DOI: http://dx.doi.org/10.15446/ing.investig.v38n2.70550

Non-Intrusive Electric Load identification using Wavelet Transform

Identificación de cargas eléctricas por medios no invasivos empleando transformada de Wavelet

José A. Hoyo-Montaño[1], Naim León-Ortega[2], Guillermo Valencia-Palomo[3], Rafael A. Galaz-Bustamante[4], Daniel F. Espejel-Blanco[5], and Martín G. Vázquez-Palma[6]

ABSTRACT

This paper shows the development of a decision tree for the classification of loads in a non-intrusive load monitoring (NILM) system implemented in a simple board computer (Raspberry Pi 3). The decision tree uses the total energy value of the power signal of an equipment, which is generated using a discrete wavelet transform and Parseval's theorem. The power consumption data of different types of equipment were obtained from a public access database for NILM applications. The best split point for the design of the decision tree was determined using the weighted average Gini index. The tree was validated using loads available in the same public access database.

Keywords: Non-Intrusive Load Monitoring, Wavelet Transform, Decision Tree.

RESUMEN

El presente artículo muestra el desarrollo de un árbol de decisión para la clasificación de cargas en un sistema de monitoreo de cargas no-invasivo (NILM) implementado en un computadora de tarjeta sencillo tipo Raspberry Pi 3. El árbol de decisión emplea el valor de energía total de una señal de potencia de los equipos, el cual es generado empleando una transformada discreta de ondoleta y el teorema de Parseval. Los datos de consumo de potencia de diferentes tipos de equipos fueron obtenidos de una base de datos de acceso público para aplicaciones NILM. El punto de mejor ruptura para el diseño del árbol de decisión se determinó empleando el índice de Gini de promedio ponderado. El árbol fue validado empleando cargas disponibles en la misma base de datas pública.

Palabras clave: Monitoreo de cargas no-invasivo, transformada de ondoleta, árbol de decisión.

Received: February 20th 2018
Accepted: May 16th 2018

Introduction

Nowadays, the World is facing several challenges regarding energy usage, such as energy sources availability, carbon emission, sustainability, among others (Aiad & Lee, 2016b). Building energy management is becoming a major issue worldwide; it is estimated that nearly 40% of all the electric power is consumed in buildings (Ma *et al.*, 2016). Several countries, like China (Zhou *et al.*, 2015), European Un-ion (Tsai & Lin, 2012), and Mexico (Honorable Congreso de la Unión, 2012), have developed public policies to mitigate these challenges

Mexico has established its commitment with the environment and rational use of energy in the Climate Change General Law, which establishes a 30% reduction in the emission of greenhouse gases by 2020. This goal requires the deployment of incentive programs to use non-fossil fuels, and increase energy efficiency as an additional

[1] Industrial Electronics Engineering, Instituto Tecnológico de Hermosillo, Mexico. M.Sc. and Sc.D. Electronic Engineering, Centro Nacional de Investigación y Desarrollo Tecnlógico, Mexico. Affiliation: Full-Time Professor-Researcher, TecNM/Instituto Tecnológico de Hermosillo, Mexico. E-mail: j.hoyo@ieee.org.

[2] Electronics Engineering, Instituto Tecnológico de Hermosillo, Mexico. Master on Electronics Engineering, TecNM/Instituto Tecnológico de Hermosillo, Mexico. Affiliation: Full-Time Graduate Student, TecNM/Instituto Tecnológico de Hermosillo, Mexico. E-mail: naim.leon@gmail.com.

[3] Electronics Engineer, Instituto Tecnológico de Mérida, Mexico. M.Sc. Electronics Engineering, Centro Nacional de Investigación y Desarrollo Tecnlógico, Mexico, Ph.D. Automatic Control and Systems Engineering, University of Sheffield, U.K. Affiliation: Full-Time Professor-Researcher, TecNM/Instituto Tecnológico de Hermosillo, Mexico. E-mail: gvalencia@ith.mx.

[4] Industrial Electronics Engineering, Instituto Tecnológico de Hermosillo, Mexico. M.Sc. Computer Science, Instituto Tecnológico de Nogales, Mexico. Affiliation: Full-Time Professor, TecNM/Instituto Tecnológico de Hermosillo, Mexico. E-mail: rafael_galaz_b@yahoo.com.mx.

[5] Industrial Electronics Engineering, Instituto Tecnológico de Hermosillo, Mexico. M.Sc. Industrial Engineering, Instituto Tecnológico de Hermosillo, Mexico. Affiliation: Full-Time Professor, TecNM/Instituto Tecnológico de Hermosillo, Mexico. E-mail: dfespejel@gmail.com.

[6] Electronics Engineering, Universidad Autónoma Metropolitana, Mexico. M.Sc. Computer Science, Instituto Tecnológico de Chihuahua, Mexico. Affiliation: Full-Time Researcher/Project Manager, Diseño e Ingeniería Sustentanble SA de CV, Mexico. E-mail: mgvpalma@gmail.com.

How to cite: Hoyo-Montaño, J.A., León-Ortega, N., Valencia Palomo, G., Galaz-Bustamante, R.A., Espejel-Blanco, D.F., Vázquez-Palma, M.G. (2018). Non-Intrusive Electric Load identification using Wavelet Transform. Ingeniería e Investigación, 38(2), 42-51.
DOI: 10.15446/ing.investig.v38n2.70550

action besides replacement of power generation technology (Honorable Congreso de la Unión, 2012).

The implementation of energy saving and/or efficiency actions, in particular regarding domestic installations, requires information on how the energy is used. Currently, energy meters provide information about total energy consumption through monthly bills, and they do not allow to determine individual equipment consumption (Aiad & Lee, 2016b).

There are studies that show a relationship between the knowledge of the amount of energy consumed by the equipment and the implementation of changes in the operating habits of the equipment by the users that promote energy savings, which may vary between 9% to 20% (Aiad & Lee, 2016a).

A suitable monitoring system is required to know the operation conditions of electrical appliances. The monitoring of these operations would facilitate the implementation of energy efficiency measures. The monitoring of loads in a general way is a process that seeks to identify and acquire the measurement of energy consumption of a particular load (I. Abubakar, Khalid, Mustafa, Shareef, & Mustapha, 2017).

Traditionally, the monitoring of the operation of connected equipment in a facility is based on the installation and operation of a large number of sensors. Each power outlet or load has a sensor, and the system is called intrusive monitoring (He, Stankovic, Liao, & Stankovic, 2016), having disadvantages of a high cost, complex installation and difficult maintenance issues (I. Abubakar *et al*., 2017).

As an alternative to the inconvenience of intrusive monitoring, some alternatives have been developed with a reduced number of sensors, forming a Non-Intrusive Load Monitoring (NILM) system. In this type of system, the goal is to disaggregate the individual load consumption from the total, and the application of voltage and current waveform analysis at a single point located at the service's point of entry (I. Abubakar *et al*., 2017). Figure 1 shows a general structure of a NILM system (H. H. Chang, Chen, Tsai, & Lee, 2012).

This paper presents an implementation of a NILM system for power consumption signature detection based on discrete wavelet transform, Parseval's theorem, and decision trees suitable to be executed in a Single Board Computer (SBC), just like Raspberry Pi 3. The system developed is part of a smart power-meter; its hardware is based on a Raspberry Pi 3 and an acquisition stage, as shown in Figures 2 and 3.

In the next section, a brief review of several approaches developed for load identification in NILM systems is presented. Wavelet transform structures suitable for NILM applications are presented afterward followed by the design process of the decision tree using data available in a public database for NILM applications. Finally, the results obtained from the validation of the design of the decision tree for the classification of loads, as well as the conclusions drawn, are presented.

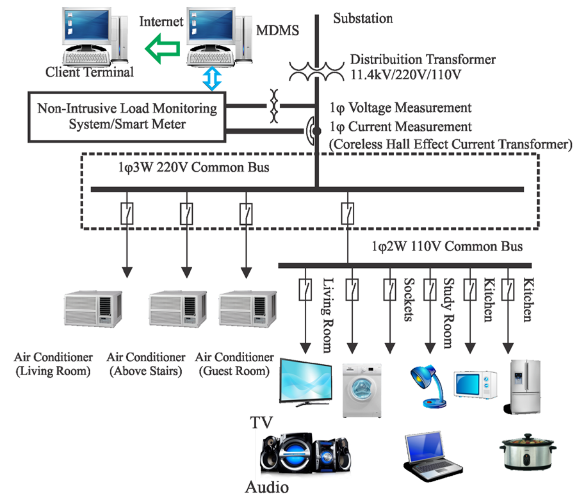

Figure 1. General structure of a NILM system.
Source: H. H. Chang *et al*. (2012)

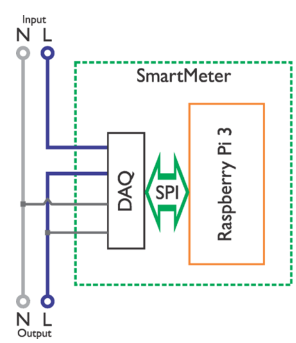

Figure 2. Simplified structure of Smart Power-Meter.
Source: Authors

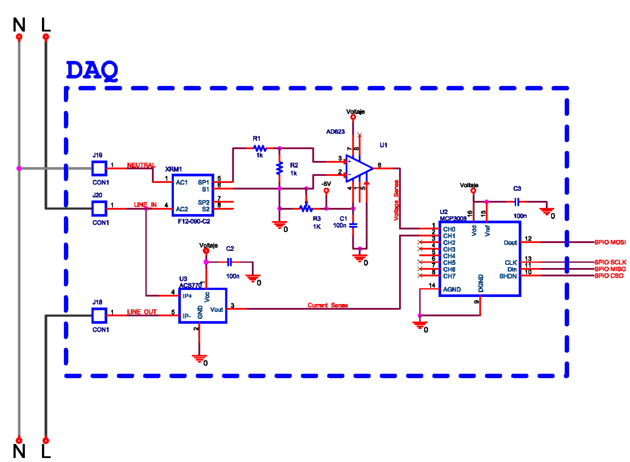

Figure 3. Simplified schematic of DAQ stage of Smart Power-Meter.
Source: Authors

Non-Intrusive Load Identification

A NILM system analyzes voltage and current waveforms trying to identify a power consumption signature that can be associated to the nature and operating state of individual devices. These power consumption signatures can be classified as steady state, transient and non-traditional signatures (I. Abubakar *et al.*, 2017).

Steady state signature is obtained when the device has completed its starting stage, and it has an steady operation, this identification uses parameters such as active power, reactive power, RMS voltage and current, power factor, and harmonic components (I. Abubakar *et al.*, 2017).

Transient signature is drawn from the analysis performed to period between the turn-on and steady states, or between the steady and turn-off state of a device, because during these periods some characteristic power consumption behaviors can be associated with specific loads (I. Abubakar *et al.*, 2017).

Non-traditional signatures, on the other hand, can be obtained using the values of non-electric variables in the load identification process. Values of temperature, lighting, time of day, start-up time, among others, are used to give context to the device usage. Information from these variables can be mixed with previous signatures to improve identification (I. Abubakar *et al.*, 2017).

To help in the identification process of the devices operating in an installation using NILM systems, a classification has been proposed (Bernard, Wohland, Klaassen, & Vom Bogel, 2016; Hart, 1992; Zoha, Gluhak, Imran, & Rajasegarar, 2012):

1. Type I. Turn-on/Turn-off

   There are only two possible operating states (turn-on and turn-off); a typical example is a lamp.

2. Type II. Finite State Machines

   They present several levels of defined consumption, and a cyclic operation. An example of these devices is the washing machine.

3. Type III. Continous Variable Consumption

   They have an infinite number of operating points when turned-on, examples of these devices are light dimmers and power tools. They are a great challenge for identification due to their power consumption nature.

4. Type IV. Continuous Consumption

   They operate during long periods of time, days or weeks, wireless phones, and any remote controlled appliance, are perfect examples of this type of devices.

Each one of these categories present its own complexity for the identification of individual devices.

Total aggregate power consumption of devices operating inside an electric installation can be described as (Hart, 1992):

$$P(t) = \sum_{i=1}^{n} a_i(t) \cdot P_i + e(t) \tag{1}$$

Where $P(t)$ is the power consumption over time, $a_i(t)$ is the activation vector of device i, with values 0 and 1 when device is off or on during time t; P_i is the power vector of device i, $e(t)$ is the error or noise term.

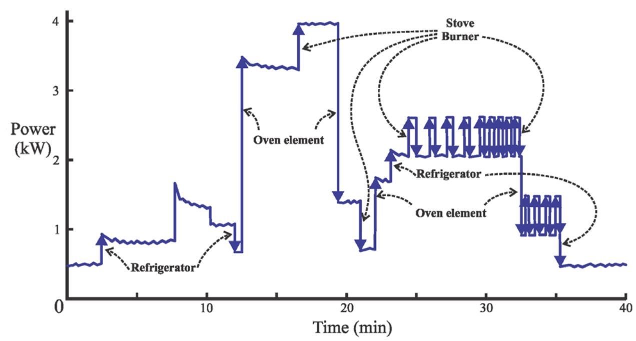

Figure 4. Aggregated power consumption from different devices.
Source: Hart (1992)

Figure 4 shows an example of aggregated power consumption from different devices.

Identification of Power consumption signature has been implemented in different schemes in literature. In general, NILM identification process requires implementing six stages of analysis (Basu, Debusschere, Douzal-Chouakria, & Bacha, 2015; Liang, Ng, Kendal, & Cheng, 2010):

1. Data Acquisition

   It is required to gather information of the steady and transient states from the power waveforms of the device. A high frequency sampling stage captures information regarding transient events; meanwhile, a low frequency sampling stage gathers steady state information of the device.

2. Data processing

   Data must be conditioned and processed in order to give meaningful information. This stage includes noise filtering, harmonic components separation, signal synchronicity, etc.

3. Event detection

   Processing and storage of all the information is inefficient and impractical process, so it is important to detect the activation and deactivation of the device. It is necessary to establish a threshold crossing detection mechanism for the detection of transient.

4. Characteristic Extraction

Electric parameters, such as active power, reactive power, harmonic components and transient waveforms, can be extracted from the event detection and data processing stages. The identified characteristics are depending of the disaggregation method used for load identification.

5. Load classification or disaggregation

Using the characteristic information gathered from the processed data, along with a known pattern, the device disaggregation can be performed from the total energy consumption, that is, the device can be identified.

6. Energy calculation

By identifying an individual device, its operation pattern and energy consumption can be estimated.

Using active power (P), reactive power (Q), RMS current and harmonic components has given good results in identifying type I and II devices; however, it has a poor performance with low power devices. High frequency alternatives have been developed to improve steady state analysis by including harmonic content. Because the transient behavior of the device turns out to be distinctive in many cases, implementing this type of analysis can facilitate the identification process, and requires the implementation of a high frequency sampling scheme (Isiyaku Abubakar, Khalid, Mustafa, Shareef, & Mustapha, 2015; Zoha *et al*., 2012).

There are some reports of performance improvement of steady-state analysis using voltage and current waveforms as a way to identify unique features of loads, such as peak and RMS values, phase difference and power factor (Zoha *et al*., 2012). These identifying methods can be complemented using harmonic analysis along with real and reactive power features to improve device detection of the algorithms, but its uses require high sampling rates of waveforms (Abubakar *et al*., 2015; Zoha *et al*., 2012).

Most devices have a distinctive transient behavior that can be suitable for device identification, using a high sampling rate is possible to capture the transient behavior (Abubakar *et al*., 2015; Zoha *et al*., 2012). Features such as transient shape and turn-on energy calculation have been used to identify individual devices (Zoha *et al*., 2012).

Several methods for transient information extraction that allow load identification have been developed. Some of these methods are based on short-time Fourier Transform (Liang, Ng, Kendall, & Cheng, 2010; Marcu & Cernazanu, 2012; Zoha *et al*., 2012), Markov Models (Aiad & Lee, 2016a; Kim, Marwah, Arlitt, Lyon, & Han, 2011; Kong *et al*., 2016; Mueller & Kimball, 2016), and Wavelet Transform (Alshareef & Morsi, 2015; H.-H. H. Chang, Lian, Su, & Lee, 2014; J. M. Gillis, Alshareef, & Morsi, 2016; J. Gillis & Morsi, 2016), among others.

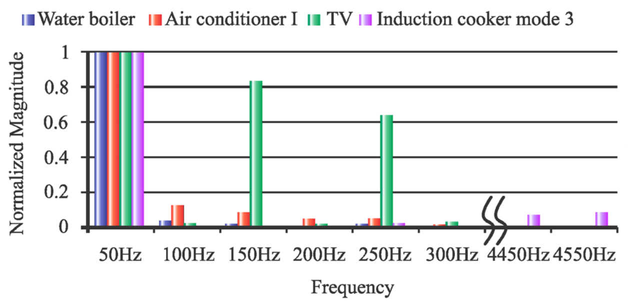

Figure 5. Frequency spectrum of diverse devices.
Source: Liang, Ng, Kendall, *et al*. (2010)

Fourier Transform usage for spectral analysis of power consumption has been proven useful to detect variable loads. To detect the operation of a device, and estimate its energy consumption, Short-time Fourier Transform, and active and reactive power calculation has been combined (Zoha *et al*., 2012). This mathematic tool performs the transformation of a time domain function into a frequency domain function (Marcu & Cernazanu, 2012). Figure 5 shows a spectral analysis reported by (Liang, Ng, Kendall, *et al*., 2010) for different devices, it can be noticed that both TV and Air Conditioners have a strong presence of low-order harmonics, meanwhile devices such as induction pots present a high content of high-order harmonics.

Markov models have become an interesting alternative to implement NILM systems due to their simplicity to model basic functions. The general scheme to implement Markov Models, in particular Hidden Markov Models (HMM), is based in the fact that a device behavior can be represented as a latent state and an observable output, usually active power. A system with trained Markov models can perform inferences regarding the most probable state sequence of a device, based in the set of measures processed. HMM have been proven useful in predicting in a precise way the behavior of devices using measurements gathered with low-frequency sampling (<1Hz). The goal of a HMM-based NILM system is to generate energy consumption profiles and to determine time of use of each devices operating in an installation. Usually this information is considered as non-critical, and its processing is performed off-line (Mueller & Kimball, 2016).

Markov modeling requires a periodic acquisition of T measurements, where each measurement is assumed to be associated with a state Q from de process, and each state of the process can assume one of N possible values. HMM have a three component structure. A Transition A matrix containing the state-transition probability values; each state probability ϕ, and a vector with the values of the initial state occupation probabilities π. Figure 6 shows the structure of a HMM for a device with only three states (Mueller & Kimball, 2016).

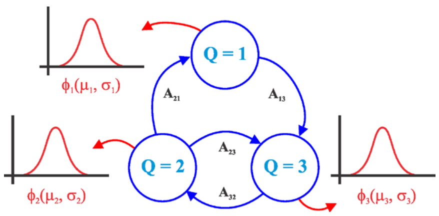

Figure 6. Structure of a HMM for a three state device.
Source: Mueller & Kimball (2016)

Identification of each device in an aggregate total power consumption requires firstly the determination of the sequence of states that can be used to compose the observation sequence. Using Matrix A and ϕ, the inferred state sequence is calculated, it represents the most probable behavior of all the devices represented as a unity. The state sequence Q can be used to determine which state sequence has the highest probability of occurrence for each individual device (Mueller & Kimball, 2016).

Wavelet Transform is another tool that has been used to perform transient analysis of a device (Zoha *et al.*, 2012). Analysis based on Wavelet Transform performs an extraction of the desired waveform applying a function translation and dilation process (Chen, Chang, & Chen, 2013). A more detailed discussion about Wavelet Transform is presented next.

Wavelet Transform and Parseval's Theorem

Wavelet Transform can be implemented in two ways: Continuous Wavelet Transform (CWT), and Discrete Wavelet Transform (DWT). DWT has a structure more suitable for digital signal analysis. DWT is derived from CWT definition, and can be expressed as (Chen *et al.*, 2013):

$$\mathrm{DWT}_{\mathrm{m,n}} = \frac{1}{\sqrt{2^{\mathrm{m}}}} \int_{-\infty}^{+\infty} x(t)\psi^{*}\left(a_0^{-m}t - nb_0\right)dt \qquad (2)$$

Where $x(t)$ is the signal analyzed, ψ is the mother wavelet applied, a_0 is the scaling factor, and b_0 is the shift factor.

Equation (2) can be transformed in:

$$DWT_{m,n} = 2^{-m/2} \sum_k x[k]\, g\left[a_0^{-m} n - kb_0\right] \qquad (3)$$

Setting, and, (3) becomes:

$$DWT_{m,n} = 2^{-m/2} \sum_k x[k]\, g\left[2^{-m} n - k\right] \qquad (4)$$

DWT performs two operations, dilation (applying scaling factors), and translation (applying shifting factors). These operations are performed to decompose a signal into a series of short duration waveforms called Mother Wavelets. Mother Wavelet has characteristics suitable for transient events analysis (H. H. Chang *et al.*, 2012). Multi-Resolution Analysis (MRA) is based on the application of DWT. MRA decompose a complex waveform or signal into several sets of simpler waveforms, this is performed by a set of low-pass g[n] and high-pass h[n] filters (Chen *et al.*, 2013). Figure 7 shows a three-level DWT filter structure.

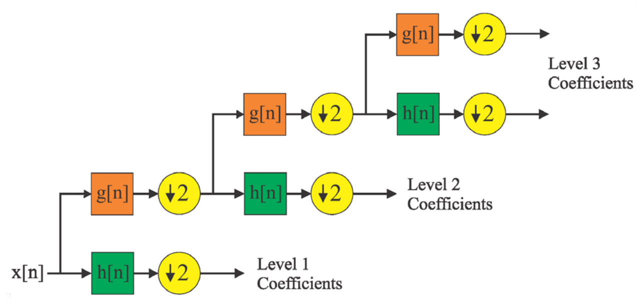

Figure 7. Three-level DWT structure.
Source: Authors

This type of structure provides a multilayer decomposition scheme, where the last low-pass filter g[n] gives an approximation value (level 3 in Figure 7), meanwhile, high-pass filters h[n] provides detail values (Figure 7 shows three values). An increase in the number of levels will provide an increase in the number of detail values, but only one approximation value will be obtained (Chen *et al.*, 2013).

MRA in conjunction with Parseval's Theorem has been used to obtain load power index as an element suitable for load identification in a NILM system (Alshareef & Morsi, 2015; H.-H. H. Chang *et al.*, 2014; J. M. Gillis *et al.*, 2016; J. Gillis & Morsi, 2016).

Parseval's Theorem is used to calculate the energy dissipated by a 1Ω resistor when a discrete current $f[n]$ flows through it. The Theorem uses the Fourier Transform coefficients (Kocaman & Özdemir, 2009)

$$\frac{1}{N}\sum_n \left|f(n)\right|^2 = \sum_k \left|a_k\right|^2 \qquad (5)$$

where N is the sampling period, and a_k are the Fourier Transform coefficients.

In order to apply (5) to DWT, it is transformed in:

$$\frac{1}{N}\sum_t \left|f(t)\right|^2 = \frac{1}{N_J}\sum_k \left|a_k\right|^2 + \sum_{j=1}^{J}\left(\frac{1}{N_J}\sum_k \left|d_j(k)\right|^2\right) \qquad (6)$$

The first term in the right part of (6) represents the energy levels of the approximation component of DWT, the second component represents the energy levels of the detail components. The total energy of the DWT is transformed in:

$$P_J = \frac{1}{N_J}\sum_k \left|d_{j,k}\right|^2 = \frac{\left\|d_j\right\|^2}{N_J} \tag{7}$$

Where $\|d_j\|$ is the norm of the expansion coefficients, and N_J is the number of samples used at level J.

Load identification

Load identification from total energy of the decomposition with DWT was performed using a Decision Tree (DT). In the NILM system implemented, six types of loads were defined for identification: Air Conditioning (class 0), Compact Fluorescent Lamp (class 1), Fan (class 2), Refrigerator (class 3), Vacuum cleaner (class 4), and Washing Machine (class 5).

In a NILM system, each type of load has hidden information, when developing a DT, the main goal is to develop a classification tree that contains an optimal entry node capable to measure impurities in the tree nodes, this can be performed using the Gini Index (Alshareef & Morsi, 2015; J. M. Gillis *et al*., 2016; J. Gillis & Morsi, 2016)

$$Gini(\sigma) = 1 - \sum_{c=0}^{C-1}\left[f(c\,|\,\sigma)\right]^2 \tag{8}$$

Where C is the number of classes, and $f(c|\sigma)$ is the probability that σ belongs to class c.

The design procedure for a DT suitable for classification can be seen in Figure 8.

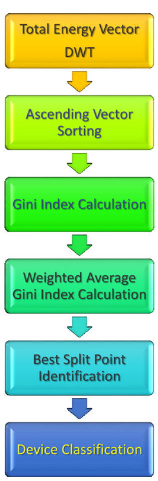

Figure 8. Procedure for best split point identification for Classification DT.
Source: Authors

Considering that there are six types or classes of devices, and that an eight-level DWT analysis was applied using a Daubechies 3 mother wavelet, the total energy values of DWT for one level of approximation (A8) and eight levels of detail (D1 to D8) were obtained. The number of levels is directly related to the harmonic spectrum covered within the sampling frequency; hence the eight-level analysis was chosen to cover a frequency analysis from 30kHz (sampling frequency) to 117,18Hz. The sampling frequency for the analysis is set in the PLAID public database (Gao, Giri, Kara, & Bergés, 2014). To calculate the Total Energy Vector (TEV) from the DWT, a Python function was developed and its results were compared with the results from a MatLab Wavelet Toolbook function. The Python function has a quadratic average error of 0,22%, this accuracy in the calculation of the TEV helps to differentiate appliances with close energy signatures, the code of the Python function is shown in Figure 9.

```
def convolucion(x,y):                # ”x” signal vector and “y” filter coeficients for convolution

    res=[]                           # starting resulting vector
                                     # convolution operations
    for n in range (0,len(x)):       # for each x[n] perform filtering
        cn=0                         # reseting value for each iteration
        for k in range (0, len(y)):
            if (n>=len(y)):          # first k values of x[n] are assumed 0,
                                     # since n-k can not be negative
                cn=cn+(x[n-k]*y[k])  # accumulated products
        if n%2 ==0:                  # downsampling
            res.insert(0,cn)         # result inserted at begining of vector
     return res

def DWTdb3AT8(s):                    # DWT calculation

    # Daubechies 3 Low-Pass coefficients
    PB=[0.035226291882101,-0.085441273882241,-0.135011020010391,
        0.459877502119331,0.806891509313339,0.332670552950957]
    # Daubechies 3 High-Pass coefficients
    PA=[-0.332670552950957,0.806891509313339,-0.459877502119331,
        -0.135011020010391,0.085441273882241,0.035226291882101]

    # 1st Level
    D1=convolucion(s,PA)    # Detail value at level 1
    A1=convolucion(s,PB)    # Aproximation Value at level 1

    # niveles 2 a 8
    for i in range (2,9):
        vars()["D%d"%i]=convolucion(vars()["A%d"%(i-1)],PA) # Detail value at level i
        vars()["A%d"%i]=convolucion(vars()["A%d"%(i-1)],PB) # Aproximation Value at level i

    return [vars()["A8"], vars()["D1"], vars()["D2"],
            vars()["D3"], vars()["D4"], vars()["D5"],
            vars()["D6"], vars()["D7"], vars()["D8"]]

def energia (v):                     # DWT energy calculation
    tot=0
    for j in range (0,len(v)):
        for i in range (0,len(v[j])):
            tot=tot+v[j][i]**2
    E=[]
    for i in range(0,len(v)):
        ene=0
        for j in range (0,len(v[i])):
            ene=ene+v[i][j]**2
        E.append(ene)
    return E
```

Figure 9. Python Code of DWT and Total Energy Vector Calculations.
Source: Authors

Table 1 shows the DWT total energy values of 48 devices taken form PLAID. Below is the procedure to find the best split point of the classification DT.

Total Energy List sorting

The first step required to find the best split point is to sort in ascending way the values of the Total Energy of DWT List before the calculation of Gini Index. The sorted List is partially shown in Column 1 of Table 2. Column 2 shows the mid-point between two adjacent values of Total Energy of DWT.

Split point calculation based on mid-points

Columns 3 and 4 from Table 2 show the number of devices of each class with values lesser (column 3) or greater/equal (column 4) to the mid-point value of column 2. It can be seen in Table 2, for instance, that the total number of devices with Total Energy values greater or equal to 1 465,72 is 34, and they are dispersed as: six class 5, eight class 4, eight class 3, one class 2, none class 1, and eleven class 0.

Table 1. Total Energy of DWT analysis

Device PLAID ID	Total Energy	Class
15	3 128,14	0
39	2 386,47	0
84	52 159,26	0
85	4 067,64	0
86	22 469,92	0
87	55 402,56	0
161	2 008,95	0
163	1 977,55	0
288	35 150,20	0
289	63 776,78	0
291	59 808,00	0
2	777,84	1
4	936,63	1
20	282,13	1
41	334,34	1
44	265,80	1
88	320,84	1
67	529,16	2
100	236,73	2
101	239,68	2
171	1 464,76	2
172	1 466,69	2
23	1 664,17	3
46	6 080,07	3
105	14 321,40	3
107	14 807,57	3
195	1 251,14	3
197	59 808,00	3
199	19 395,88	3
333	13 712,12	3
334	716,24	3
528	3 979,71	3
55	23 646,49	4
56	22 491,94	4
698	21 753,79	4
726	4 291,99	4
730	3 678,12	4
811	3 328,17	4
812	3 169,71	4
840	21 465,38	4
80	247,05	5
185	31 078,20	5
186	24 502,86	5
188	61 448,42	5
492	22 979,34	5
535	1 008,55	5
620	8 064,26	5
622	9 299,02	5

Source: Authors

Table 2. Sorted List of Total Energy of DWT

Total Energy	Mid-Point	<	>=	Class	Gini_WA
236.7		0	8	5	
		0	8	4	
	238.20	0	10	3	0.8014
		1	4	2	
		0	6	1	
		0	11	0	
			.		
			.		
			.		
		2	6	5	
		0	8	4	
		2	8	3	
	1465.72	4	1	2	0.7354
		6	0	1	
		0	11	0	
1466.7					
		2	6	5	
		0	8	4	
		2	8	3	
	1565.42	5	0	2	0.7242
		6	0	1	
		0	11	0	
1664.2					
		2	6	5	
		0	8	4	
		3	7	3	
	1820.85	5	0	2	0.7278
		6	0	1	
		0	11	0	
.	.		.		.
.	.				.
.	.				.
.	.				.
			.		
		8	0	5	
		8	0	4	
		10	0	3	
	62612.6	5	0	2	0.8067
63776.8		6	0	1	
		10	1	0	

Source: Authors

Gini Index and its weighted average

The Gini Index shows a measurement of the impurity of a node, when its value gets to a minimum, the point of best split is found. Since there are two columns of membership (columns 3 and 4), a value that includes them must be obtained, this is done by calculating the weighted average of the Gini Index using:

$$Gini_{WA}\left(\sigma\right)=\frac{\text{Ó}_1}{N}\cdot Gini\left(\sigma\right)_a+\frac{\text{Ó}_2}{N}\cdot Gini\left(\sigma\right)_b \qquad (9)$$

Where Σ_1 and Σ_2 are the number of devices with Total Energy values lesser than, and greater/equal than mid-point value; $Gini(\sigma)_a$ is the total number of devices; and $Gini(\sigma)_b$ are the Gini Index for devices with Total Energy values lesser than, and greater/equal than mid-point value. The results of this calculation are shown in column 6 of Table 2.

Best split point identification

Because the Gini Index measures the impurity of a node, the minimum value of this index corresponds to the best split point, and identifying the entry point of the DT. By inspection of Table 2, it is found that the minimum value of $Gini_{WA}(\sigma)$ is 0,7242; which corresponds to an average point of total energy of 1 565,42. The tree developed from this entry point is shown in Figure 10.

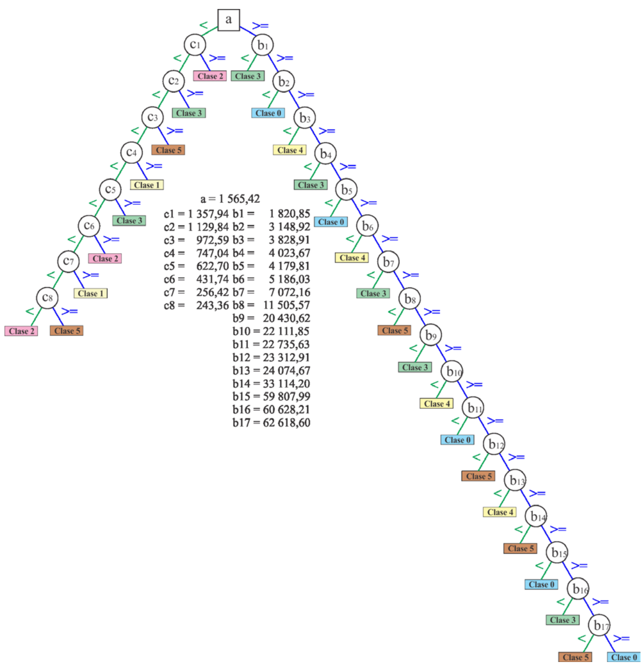

Figure 10. Classification DT based on the values of Table 2.
Source: Authors

The Classification DT developed presents an unbalanced structure; this is not rare when the Gini Index is used. There are eight nodes at the left (Energy lesser than), and seventeen to its right (Energy greater/equal than) of the entry node. The implementation of the Classification DT as a Python function is shown in figure 11.

```
# Desition Tree Classifiction function

def Clasificador(EN,DI):
    # EN es la energía del DWT
    # DI es el nivel de energía del componente D4 de DWT
    a=1565.42
    c=(1357.94,1129.84,972.59, 747.04, 622.70, 431.74, 256.42, 243.36)
    c1=(2,3,5,1,3,2,1,5,2)
    b=(1820.85, 3148.92, 3828.91, 4023.67, 4179.81, 5186.03, 7072.16,
       11505.57, 20430.62, 22111.85, 22735.63, 23312.91, 24074.67, 33114.20,
       59807.99, 60628.21, 62618.60)
    c2=(3,0,4,3,0,4,3,5,3,4,0,5,4,5,0,3,5,0)

    if DI > 0:
        return 1                    # Lámpara fluorescente compacta
    else:
        if EN < a:                  # ir a rama izquierda del árbol de decisión
            for i in range (0, 8):
                if EN >= c[i]:
                    return c1[i]
            return c1[8]
        else:                       # ir a rama derecha del árbol de decisión
            for i in range (0,17):
                if EN < b[i]:
                    return c2[i]
            return c2[17]
```

Figure 11. Python function code for the Classification DT.
Source: Authors

Simulation Results

The classification DT was tested using 30 devices from the PLAID database (Gao *et al.*, 2014), these devices were not included in the design process. Table 3 shows the total energy values used for testing, the identification result from de DT, the device class, and if there was an error in the identification process.

As it can be seen in Table 3, only three devices were wrongly identified, this means that the proposed classification DT has a 90% success rate identifying load class. Some HMM solutions had reported a success rate between 51,66% and 87% (Aiad & Lee, 2016a, 2016b) using REDD database; using DWT as part of the analysis process. Alshareef (2015) using a Daubechies 3 mother wavelet reached a 95,83% using 1 000 decision trees; Chang (2014) reported a DWT with an Artificial Neural Network reaching identification levels between 86,16 and 96,82%; Gillis (2016) reported a DWT plus Decision Tree, using a Daubechies 3 mother wavelet and six levels of decomposition, having a 96,18% success in load identification.

Table 3. Simulation Results from Classification DT for Load ID

ID	Total Energy	DT Class	Device Class	Error
17	2 691,72	0	0	
40	2 462,63	0	0	
42	268,07	1	1	
62	380,35	1	1	
65	525,36	2	2	
81	424,36	1	1	
83	57 643,11	0	0	
104	236,59	2	2	
106	13 384,11	3	3	
129	13 260,85	3	3	
162	1 974,22	0	0	
170	1 504,23	2	2	
196	21 084,15	4	3	1
290	37 474,11	0	0	
460	10 776,41	5	3	1
461	15 290,87	3	3	
461	15 290,87	3	3	

488	27 902,67	5	5	
490	22 838,07	5	5	
492	22 979,34	5	5	
531	1 023,33	5	5	
534	976,32	5	5	
621	9 604,69	5	5	
623	9 713,81	5	5	
699	22 355,30	0	4	1
727	4 319,89	4	4	
729	3 786,51	4	4	
813	3 365,17	4	4	
814	3 395,05	4	4	
840	21 465,38	4	4	

Source: Authors

Conclusions

A Classification Decision Tree was developed to identify six different type of devices included in the PLAID public database. Device data was processed using a Python script in a Raspberry Pi 3 platform to obtain the total energy values extracted from DWT analysis.

The classification DT design was performed applying weighted average Gini Index to identify the best split point, leading to an unbalanced tree, something not rare in this kind of design procedure.

The percentage of success in the identification of the classes of equipment of 90%, this implies that it is necessary to improve the structure of the decision tree. However, the proposed approach shows a better performance than the HMM presented by Aiad and Lee (2016a, 2016b), this must be taken with the proper reservation since two different appliance databases were used for validation; both Alshareef (2015) and Gillis (2016), implemented a DWT using Daubechies 3 mother wavelet to identify appliances with a better identification rate, but both solutions are more complex and requiere more computational effort. For this matter, it is planned to improve the measurement of the total energy of the equipment during the turn-on transient period to identify additional parameters that would help the implementation of a better process of identification of the loads.

Acknowledgements

The authors would like to thank the support received from Tecnológico Nacional de México, and CONACyT through Support for Scientific Research, Applied, Technological Development and Innovation 2016 program, code 5751.16-P, and the Research Stimulus Program 2017, modality PROINNOVA, code 240445.

References

Abubakar, I., Khalid, S. N., Mustafa, M. W., Shareef, H., & Mustapha, M. (2015). An Overview of Non-Intrusive Load Monitoring Methodologies. In *2015 IEEE Conference on Energy Conversion (CENCON)*, (pp. 54–59). Johor Bahru, Malaysia: IEEE.

Abubakar, I., Khalid, S. N., Mustafa, M. W., Shareef, H., & Mustapha, M. (2017). Application of load monitoring in appliances' energy management – A review. *Renewable and Sustainable Energy Reviews*, *67*, 235–245. http://doi.org/10.1016/j.rser.2016.09.064

Aiad, M., & Lee, P. H. (2016a). Non-intrusive load disaggregation with adaptive estimations of devices main power effects and two-way interactions. *Energy and Buildings*, *130*, 131–139. http://doi.org/10.1016/j.enbuild.2016.08.050

Aiad, M., & Lee, P. H. (2016b). Unsupervised approach for load disaggregation with devices interactions. *Energy and Buildings*, *116*, 96–103. http://doi.org/10.1016/j.enbuild.2015.12.043

Alshareef, S., & Morsi, W. G. (2015). Application of wavelet-based ensemble tree classifier for non-intrusive load monitoring. In *2015 IEEE Electrical Power and Energy Conference (EPEC)* (pp. 397–401). London, ON, Canada: IEEE. http://doi.org/10.1109/EPEC.2015.7379983

Basu, K., Debusschere, V., Douzal-Chouakria, A., & Bacha, S. (2015). Time series distance-based methods for non-intrusive load monitoring in residential buildings. *Energy and Buildings*, *96*, 109–117. http://doi.org/10.1016/j.enbuild.2015.03.021

Bernard, T., Wohland, D., Klaassen, J., & Vom Bogel, G. (2016). Combining several distinct electrical features to enhance nonintrusive load monitoring. In *Proceedings - 2015 International Conference on Smart Grid and Clean Energy Technologies, ICSGCE 2015* (pp. 139–143). Offenburg, German. http://doi.org/10.1109/ICSGCE.2015.7454285

Chang, H.-H. H., Lian, K.-L. L., Su, Y.-C. C., & Lee, W.-J. J. (2014). Power-spectrum-based wavelet transform for nonintrusive demand monitoring and load identification. *IEEE Transactions on Industry Applications*, *50*(3), 2081–2089. http://doi.org/10.1109/TIA.2013.2283318

Chang, H. H., Chen, K. L., Tsai, Y. P., & Lee, W. J. (2012). A new measurement method for power signatures of nonintrusive demand monitoring and load identification. *IEEE Transactions on Industry Applications*, *48*(2), 764–771. http://doi.org/10.1109/TIA.2011.2180497

Chen, K.-L., Chang, H.-H., & Chen, N. (2013). A new transient feature extraction method of power signatures for Nonintrusive Load Monitoring Systems. In *2013 IEEE International Workshop on Applied Measurements for Power Systems (AMPS)* (pp. 79–84). Aachen: IEEE. http://doi.org/10.1109/AMPS.2013.6656230

Gao, J., Giri, S., Kara, E. C., & Bergés, M. (2014). PLAID: a public dataset of high-resoultion electrical appliance measurements for load identification research. In *Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings - BuildSys '14* (pp.

198–199). New York, New York, USA: ACM Press. http://doi.org/10.1145/2674061.2675032

Gillis, J. M., Alshareef, S. M., & Morsi, W. G. (2016). Nonintrusive Load Monitoring Using Wavelet Design and Machine Learning. *Smart Grid, IEEE Transactions On*, *PP*(99), 320–328. http://doi.org/10.1109/TSG.2015.2428706

Gillis, J., & Morsi, W. G. (2016). Non-intrusive load monitoring using orthogonal wavelet analysis. In *2016 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)* (pp. 1–5). Vancouver, BC, Canada: IEEE. http://doi.org/10.1109/CCECE.2016.7726786

Hart, G. W. (1992). Nonintrusive appliance load monitoring. *Proceedings of the IEEE*, *80*(12), 1870–1891. http://doi.org/10.1109/5.192069

He, K., Stankovic, L., Liao, J., & Stankovic, V. (2016). Non-Intrusive Load Disaggregation using Graph Signal Processing. *IEEE Transactions on Smart Grid*, *PP*(99), 1–9. http://doi.org/10.1109/TSG.2016.2598872

Honorable Congreso de la Unión. Ley general de cambio climático (2012). Mexico: Camara de Diputados del H. Congreso de la Unión.

Kim, H., Marwah, M., Arlitt, M. F., Lyon, G., & Han, J. (2011). Unsupervised Disaggregation of Low Frequency Power Measurements. In *11th SIAM International Conference on Data Mining* (pp. 747–758). Phoenix, Arizona, USA: SIAM. Retrieved from http://epubs.siam.org/doi/pdf/10.1137/1.9781611972818.64

Kocaman, C., & Özdemir, M. (2009). Comparison of statistical methods and wavelet energy coefficients for determining two common PQ disturbances: Sag and swell. In *International Conference on Electrical and Electronics Engineering, 2009. ELECO 2009.* (pp. 80–84). Bursa, Turkey: IEEE. http://doi.org/10.1109/ELECO.2009.5355235

Kong, W., Dong, Z. Y., Ma, J., Hill, D., Zhao, J., & Luo, F. (2016). An Extensible Approach for Non-Intrusive Load Disaggregation with Smart Meter Data. *IEEE Transactions on Smart Grid*, *PP*(99), 1–1. http://doi.org/10.1109/TSG.2016.2631238

Liang, J., Ng, S. K. K., Kendal, G., & Cheng, J. W. M. (2010). Load signature study - part II: Disaggregation framework, simulation and applications. *IEEE Transactions on Power Delivery*, *25*(2), 561–569. http://doi.org/10.1109/PES.2010.5589424

Liang, J., Ng, S. K. K., Kendall, G., & Cheng, J. W. M. (2010). Load Signature Study - Part I : Basic Concept , Structure and Methodology. *IEEE Transactions on Power Delivery*, *25*(2), 551–560.

Ma, L., Liu, N., Wang, L., Zhang, J., Lei, J., Zeng, Z., … Cheng, M. (2016). Multi-party energy management for smart building cluster with PV systems using automatic demand response. *Energy and Buildings*, *121*, 11–21. http://doi.org/10.1016/j.enbuild.2016.03.072

Marcu, M., & Cernazanu, C. (2012). Dynamic Analysis of Electronic Devices' Power Signatures. In *2012 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)* (pp. 117–122). Graz: IEEE.

Mueller, J. A., & Kimball, J. W. (2016). Accurate Energy Use Estimation for Nonintrusive Load Monitoring in Systems of Known Devices. *IEEE Transactions on Smart Grid*, *PP*(99), 1–12. http://doi.org/10.1109/TSG.2016.2620120

Tsai, M.-S. S., & Lin, Y.-H. H. (2012). Modern development of an Adaptive Non-Intrusive Appliance Load Monitoring system in electricity energy conservation. *Applied Energy*, *96*, 55–73. http://doi.org/10.1016/j.apenergy.2011.11.027

Zhou, Z., Zhang, S., Wang, C., Zuo, J., He, Q., & Rameezdeen, R. (2015). Achieving energy efficient buildings via retrofitting of existing buildings: a case study. *Journal of Cleaner Production*, *112*, 3605–3615. http://doi.org/10.1016/j.jclepro.2015.09.046

Zoha, A., Gluhak, A., Imran, M. A., & Rajasegarar, S. (2012). Non-intrusive Load Monitoring approaches for disaggregated energy sensing: A survey. *Sensors (Switzerland)*, *12*(12), 16838–16866. http://doi.org/10.3390/s121216838