This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2015.2501422, IEEE Sensors Journal

Low Cost Disaggregation of Smart Meter Sensor Data

George Koutitas, *Member, IEEE* and Leandros Tassiulas, *Fellow, IEEE*

***Abstract*—This paper proposes a novel load disaggregation algorithm, based on non-intrusive appliance monitoring that provides accurate results without the need of the reactive power component. The proposed technique requires the minimum possible number of sensor meters within the smart meter hardware and reduces data traffic of the Advanced Metering Infrastructure (AMI) network since it processes only low frequency smart meter data of the active power component of a residential unit. We propose a three stage process for energy usage analysis, based on pulse extraction, pulse clustering and classification and pulse to appliance association. The smart meter data are decomposed into a discrete set of pulses and each pulse is associated to the operation of the appliances. The outcome of the algorithm is a pulse to appliance association that creates a daily load disaggregation map for the residential unit. In addition, the paper presents an overview and comparison of existing solutions that will help the reader obtain a good overview of the landscape of load disaggregation from smart meter sensor readings.**

***Index Terms*—smart grids, smart meter data, signal processing, load disaggregation, non intrusive appliance load monitoring, energy data, big data**

I. Introduction

OVER the recent years, the evolution of Smart Grid has created all the necessary foundations for data and command flow in the power grid network. It can now be regarded as an application development platform. New services and applications have emerged to increase customer service offerings but also to enrich demand response programs. Since management is based upon knowledge, one of the most promising directions of investigation is the appliance usage disaggregation. Load disaggregation (also known as appliance usage disaggregation or non-intrusive appliance load monitoring NIALM) is considered as an important service since it provides information about the appliance usage in the household that can help utilities' demand response programs.

Load disaggregation is a methodology for recognizing individual appliance signal signatures from aggregated circuit readings. The classification of existing algorithms is based on two general categories. *Intrusive* monitoring performs appliance disaggregation based on individual meter readings where in the *non-intrusive* approach, sophisticated signal processing techniques are required to disaggregate individual appliance signals from the total household meter reading [1, 2]. *Non-intrusive* approach is subdivided in *supervised* and the *unsupervised* training algorithms [3, 4].

A. Current Business Models

Utilities or smart grid application providers usually offer free smart meter hardware equipment to their customers and they allow access to energy services under a fixed subscription fee. This business model requires low hardware costs and computational complexity at user level. Thus, smart meters operate as data transmitters or command receivers in the Advanced Metering Infrastructure (AMI) network and all processing is migrated in the cloud. In addition, data traffic needs to be limited to avoid interference with existing wireless networks. These constraints bound the applicability of *High Frequency* (HF) data sampling (>kHz) of active and reactive power readings. The most preferred methodology is to implement non-intrusive techniques of *Low Frequency* (LF) data (0.1-5Hz) of the active power component, to reduce the cost of offered services.

B. Novelty of Proposed Solution

This paper presents a non-intrusive load disaggregation algorithm based on Fuzzy logic and pattern recognition that considers LF data sampling of active power readings. The achieved accuracy is similar to the case of considering the reactive power since we extract knowledge regarding the presence of the reactive component from the variance of the active power component. The proposed solution considers a *n-dimensional* space for decision making taking into account appliance characteristics, pulse characteristics, user activity and external condition. In the proposed algorithm the user is required to provide information about the number, type, labeled power value of the appliances in the residential unit. The proposed solution is tested over a real smart meter environment, implementing low cost smart meter equipment provided by Kimatica Ltd. The observed accuracy varies between 70%-99% according to the simultaneous appliance usage in the residential unit. The novelty of the proposed algorithm can by summarized as:

- no need of reactive power since this information is

This work was funded from the European Social Fund and Greek national funds ARISTEIA, SOFON, MIS 37262 through the Operational Program "Education and Lifelong Learning" of the NSRF.

G. Koutitas is with the Dept. Computer Engineering and Telecommunications, University of Thessaly, Volos, Greece (e-mail: george.koutitas@gmail.com).
L. Tassiulas is with Yale University, Network Science Department, New Haven, USA (e-mail: leandros.tassiulas@yale.edu)

1530-437X (c) 2015 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2015.2501422, IEEE Sensors Journal

extracted from the variance of the active power

- outcome is a daily pulse to appliance association per user that is scalable in terms of storage
- considers human behavior and external data information to minimize error propagation
- is ideal for real applications of current business models since it minimizes costs of hardware equipment

II. State of the Art and Differentiation

A feature extraction and pattern recognition approach for load disaggregation is given in [5]. The authors focused on event extraction from appliance usage. Event was detected from the switch on/off state of the appliance. In [6] the authors implemented a decision making algorithm, called Committee Decision Mechanism (CDM) to detect appliances over a typical household. An additive Factorial Hidden Markov Model (FHMM) for appliance detection is given in [7]. The proposed solution incorporates a difference FHMM to consider past recording in the detection process. In [8] a taxonomy of load signatures of typical appliances is presented accompanied with environmental data used to add new layers on appliance detection and increase accuracy. In [9] the authors propose an application for load disaggregation by considering reactive power component. A HF approximation is given in [10] where current harmonics are used to classify appliances in the household recordings. This approximation can increase accuracy of predictions but requires expensive hardware equipment but also increases communication costs. A classification algorithm for appliance detection based on non-intrusive approximation in presented in [11, 12]. The authors implement clustering and classification algorithms over LF signals. A different approach based on neural network (NN) and supervised training is given in [3]. Neural network is considered an accurate process but it requires high processing power in the cloud. Finally, a survey of load disaggregation techniques and applications is given in [13].

In general, the aforementioned techniques can provide an adequate framework for practical implementations of load disaggregation but their main drawbacks are:

- HF solutions can provide very accurate predictions but they increase system's costs
- LF solutions require additional reactive power readings that increase hardware costs
- HMM is a preferred solution for real time applications but creates error propagation
- Neural Network approximation can be considered accurate but they require high processing power

III. Energy Data Characteristics

A. Capturing Smart Meter Power Signals

A wireless smart meter is connected to the electrical panel of the residential unit (Fig. 1). The smart meter captures and transmits power samples $y(t)$ (Watt) using a clamp sensor and a radio unit. The sampling frequency was $\lambda=7$ seconds (1 sample every 7 secs ~ 0.14Hz). Data is transmitted over an RF

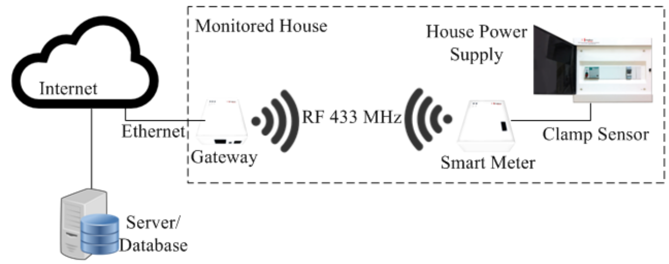

Fig. 1. Monitoring of smart meter energy data

433MHz dedicated wireless channel to the gateway. The gateway is attached to an Ethernet port of user's router and is responsible to push a 15 minute data set to the server/database over the internet. From the database, the third party application provider can use the raw data to perform load disaggregation. Raw data that represents the power consumption of the residential unit is a matrix $\mathbf{Y}=\left(y_i(t): t=1,...,T\right)$, spanning over $T=24$ hours and $i \in N$, where $N = 3600 \cdot T/\lambda$ is the set of samples per day. The energy consumption over time T is:

$$E = \lambda/3600 \cdot \sum_{t=1}^{T} y_i(t) \tag{1}$$

B. Appliances Footprints

A typical daily power consumption plot of vector $\mathbf{Y}$ is presented in Fig. 2. The plot is the superposition of power values of appliances that operate during the day. Assuming that there is a set of M appliances with identifiers $j \in M$, and each appliance consumes power equal to $x_j(t)$, then:

$$y_i(t) = \sum_{j=1}^{m<|M|} x_j(t) \tag{2}$$

The power consumption footprint of the appliance $\mathbf{X}$ is usually unique and holds important information for the algorithm.

Based on Fig. 2 of the paper and Fig. 3 of [15], we distinguish the following main characteristics of the power curve:

Periodicity: thermostatic appliances that create periodicity in the power curve. During their operation, they switch between *on/off* states (t_{ON}, t_{OFF}). The most common example of such appliances are the refrigerators and freezers.

Periodic with preparation: thermostatic appliances, like ovens and air condition units that when they are switched *on*, present a preparation time period (T_{ON}) with constant power consumption $x_j(T_{ON})$ that is followed by on/off states with $T_{ON}>t_{ON}$ and $T_{ON}>t_{OFF}$, where $x_j(T_{ON}) \approx x_j(t_{ON})$.

Static: appliances that create constant power consumption $x_j(t)=const.$ while they are *on*, such as lights and PCs.

Multi State: appliances that are active and present a preparation period with constant power consumption $x_j(T_{ON})$ that is followed by different power states of smaller amplitudes $x_j(T_{ON}) > x_j(t_{HOP})$. These are usually appliances with different cycles of operation like washing machines.

Spike: once an appliance is *on*, there might be a spike associated to its operation due to motors of inductive appliances.

1530-437X (c) 2015 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2015.2501422, IEEE Sensors Journal

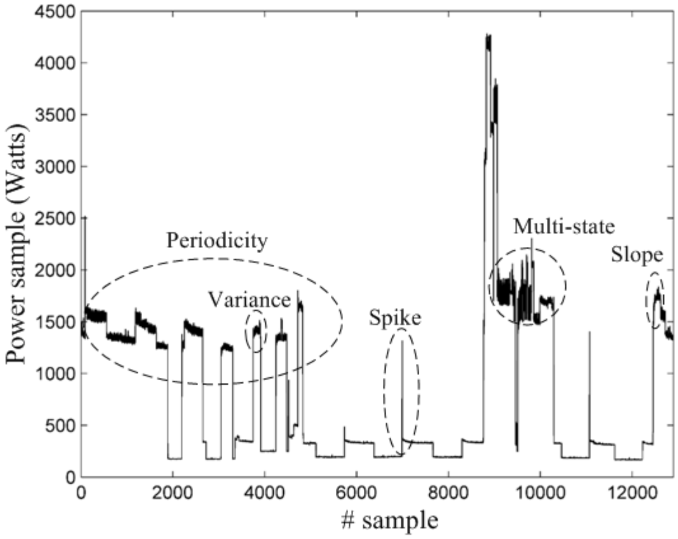

Fig. 2. Raw energy data and information extraction

Variance: a fluctuation of power around the average power consumption. This characteristic is met at inductive appliances and its presence is correlated to a positive reactive power component.
Slope: once an inductive appliance is *on*, a slope of the power curve is observed for a short time interval. This slope is a unique characteristic of a motor.

An interesting survey of load profiles of typical appliances is also given in [14].

C. Decomposition of Appliance Footprint to Pulses

Based on these observations, we decompose the appliance footprint into a set of discrete rectangular pulses Π_n with $n \in \Omega_M$, as shown in Fig. 3:

$$\Pi_n(t) = \begin{cases} P_n, & t_n \le t < t_n + \tau \\ 0, & elsewhere \end{cases} \tag{3}$$

where t_n is the initiation of the pulse with $1 < t_n < T_M$, τ is the duration of the pulse and P_n is the amplitude of the pulse.

D. Integrating Human Behavior Information to Pulses

The performance of the load disaggregation algorithm can be enhanced by considering human behavior in the household. This can be modeled as an additional layer of information to the load disaggregation algorithm based on expected usage characteristics. For example, it is almost impossible to operate an oven at 03.00am at night but is most probable to operate the fridge or the air condition unit. Furthermore, the duration of specific appliances is bounded by a maximum threshold. For example an oven cannot operate for more than 3 hours.

Based on this information, we add five more layers of information in our solution, the *time of use* probability, *look for neighbor pulses* probability, *duration of pulse*, *sequence of operation* and *external conditions* as discussed in *Section V*.

IV. Algorithm Description

A. User Input

This process refers to appliance registration. The user is required to register each appliance in the residential unit by

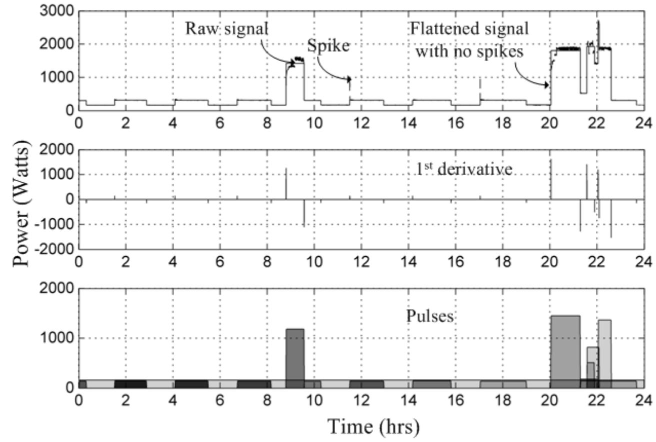

Fig. 3. Processing of power signal. From raw smart meter data to pulse extraction.

indicating a tabulated power value. In case a user wants to add or remove a new appliance in the system, then the registration needs to be updated. We then create a known power table of each registered appliance named as $\mathbf{P}^{apriori} = \left(P_j^{apriori} : j \in M\right)$.

B. Data Incorporation

There are two types of information that is automatically attached during the registration. The first corresponds to values that include personalized user activity behavior, such as *Time of Use Probability* and *Duration of Pulse*. *Time of Use Probability* is defined as a binary vector that indicates the expected time of operation of the appliances, with $T_j^U(t) = \{0, \ 1\}$ indicating if an appliance is most probable to operate a specific time, *t* of the day or not. *Duration of Pulse* is a maximum threshold of the expected duration of a power pulse. For example, a power pulse created by an oven cannot exceed 1 hour of duration.
The second category corresponds to characteristics that exist in typical appliances and are the *Look for Neighbors Pulses*, *Multi State*, *Periodicity*, *Spike*, *Slope*, *Variance* as described in Table I and discussed later in the paper (*Section V*).

C. Spike Processing

Spikes can be easily detected since they create a high power value on the smart meter signal with a very short time duration. In all cases, spikes have duration smaller than *5 seconds* that correspond to one discrete measurement. We detect the spikes of the power signal and store them as information that will be used at the load disaggregation algorithm. We create a binary vector matrix to express this information during the day *T*:

$$s_i = \left(s_i(t) : t = 1,....T\right) \tag{4a}$$

By denoting as t_s the point in time when a spike is detected, then:

$$\begin{gathered} s_i(t_s) = 1 : \forall t_s \\ \left|y_i(t_s) - y_i(t_s+1)\right| > \delta \quad and \quad \left|y_i(t_s) - y_i(t_s-1)\right| > \delta \end{gathered} \tag{4b}$$

where δ is a threshold and can be safely set to δ*=1000Watts*.

Spikes need to be removed since they distort the pulse extraction process. Removal of spike is simple and it can be based on the following smoothing process:

$$y_i(t_s) = y_i(t_s+1), \quad \forall t_s \in T \tag{5}$$

The effect of spike removal is given in Fig. 3 (upper subplot).

1530-437X (c) 2015 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2015.2501422, IEEE Sensors Journal

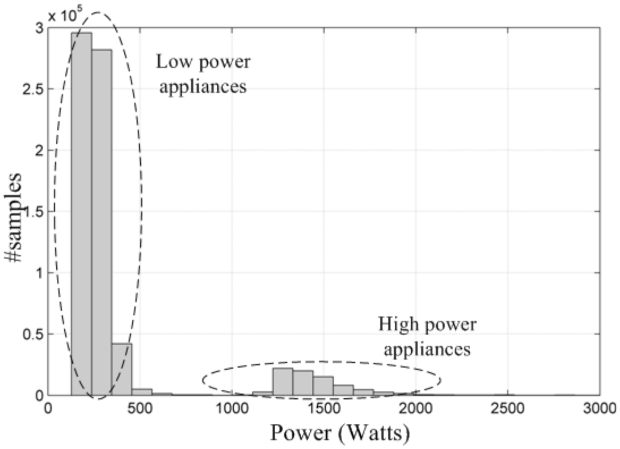

Fig. 4. Histogram of power samples over a two month period for a typical residential unit.

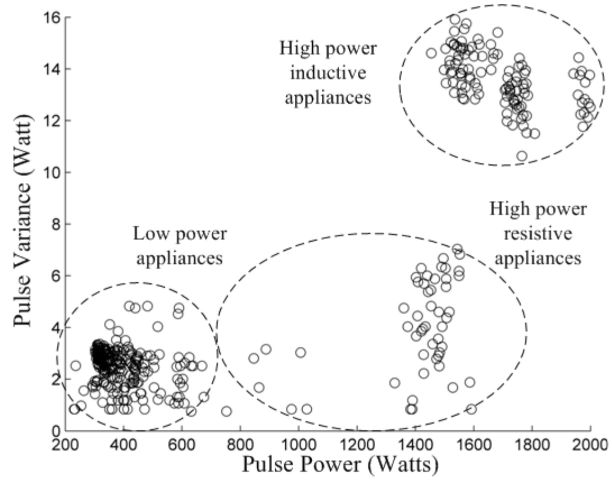

Fig. 5. Scatter plot of power and variance of measured power data

D. Flattening

The simplest and most efficient way to flatten the power signal is to perform a moving average filter. It is observed that most households have two types of appliances. The *low power* appliances, such as fridge, entertainment equipment, lights, etc., and the *high power* appliances such as ovens, heat water, air condition units, etc. In most cases, the low power appliances have operational power in the range of 50-300Watts whereas high power appliances operate in the range above 800Watts. This observation is presented in Fig. 4.The flattening algorithm decomposes the power signal in the two components, the high and the low power component, and performs a smoothing filter respectively to minimize signal distortion. The outcome of this process is presented in Fig. 3 (upper subplot). The process creates and stores a new time series of the power signal, which is the flattened power signal $\mathbf{Y}^F = \left(y_i^F(t): t = 1,...,T\right)$.

E. Pulse Extraction

Pulse extraction replaces the time series of power signal $\mathbf{Y} = \left(y_i(t): t = 1,...,T\right)$ with a set of discrete pulses. Pulses are computed according to the first derivative of the flattened power signal. For simplicity we refer to $\mathbf{Y}$ as $\mathbf{Y^F}$ of the previous section. The first derivative is computed according to:

$$d\mathbf{Y} = \left(\frac{dy_i(t)}{dt}: t = 1,...,T\right) \tag{6}$$

From this matrix we create two matrices with the positive and negative values of d$\mathbf{Y}$:

$$dy_k^+ = \left(\frac{dy_i(t)}{dt} > 0 : k = 1,..., K\right), K \subset N \tag{7}$$

$$dy_q^- = \left(\frac{dy_i(t)}{dt} < 0 : q = 1,..., Q\right), Q \subset N \tag{8}$$

with $K = Q, K + Q = N$ and N is the set of samples of the power signal. We name as an *event* the detection of the switch *on* and *off* of an appliance. The *event* is regarded as a power *pulse* of an appliance as shown in Fig. 3. We start the detection of the event from the first negative value of dy_q^-, $q=1$, that indicates an appliance has stopped its operation (is switched *off*) and we need to associate this negative value with the switch *on* of the appliance dy_k^+. We use a weighted *Euclidean* distance function to associate q to k index, taking into account the power and time distance of dy_k^+ and dy_q^-. The power component indicates that when an appliance is switched *on* and *off* creates almost identical power reactions in the smart meter signal, and the time component indicates that in most cases these positive and negative reactions of the power signal are in close vicinity. Since the power component is most important we use heuristic weight factors $\omega=0.75$ and $\psi=0.25$ for the power and time components respectively. The association function for the pulse is given by:

$$f : q \mapsto k = \min_k \sqrt{\omega \cdot \left(\left|dy_q^-\right| - \left|dy_k^+\right|\right)^2 + \psi \cdot \left(\left|t_q^-\right| - \left|t_k^+\right|\right)^2} \tag{9}$$

where $t_k^+ < t_q^-$ and values should be normalized.

F. Carried Information by Pulses

Each pulse carries the following information:

Pulse amplitude: it is given by

$$P_n = \frac{\left|dy_q^-\right| - \left|dy_k^+\right|}{2} \tag{10}$$

Pulse initiation: it is given by

$$t_n = t_k^+ \tag{11}$$

Pulse duration: it is given by

$$\tau_n = t_q^- - t_k^+ \tag{12}$$

Pulse spike: a binary value, indicating if the pulse n has a spike:

$$s_n = s_i(t_n) \tag{13}$$

Pulse slope: assuming $y_i^n = \left(y_i(t): t = t_n,....t_{n+\delta}\right)$ where $\delta=14$ times steps (~1 minute of smart meter recordings following the initiation of the pulse) then the *pulse slope* is the slope of the linear regression found as:

1530-437X (c) 2015 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2015.2501422, IEEE Sensors Journal

TABLE I
APPLIANCE CHARACTERISTICS MAP

Appliance Type	Name of appliance	Variance	Spike	Slope	Periodicity	Multi State	Neighbor pulses	Sequence of Operation
Resistive	Electric Storage Heater	X	X	X	√	X	X	X
	Oven	X	X	X	√	X	X	√
	Hotplate	X	X	X	√	X	X	√
	Heat Water	X	X	X	√	X	X	√
	Toaster	X	X	X	√	X	X	X
	Hair Dryer	X	X	X	X	X	√	√
	Coffee	X	X	X	√	X	X	X
	Lights	X	X	X	X	X	X	X
Inductive	Vacuum Cleaner	X	X	X	X	X	√	X
	Fridge/Freezer	√	√	X	√	X	X	X
	Air Condition	√	√	√	√	X	X	X
	Washing / Dryer	√	X	X	X	√	X	√
	Dish Washer	√	X	X	X	√	X	√
Capacitive	TV	X	X	X	X	X	X	√
	Stereo	X	X	X	X	X	X	X
	Game/Entertainment	X	X	X	X	X	X	X
	PC	X	X	X	X	X	X	X

$$\beta = \frac{d \cdot \left(\sum_{i=1}^{i+d} y_i^n \cdot \left(t_i^n - t_n \right) \right) - \sum_{i=1}^{i+d} y_i^n \cdot \sum_{i=1}^{i+d} t_i^n}{d \cdot \sum_{i=1}^{i+d} \left(t_i^n \right)^2 - \left(\sum_{i=1}^{i+d} t_i^n \right)^2} \tag{14}$$

We use a binary indicator and a threshold to detect the presence of slope in the initiation of a pulse and this is given by:

$$\beta_n = \begin{cases} 1, & \beta \geq \delta \\ 0, & \beta < \delta \end{cases} \tag{15}$$

where δ is a threshold and was set $\delta=5$. This threshold needs to be calibrated for each household.

Pulse variance: we consider the variance of the pulse to model the presence of inductive appliances. This is indicated in Fig. 5. It is observed that high power inductive appliances create a variance that is significantly larger compared to those of resistive appliances. This observation helps us eliminate the need of the reactive power component of the signal that would require additional hardware installations. The standard deviation of the raw smart meter data for the time interval equal to the duration of the pulse models the variance as:

$$r_n = \frac{1}{k} \left(\sum_{t=t_n}^{t_n+\tau_n} \left(y_i(t) - \frac{1}{k} \sum_{t=t_n}^{t_n+\tau_n} \left(y_i(t) \right) \right) \right)^2 \tag{16}$$

where k are the number of samples in the time interval of the pulse given as $k = \tau_n / \lambda$, with λ the sampling rate.

We denote by r_0 the variance observed when no appliances are operating, and this value is used for comparison. We use a binary indicator and a threshold to detect the presence of variance in the pulse and this is given by:

$$v_n = \begin{cases} 1, & r_n / r_0 \geq \delta \\ 0, & r_n / r_0 < \delta \end{cases} \tag{17}$$

where δ is a threshold and was set $\delta=3$.

V. LOAD DISAGGREGATION

With the aforementioned process, the raw smart meter recordings are decomposed to a discrete set of pulses and each pulse carries unique information that models an appliance footprint. The outcome of the proposed algorithm is a pulse to appliance association. Each appliance has a discrete set of pulses that models its daily operation.

A. The Scoring System

A scoring system was used to model the pulse to appliance association (18). The appliance with the highest score will be associated to the relevant pulse. Coefficient C_j, includes normalized metrics in the range $0 \leq M \leq 1$, that model the correlation of the pulse characteristics to the appliances.

$$\begin{aligned} C_j &= w_1 \cdot \left(M_P + M_T + M_D + M_{PR} \right) + \ldots \\ &\ldots + w_2 \cdot \left(M_M + M_V + M_{SL} + M_S + M_N + M_A \right) - \ldots \\ &\ldots - \left(\Phi_{temp} - \Phi_{energy} \right) \end{aligned} \tag{18}$$

The weighting factor w_1 was set to 0.7 and w_2 to 0.3 to give priority to metrics that hold *a priori* information of the appliances. In addition, two penalties were used to correct inappropriate pulse to appliance association.

B. Clustering and Classification

A *K-Nearest Neighbors* (KNN) clustering algorithm was used as a first step to create a cluster of the most probable candidate appliances for each pulse. Once the most probable appliances are detected, then the algorithm implements the metrics and scoring system described in this section to provide an accurate pulse to appliance association. Clustering is based on the detected pulse power P_n and the registered power values, $\mathbf{P}^{apriori}$. Thus, a pulse $n \in K$ is first associated to a cluster of appliances $j \in L$ which are regarded as first candidate appliances for the load disaggregation. The KNN algorithm used $K=5$, to indicate the closest 5 appliances taking into account the *Euclidean* power distance of P_n and $P_j^{apriori}$.

Power metric: M_j^P correlates the power amplitudes. We used a linear relation to normalize the metric as shown below:

1530-437X (c) 2015 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2015.2501422, IEEE Sensors Journal

$$M_j^P = \begin{cases} -\dfrac{\omega}{P_n}\cdot\left(P_n - P_j^{apriori}\right)+1, & P_n > P_j^{apriori} \\ \dfrac{\omega}{P_n}\cdot\left(P_n - P_j^{apriori}\right)+1, & P_n \le P_j^{apriori} \end{cases} \quad (19)$$

The value $\omega=3$, indicates that the metric becomes zero if the pulse power is 1/3 smaller than the registered appliance power.

Time metric: M_j^T is the metric that correlates the operation of specific appliances with the expected *Time of Use*. Similar to the power metric, it was computed based on the *a priori* information of appliance usage taking into account typical human behavior. It is computed according to:

$$M_j^T = \begin{cases} -\dfrac{\lambda}{3600}\cdot\left(t_n - t_{j,2}^{apriori}\right)+1, & t_n > t_{j,2}^{apriori} \\ \dfrac{\lambda}{3600}\cdot\left(t_n - t_{j,1}^{apriori}\right)+1, & t_n \le t_{j,1}^{apriori} \end{cases} \quad (20)$$

where $t_j^{apriori}$ represents a discrete time set of the expected time of use. The value 3600 indicates that the metric becomes zero in case the pulse initiates with 1 hour difference from the expected time of use.

Duration metric: M_j^D correlates the duration of the detected pulse with the expected pulse duration of the appliances.

$$M_j^D = \begin{cases} -\dfrac{\psi}{\mathrm{T}_j^{apriori}}\cdot\left(\tau_n - \mathrm{T}_j^{apriori}\right)+1, & \tau_n > T_j^{apriori} \\ \dfrac{\psi}{\mathrm{T}_j^{apriori}}\cdot\left(\tau_n - T_j^{apriori}\right)+1, & \tau_n \le T_j^{apriori} \end{cases} \quad (21)$$

where $T_j^{apriori}$ represents the expected duration of the pulses of the appliance. The value $\psi=10$, indicates that the metric becomes zero in case the duration of the pulse is smaller than the 1/10 of the expected value.

Periodicity metric: M_j^{PR} detects the periodicity of the detected pulse. For each pulse, we search if there are similar pulses that can be considered as parts of the same appliance. The search takes place for a small time horizon T_h in the forward and backward time around the pulse. An example is given in Fig. 6. For each appliance, we use a binary metric to model if there are periodic pulses detected or not $pr_n = \{0, \; 1\}$. The condition to set $pr_n = 1$ is to find at least 2 identical pulses within the time horizon T_h that satisfy:

For forward time:

$$\begin{aligned} p_n - \delta_p < p_{n+i} < p_n + \delta_p \\ \tau_{n+i} << \tau_n \end{aligned} \quad (22a)$$

where i indicates all the pulses that initiates in the time interval $t_n < t_i < t_n + T_h$

For backward time:

$$\begin{aligned} p_n - \delta_p < p_{n-i} < p_n + \delta_p \\ \tau_{n-i} >> \tau_n \end{aligned} \quad (22b)$$

where i indicates all the pulses that initiates in the time interval $t_n - T_h < t_i < t_n$

For both time directions:

$$\begin{aligned} p_n - \delta_p < p_{n\pm i} < p_n + \delta_p \\ \tau_n - \delta_\tau < \tau_{n\pm i} < \tau_n + \delta_\tau \end{aligned} \quad (22c)$$

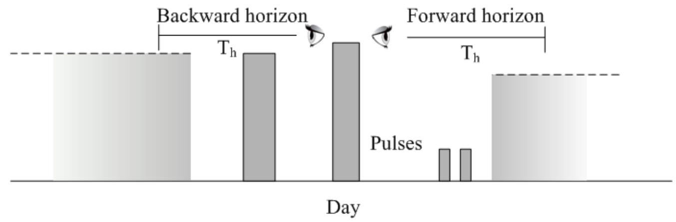

Fig. 6. Extracted pulses during a day observation

where i indicates all the pulses that initiates in the time interval $t_n - T_h < t_i < t_n + T_h$. $\delta_p = 0.2 \cdot p_n$ and $\delta_\tau = 0.2 \cdot \tau_n$ are thresholds to indicate similarities within 20% interval Equations (22a) and (22b) detect the *periodic with preparation* characteristic of the appliances where (22c) detects the general *periodic* operation of the appliance as discussed in *Section III,B*. Thus, the periodicity metric can be computed as:

$$M_j^{PR} = \begin{cases} 1, & if \quad pr_n = PR_j^{apriori} \\ 0, & if \quad pr_n \ne PR_j^{apriori} \end{cases} \quad (23)$$

where $PR_j^{apriori}$ is a binary value to represent if appliance j is expected to create periodic pulses or not. This information is indicated in *Table I*.

Multi-state metric: M_j^M models the multi-state operation of the appliance. A time horizon T_h=*1hour* was used to capture the operating cycles of an appliance. The condition to detect the multi-state operation was set:

$$\begin{aligned} \delta_p < p_{n\pm i} < p_n \\ \delta_\tau < \tau_{n\pm i} < \tau_n \end{aligned} \quad (24)$$

where i indicates all the pulses that initiate in the time interval $t_n < t_i < t_n + T_h$ and $\delta_p = 300W$ and $\delta_\tau = 10\,\text{min}$ are the thresholds used.

We use a binary metric $m_n = \{0, \; 1\}$ to model if the pulse is part of a multi-state operation or not.

$$M_j^M = \begin{cases} 1, & if \quad m_n = M_j^{apriori} \\ 0, & if \quad m_n \ne M_j^{apriori} \end{cases} \quad (25)$$

where $M_j^{apriori}$ is a binary value to represent if appliance j is expected to create multi state pulses or not. This information is indicated in *Table I*.

Variance metric: M_j^V is the metric that models the existence of variance during the operation of an appliance. This information was integrated within the pulse as given in (17). To compute the metric we use the following relationship:

$$M_j^M = \begin{cases} 1, & if \quad v_n = V_j^{apriori} \\ 0, & if \quad v_n \ne V_j^{apriori} \end{cases} \quad (26)$$

where $V_j^{apriori}$ is a binary value to represent if appliance j is an inductive load or not. This information is indicated in *Table I*.

Slope metric: M_j^{SL} is the metric that models the existence of a slope at the initiation of an appliance. This information was integrated within the pulse as given in (15). To compute the metric of the slope we use the following relationship:

$$M_j^{SL} = \begin{cases} 1, & if \quad \beta_n = B_j^{apriori} \\ 0, & if \quad \beta_n \ne B_j^{apriori} \end{cases} \quad (27)$$

1530-437X (c) 2015 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2015.2501422, IEEE Sensors Journal

where $B_j^{apriori}$ is a binary value to determine if appliance *j* is an inductive load or not. This information is indicated in *Table I*.

Spike metric: M_j^S is the metric that models the existence of a spike at the initiation of an appliance. This information was integrated within the detected pulse as given in (13). To compute the metric we use the following relationship:

$$M_j^S = \begin{cases} 1, & if \quad s_n = SP_j^{apriori} \\ 0, & if \quad s_n \neq SP_j^{apriori} \end{cases} \tag{28}$$

where $SP_j^{apriori}$ is a binary value to represent if appliance *j* is expected to present a spike or not (*Table I*).

Neighbors metric: M_j^N is the metric that examines if there are similar pulses in the close vicinity of the detected pulse *n* that can model the expected human activities named as *look for neighbor pulses* (introduced in *Section III, D*). This metric captures appliance usage with non periodic characteristics, i.e. a vacuum cleaner that is switched on/off during its operation. For each appliance we use a binary metric to model if there are neighbor pulses detected or not $\ln_n = \{0, \ 1\}$. The condition to set $\ln_n = 1$ is to find at least 1 identical pulse within the time horizon $T_{h=}10mins$ that satisfy:

$$p_n - \delta_p < p_{n+i} < p_n + \delta_p \tag{29}$$

where *i* indicates all the pulses that have initiation in the time interval $t_n - T_h < t_i < t_n + T_h$.

$$M_j^N = \begin{cases} 1, & if \quad \ln_n = LN_j^{apriori} \\ 0, & if \quad \ln_n \neq LN_j^{apriori} \end{cases} \tag{30}$$

where $LN_j^{apriori}$ is a binary value to represent if appliance *j* is expected to create neighbor pulses or not (*Table I*).

Appliance sequence metric: M_j^A detects an expected sequence of operation of specific appliances in the close vicinity of the pulse *n*, $t_n + T_{h=}180mins$. The metric models the characteristic named as *sequence of operation* which is discussed in *Section III,D*. It is computed according to:

$$M_j^N = \begin{cases} 1, & if \quad C_{j,n}^{MAX} = C_{k,i}^{MAX} \\ 0, & elsewhere \end{cases} \tag{31}$$

where *i* indicates all the pulses that have initiation in the time interval $t_n - T_h < t_i < t_n + T_h$. Equation (31) was used to model the sequence of operation for the following combination of appliances: *j ~ Washing machine* and *k ~ dryer*, or *j ~ oven* and *k ~ dishwasher*, or *j ~ Heat water* and *k ~ Hairdryer*.

Penalty for Temperature: Φ_j^{temp} corrects the wrong association of an appliance to a detected pulse according to the external temperature. This is used for heating/cooling appliances. Data regarding external temperature can be easily integrated in the load disaggregation algorithm according to web based weather APIs. It is computed according to:

For cooling (*j* refers to air condition unit):

$$\Phi_j^{temp} = \begin{cases} -1, & if \quad Temp(t_n) < 20^o C \\ 0, & elsewhere \end{cases} \tag{32a}$$

For heating (*j* refers to aircondition, remote heater, heat pump)

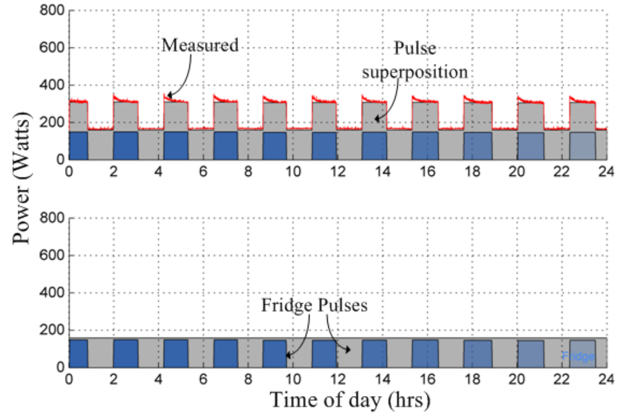

Fig. 7. Pulse to appliance disaggregation for a low energy day

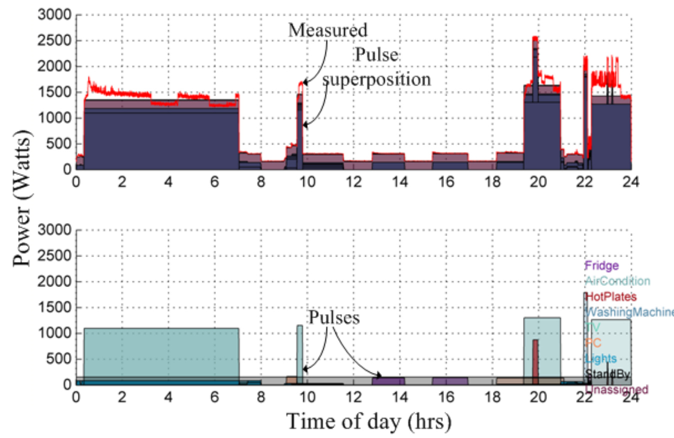

Fig. 8. Pulse to appliance disaggregation for a typical day

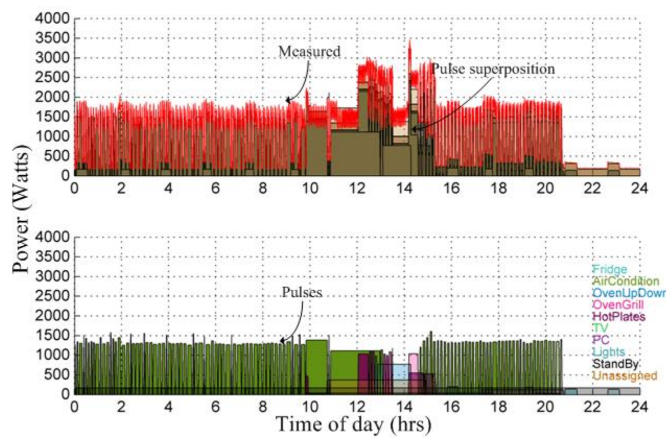

Fig. 9. Pulse to appliance disaggregation for a high energy day

$$\Phi_j^{temp} = \begin{cases} -1, & if \quad Temp(t_n) > 20^o C \\ 0, & elsewhere \end{cases} \tag{32b}$$

Penalty for Energy: Φ_j^{energy} is the penalty that corrects the wrong association of an appliance to a pulse according to maximum expected daily energy consumption of appliances. For example, an oven cannot consume more than 10kWh of energy per day, or a hairdryer more than 1kWh per day. This is *a priori* information that is integrated to an appliance once a user registers the appliance in the system (discussed in *Section IV, B*). The metric is computed as:

1530-437X (c) 2015 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2015.2501422, IEEE Sensors Journal

TABLE II
ACCURACY OF PROPOSED ALGORITHM

Daily	Pulse to Appliance association accuracy
Extreme low energy scenario	100%
Typical day scenario	85%
Extreme high energy scenario	70%
Monthly	Pulse to Appliance association accuracy
Month 1	78%
Month 2	83%

$$\Phi_j^{energy} = \begin{cases} -1, & if \quad \sum_{n\in\Omega_j} \frac{\lambda \cdot P_n \cdot \tau_n}{3600} > E_j^{apriori} \\ 0, & if \quad \sum_{n\in\Omega_j} \frac{\lambda \cdot P_n \cdot \tau_n}{3600} \le E_j^{apriori} \end{cases} \tag{33}$$

where $n \in \Omega_j$ denotes all the pulses *n* that have been associated to the specific appliance *j*.

C. Pulse to Appliance Association

A pulse *n* is associated to an appliance *j* according to the coefficient C_j. We select the appliance with the maximum coefficient C_j as described in (34) to indicate the most probable candidate appliance for the pulse:

$$f : n \mapsto \Omega_j = \max_j \lfloor C_j \rfloor, \; j \in L \tag{34}$$

VI. SIMULATION RESULTS

A. Daily results

We examine three case studies. The first refers to an *extreme low energy* scenario where there was no human activity in the residential unit. Therefore, the energy consumption footprint was a result of the standby power consumption and refrigerator operation. The predictions are presented in Fig. 7 and Table II.

The second scenario concerns a *typical day* with an average human activity (Fig. 8). In most occasions, the proposed algorithm successfully detected the operation of the air-condition unit, the hot-plates and the heat water preparation of the washing machine. The multi-state pulses generated by the *rinse mode*, the entertainment equipment of the house (TV, PC, lights) and the refrigerator were not successfully detected at all times for reasons stated in Table III.

The third scenario refers to an *extreme high energy* scenario where the majority of the appliances were operating during the day (Fig. 9). It should be highlighted that this scenario is not met in reality but was used to test the performance of the algorithm. Hotplates and oven operation was successfully detected between 09:55am to 10:04am and 12:12pm to 01:40pm and 02:24pm to 03:35pm. The proposed algorithm made a false detection of the aircondition unit and low power devices such as refrigeration, lights, PC and TV for reasons stated in Table III.

B. Activity detection

This section describes the performance of the algorithm for a two month period. Fig. 10 presents the probabilities of using specific appliances during one hour time blocks of a day. The

TABLE III
APPLIANCE DETECTION

Appliance Type	Time of Detection	Reason for False Detection
Typical Day Scenario		
Refrigerator	00:05am-08:00pm	High power pulses from air-conditioning unit, hotplates/oven
Washing machine/ Rinse mode	10:15pm-10:45pm	High power pulses from air-conditioning unit
PC/TV	08:00pm-11:00pm	High power pulses from air-conditioning unit and rinse mode of washing machine
Extreme High Energy Scenario		
Refrigerator	00:05am-07:00am	High power pulses from air-conditioning unit
Air-conditioning unit	12:30pm-02:05pm	High power pulses from hotplates and oven
PC/TV/Lights	06:00pm-08:00pm	Power pulses from high power appliances

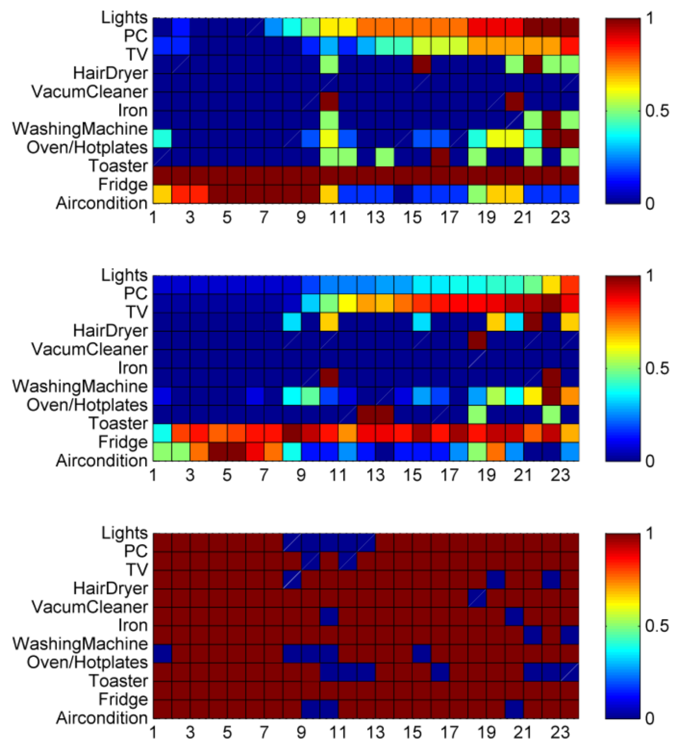

Fig. 10. User activity probabilities over a month period. First graph represents measured results, second graph the simulated results and third graph the correct detection of appliance usage

first two graphs present the measured and the simulated results. For each time block, the color plot associates the probability that a specific appliance was used by the user or detected by the algorithm. The third subplot gives a comparison of the successful detection of the appliances, assuming a 25% deviation. Red time zones indicate correct detection, whereas blue time zones indicate fault detections. It is observed that the simulated results gave a good approximation of the user activity in the household.

1530-437X (c) 2015 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSEN.2015.2501422, IEEE Sensors Journal

VII. Conclusions

The paper introduced a novel load disaggregation algorithm that decomposes smart meter power readings into a set of discrete pulses. The pulses are then associated to a registered appliance according to maximum likelihood procedure. The proposed algorithm also considers important characteristics of the appliances signatures, external environmental parameters and human behavior. These factors are introduced in a *n dimensional* space for decision making. An accuracy of more than 85% was observed. The proposed solution is cost efficient since it does not require reactive power readings at smart meter level and is based on low frequency sampling rates. These are important factors for the real implementation of smart grid energy services in modern business models.

VIII. Acknowledgements

The author would like to thank Kimatica Ltd and Dr. C. Tzaras for providing the metering infrastructure. This work was funded by the European Social Fund - ESF and Greek national funds through the Operational Program "Education and Lifelong Learning" ARISTEIA, MIS 372620, SOFON.

References

[1] J. ZicoKolter, Matthew J. Johnson, "REDD: A public data set for energy disaggregation research", in Proc. *SustKDD workshop, Data Mining Applications in Sustainability*, 2011.

[2] G. Hart, "Nonintrusive appliance load monitoring" *Proceedings of the IEEE*, vol. 80, no. 12, pp. 1870-1891, 1992.

[3] A. G. Ruzzelli, C. Nicolas, G. M. P. O'Hare, "Real-Time Recognition and Profiling of Appliances through a Single Electricity Sensor", *Sensor Mesh and Ad Hoc Communications and Networks* (SECON), 7th Annual IEEE Communications Society Conference on, 2010.

[4] H. Kim, M. Marwah, M. F. Arlitt, G. Lyon, J. Han, "Unsupervised Disaggregation of Low Frequency Power Measurements", 11th Int. Conf. *on Data Mining*, pp. 747-758, 2011.

[5] M. Figueiredo, A. De Almeida, B. Ribeiro, "Home electrical signal disaggregation for non-intrusive load monitoring (NILM) systems", *Journal Neurocomputing*, vol. 96, pp. 66-73, 2012.

[6] J. Liang, Simon K. K. Ng, G. Kendall and J. W. M. Cheng, "Load Signature Study—Part II: Disaggregation Framework, Simulation, and Applications", *IEEE Transactions on Power Delivery*, vol. 25, no. 2, April 2010.

[7] J. ZicoKolter, TommiJaakkola, "Approximate Inference in Additive Factorial HMMs with Application to Energy Disaggregation", *Proceedings of the International Conference on Artificial Intelligence and Statistics*, 2012.

[8] Lam, H. Y., Fung, G. S. K., Lee, W. K., 2007. "A Novel Method to Construct Taxonomy of Electrical Appliances Based on Load Signatures", *IEEE Trans. on Consumer Electronics*, vol. 53, no. 2, pp. 653-660, 2012.

[9] M. Weiss, A. Helfenstein, F. Mattern, T. Staake, "Leveraging smart meter data to recognize home appliances", in *Proc. IEEE Conf. Pervasive computing and communications* (PERCOM), March 2012.

[10] K. Chahine, et. al., "Electric Load Disaggregation in Smart Metering Using a Novel Feature Extraction Method and Supervised Classification", *Energy Procedia*, vol. 6, pp. 627-632, 2011.

[11] Z. Wang and G. Zheng, "Residential Appliances Identification and Monitoring by a Nonintrusive Method", *IEEE Trans. on Smart Grids*, vol. 3, no. 1, March 2012.

[12] M.D. Gillman, J.S. Donnal, J. Paris, S.B. Leeb, "Energy Accountability Using Nonintrusive Load Monitoring", *Sensors Journal, IEEE*, Vol. 14, no. 6, pp. 1923-1931, 2014.

[13] J. Paris, J.S. Donnal, S.B. Leeb, "NilmDB: The Non-Intrusive Load Monitor Database" *Smart Grid, IEEE Transactions on*, vol.5, no.5, pp.2459,2467, Sept. 2014.

[14] M. Pipattanasomporn, M. Kuzlu, S. Rahman, Y. Teklu, "Load Profiles of Selected Major Household Appliances and Their Demand Response Opportunities", *Smart Grid, IEEE Transactions on*, vol. 5, no. 2, pp. 742-750, 2014.

[15] G. Koutitas and L. Tassiulas, "Periodic Flexible Demand: Optimization and Phase Management in the Smart Grid", *Smart Grid, IEEE Transactions on*, vol. 4, no. 3, pp. 1305-1313, 2013.

George C. Koutitas (M'05) is an academic and entrepreneur in Wireless Networks and Smart Grids. He received the B.Sc. degree in Physics from Aristotle University of Thessaloniki Greece and the M.Sc. degree (with distinction) and PhD degree (EPSRC scholarship) from the University of Surrey UK. During his studies, he received the "Nokia Prize" for the best overall performance. His main research interests are in the area of Wireless Communications, Energy Efficient Networking and Smart Grids. He is also the founder of Gridmates, a Transactive Energy Platform designed to end energy poverty, a postdoc at the University of Thessaly and a visiting professor at Texas State University.

Leandros Tassiulas (S'89, M'91, SM/05 F/07), the John C. Malone Professor of Electrical Engineering at Yale University, obtained the Diploma in Electrical Engineering from the Aristotelian University of Thessaloniki, Greece in 1987, and the M.S. and Ph.D. degrees in Electrical Engineering from the University of Maryland, College Park, in 1989 and 1991, respectively. He has held positions as Assistant Professor at Polytechnic University New York (1991–1995), Associate Professor University of Maryland College Park (1995–2001), and Professor at University of Ioannina (1999–2001) and University of Thessaly (2002-2015), Greece. His research interests are in the field of computer and communication networks with emphasis on fundamental mathematical models, architectures and protocols of wireless systems, sensor networks, high-speed internet and satellite communications. Dr. Tassiulas is a Fellow of IEEE. He received a National Science Foundation (NSF) Research Initiation Award, an NSF CAREER Award an Office of Naval Research, Young Investigator Award and a Bodosaki Foundation award. He received the INFOCOM 1994 best paper award, the INFOCOM 2007 achievement award, and the IEEE 2016 Koji Kobayashi Computers and Communication Award.

1530-437X (c) 2015 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.