrated expert system applied to the analysis of non-technical losses in power utilities. *Expert Systems with Applications*, 38(8):10274–10285, 2011.

[14] D. Mashima and A. A. Cárdenas. Evaluating electricity theft detectors in smart grid networks. In *Research in Attacks, Intrusions, and Defenses - 15th International Symposium, RAID 2012, Amsterdam, The Netherlands, September 12-14, 2012. Proceedings*, pages 210–229, 2012.

[15] I. Monedero, F. Biscarri, C. Len, J. I. Guerrero, J. Biscarri, and R. Milln. Detection of frauds and other non-technical losses in a power utility using pearson coefficient, bayesian networks and decision trees. *International Journal of Electrical Power & Energy Systems*, 34(1):90–98, 2012.

[16] J. Nagi, A. Mohammad, K. Yap, S. Tiong, and S. Ahmed. Non-technical loss analysis for detection of electricity theft using support vector machines. In *Power and Energy Conference, 2008. PECon 2008. IEEE 2nd International*, pages 907–912, 2008.

[17] J. Nagi, K. Yap, S. K. Tiong, S. Ahmed, and M. Mohamad. Nontechnical loss detection for metered customers in power utility using support vector machines. *IEEE Transactions on Power Delivery*, 25(2):1162–1171, 2010.

[18] J. Nagi, K. S. Yap, S. K. Tiong, S. Ahmed, and F. Nagi. Improving SVM-based nontechnical loss detection in power utility using the fuzzy inference system. *IEEE Transactions on Power Delivery*, 26(2):1284–1285, 2011.

[19] A. Nizar, Z. Dong, M. Jalaluddin, and M. Raffles. Load profiling method in detecting non-technical loss activities in a power utility. In *Power and Energy Conference, 2006. PECon '06. IEEE International*, pages 82–87, 2006.

[20] A. Nizar, Z. Y. Dong, and P. Zhang. Detection rules for non technical losses analysis in power utilities. In *2008 IEEE Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century*, pages 1–8, 2008.

[21] S. Papadimitriou, J. Sun, and S. Y. Philip. Local correlation tracking in time series. In *Sixth International Conference on Data Mining (ICDM'06)*, pages 456–465. IEEE, 2006.

[22] A. Papoulis. *Signal analysis*, volume 191. McGraw-Hill New York, 1977.

[23] L. Pereira, L. Afonso, J. Papa, Z. Vale, C. Ramos, D. Gastaldello, and A. Souza. Multilayer perceptron neural networks training through charged system search and its application for non-technical losses detection. In *Innovative Smart Grid Technologies Latin America (ISGT LA), 2013 IEEE PES Conference On*, pages 1–6, 2013.

[24] C. Ramos, A. de Sousa, J. Papa, and A. Falcao. A new approach for nontechnical losses detection based on optimum-path forest. *IEEE Transactions on Power Systems*, 26(1):181–189, 2011.

[25] T. B. Smith. Electricity theft: a comparative analysis. *Energy Policy*, 32(18):2067–2076, 2004.

[26] M. Sokolova and G. Lapalme. A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4):427–437, 2009.

[27] D. Suriyamongkol. Non-technical losses in electrical power systems. phdthesis, Ohio University, 2002.

[28] H. Tasdoven, B. A. Fiedler, and V. Garayev. Improving electricity efficiency in turkey by addressing illegal electricity consumption: A governance approach. *Energy Policy*, 43:226–234, 2012.

[29] V. Tudor, M. Almgren, and M. Papatriantafilou. Analysis of the impact of data granularity on privacy for the smart grid. In *Proceedings of the 12th ACM workshop on Workshop on privacy in the electronic society*, pages 61–70. ACM, 2013.

[30] M. Vlachos, P. S. Yu, and V. Castelli. On periodicity detection and structural periodic similarity. In *Proceedings of the 2005 SIAM International Conference on Data Mining, SDM 2005, Newport Beach, CA, USA, April 21-23, 2005*, pages 449–460, 2005.