University of Alberta

Decomposition Techniques for Power System Load Analysis

by

Ming Dong

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Energy Systems

Department of Electrical & Computer Engineering

Abstract

In recent years, the increased public awareness of energy conservation has attracted serious attention to detailed energy consumption monitoring and management for the end users of power system. Load decomposition is a technique that can extract detailed sub-load information from compound load information. This technique decomposes a compound load such as an entire residential house into specific sub-load levels such as different home appliances by using only the aggregated metering data of the compound load. Through load decomposition, users can better understand the usage patterns of individual loads or load groups and therefore decide on how to save energy. On the utility side, load decomposition can be very helpful for load forecast, demand response program development, and Time-Of-Use price design.

In the past, traditional methods are either too costly or inaccurate. Therefore, some researchers proposed a non-intrusive load monitoring (NILM) approach that can identify and track major sub-loads based on only the total signal collected from the meter-side with acceptable error. Recently, the vast deployment of smart meters has raised considerable interests in this approach. However, many critical problems still need to be solved before it truly becomes technically available for ordinary end-users.

To solve the above problems, at the beginning, this thesis presents a novel NILM method based on event detection and load signature studies. The key idea is to model the entire operating cycle of a load and make identification based on event-window candidates. The proposed technique makes NILM more applicable for complex loads, more robust for load inventory change and can also simplify the training process; on the other hand, the thesis addresses a new and critical problem that previous researchers ignored---the non-intrusive extraction of load signatures. The proposed approach is an unsupervised non-intrusive approach which can automatically extract load signatures by using the meter-side data and requires almost zero effort from users. This thesis also discusses how to estimate the energy for several key components in a residential house such as the NILM identified appliances, load groups and background power. Based on estimation, residential energy characteristics are discussed with respect to the Time-of-use price.

Acknowledgement

I express my sincere appreciation to Dr. Wilsun Xu. This research and dissertation would not have been possible without his patient guidance and great supervision. I will always be proud to say I had the privilege of being his student.

I also thank my parents, Liping Guo and Ziqing Dong, for supporting me in so many different ways during my past 27 years. My parents turned my desire to do my graduate studies at a great Canadian university into a reality.

Finally, I also thank my workmates Charles Shi, Pengfei Gao, Shane Long, Kirash Shaloudegi and Mostafa Tabatabaei, who helped me in both working and living.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This chapter clarifies several basic questions regarding the thesis's research subject: decomposition techniques for power system load analysis. Firstly, the background and importance of this research are introduced. Secondly, the research problem is properly defined. Thirdly, three existing research directions that may solve the defined problem are individually reviewed by investigating the literature details. Finally, this chapter discusses the scope of the thesis and presents the thesis outline.

1.1 Background

The increased public awareness of energy conservation in recent years has created a huge interest in energy consumption monitoring at the end-user side of power system. Conventionally, meters installed in the downstream of power system can obtain only the aggregated compound load information. However, according to a recent market research report [1], end-users have become interested in tools that can help them understand and manage the details of energy use and its expense. This trend is especially obvious for the residential end-users [1]-[9] since commercial and industry end-users may already have advanced energy auditing tools and protocols.

For residential customers, a critical link to address the above need is smart meter. According to [2], smart meter is defined as an electric meter that records the consumption of electric energy in very short intervals such as an hour or less and communicates that information at least daily back to the utility for monitoring and billing purposes. The above two main features of smart meter enable a deeper and clearer view of home electricity usage by the end-users. However, the smart meters currently available in the market can provide only the compound load information or the information about an entire residential house. They cannot tell which appliance loads in the household consume the most energy or are least efficient. If households can also understand their usage patterns of concrete appliances, the following benefits can be achieved:

Recently, instead of using constant retail price for electricity, substantial variable electricity rates have been proposed and started to be used in some areas of North America. The reason to adopt such pricing structures is that they can more closely reflect the actual cost of electricity at a given time or period, which may potentially lead to user's adjustment of electricity usage accordingly. One example is the hourly Real-Time Pricing (RTP) used in Illinois [12]-[13]. In Canada, a lower-resolution variable price rate---Time of Use (TOU) price has been widely used in the province of Ontario [14]-[19]. TOU is the electricity price that is pre-set for a specific time period on an advance or forward basis [15]. In Ontario, instead of varying the price hourly, only three TOU prices are applied to three corresponding periods in a day. Overall speaking, to take full advantage of the above variable rates, householders need to be informed of the schedule patterns of individual appliance activities. Therefore, the compound load information needs to be decomposed into the sub-load level.

Besides, even with flat electric pricing, breaking compound load down to individual component level can still help customers control and save their energy usage better. For example, customers can compare the efficiency of a certain load with his neighbors in the same geographic area. Also, they will naturally pay more attention to heavy power consumers after identifying them. A preliminary study from a pilot program between IBM and the City of Dubuque, Iowa has indicated strong engagement by residents and energy savings of up to 11% by making comparison among residential profiles [20].

For the utility side, understanding the decomposed and detailed usage patterns can be very helpful for load forecast, load settlement and other load studies [10]. Also, it can be used for electricity rate design and residential demand response management planning [10],[91]. In addition, potentially reducing power demand in peak hours can lead to a cost reduction for utility companies [11] .

1.2 Problem definition

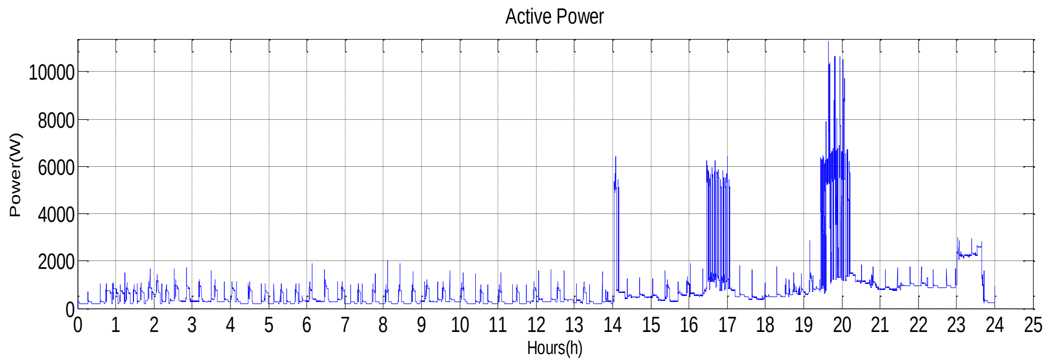


Figure 1.1: Example of compound load power

In Figure 1.1, the compound load power of a residential house throughout a day is shown. The compound power is acquired from the meter-side, and as can be seen, all different appliances' powers aggregate together. This is because for any circuit, power is physically additive regardless of the individual component being resistive, inductive or capacitive [98].The problem to be solved can be stated as follows: individual loads' signals aggregate at the entry point of a house as $P(t)$ and the intent of this research is to do the reverse or to decode the overall signal into various components $P_i(t)$ that are attributed to specific loads (appliances) *i*.

$$P(t) = P_1(t) + P_2(t) + \ldots. + P_n(t) \tag{1.1}$$

Assuming each load has its specific operation state, the above problem is equivalent to the following equiation:

$$\begin{aligned} P(t) &= P_1(t) + P_2(t) + \ldots. + P_n(t) \\ &= f_1(S_1(t), p_1) + f_2(S_2(t), p_2) + \ldots + f_n(S_n(t), p_n) \end{aligned} \tag{1.2}$$

where $S_n(t)$ is the operation state of individual load *n* and p_n is the power drawn from a certain state of load *n*. Equation (1.2) actually implies the problem of load decomposition can be converted to the identification of activities or states of individual loads based on the aggregated signal. For example, an incandescent light bulb has only two states: ON and OFF. For a given instant, once its state is known, the instant power the bulb is drawing can also be determined. This is a very important conclusion because many previous methods that will be reviewed in this chapter are actually based on equation (1.2).

In addition, when a time period is interested instead of a particular time point, the problem can be defined by using the following formula:

$$E(t_0,t_1)=E_1(t_0,t_1)+E_2(t_0,t_1)+\ldots+E_n(t_0,t_1) \tag{1.3}$$

where t_0 and t_1 are the beginning and ending time of a specific time period and E_n is the energy consumed by load or load group *n* during this period.

Equation (1.3) is especially suitable when dealing with certain load groups or long-term energy components such as the energy consumed by all the in-home lights in a day and the stand-by energy. In these cases, the instant power of the sub-load is not of interest whereas the energy consumed during this period as a whole is of more concern.

1.3 Review of statistic data based methods

This research direction aims to provide the estimate of sub-level loads based on statistical samples and certain given parameters of compound loads. The common procedure is as follows: As prior knowledge, statistical data such as house occupant behavior, demographic characteristic, proclivity characteristic and appliance power characteristics have already been collected from extensive surveys and measurements done by previous organizations and researchers [21]-[30]. Secondly, for a new compound load of interest, specific parameters such as the number and type of occupants and building structure information are also given as calibration inputs. Then, the above two parts of information are both fed into a simulation module. Finally, its decomposed energy information can be estimated [31]-[35].

For example, in [21], a nationwide time use survey was conducted by ISTAT(Italian Central Institute of Statistics) covering a sample of 40000 persons between June 1988 and May 1989. Participants were asked to keep a diary listing all their activities for a specific day. The survey's questions concern about the starting and finishing times, brief description and place etc. for daily activities of a household. Based on the survey results and appliance penetration analysis, [33] generated different home load activity probability profiles such as housework loads (clothes-washer, dish-washer), cooking loads, leisure time loads (TV).

[23] is the UK 2000 Time Use Survey which collected 20,981 1-day diaries recorded at a 10-min resolution. 19,898 of them were considered qualified for statistical analysis. This survey includes detailed questions regarding how people spend their time at home. For example, for home cooking category, the related questions involve "unspecified food management", "food preparation", "baking", "dishwashing", "preserving" and "other unspecified food management". Based on this survey, researchers constructed high-resolution daily activity profiles [31] and these profiles for different appliances can be downloaded online [24]. One example is shown in Figure 1.2.

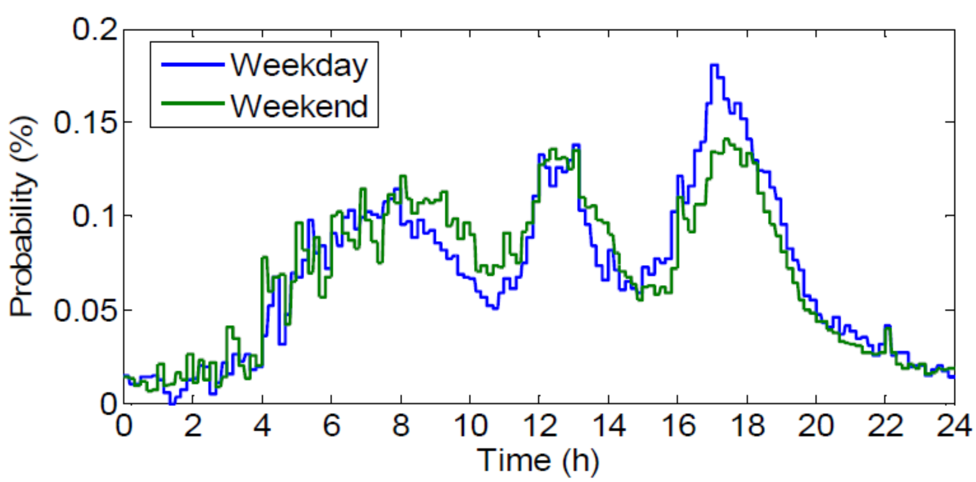


Figure 1.2: Time use probability profile for stove

[26] discusses the derived profiles of electric loads in Canadian houses. It adopts the statistical study results obtained by Pacific Northwest National Laboratory in 1989 [27]. Energy use profiles of air infiltration, ventilation, lighting equipment, appliances and miscellaneous electric loads were obtained from this study in which measurements in some benchmark houses were conducted.

Besides statistical time use profiles of loads, other statistical data such as appliance duration characteristics can be found or derived from Canadian Center for Housing Technology [25] and standard appliance test methods of the Canadian Standards Association [28]-[29] .Appliance power characteristics can be obtained from available measurement data [30],[102].

After statistical data are obtained as inputs, concrete compound load inputs are also provided to “calibrate” the statistical models. For example, in [31]-[33], inputs such as the number and type of household members, availability of household member during a day, appliance ownership are taken into consideration.

There are already a few commercial programs based on the above procedure. For example, DOE-2 is a widely used building energy simulation analysis program [34]. Its flowchart is illustrated as Figure 1.3. Its library is a huge database which stores information of specific sub-loads, compound loads and relevant statistical parameters such as regions, weathers, building structure, occupant behavior and even materials of enclosures. With user input, specific building descriptions of compound load can be generated. Along with statistical weather data, complicated and dynamic simulations can be performed. This program is especially useful for energy estimation of heating, ventilating, and air-conditioning (HVAC) equipment. Finally, the output of different sub-loads and other detailed energy predictions can be obtained.

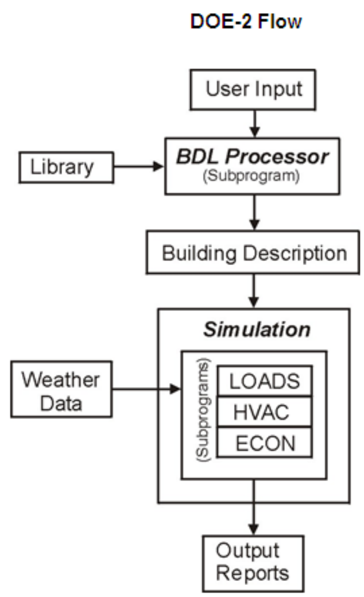


Figure 1.3: Flowchart of DOE-2

Similarly, [35] is a recently developed web-based service aiming to provide detailed consumption information for ordinary residential houses in the US. To accurately profile the house, users are inquired with numerous questions related to their house structures, insulations, family compositions and major appliances. Examples are questions about the roof, ceiling, siding, floor materials, foundations of houses, types of heating and cooling system, typical hours of use of stove, oven and dishwasher and so on. In total, more than 100 questions have to be answered by an ordinary householder. In the end, energy consumption estimates on different sectors such as water heating, major appliances, other appliances and lighting are generated.

From the above examples, the advantages and disadvantages of the method based on survey data can be seen:

- The estimate is based only on statistical data and simulation. No real consumption information about the compound load is measured and taken into account. Conceptually, the error between simulation result and the real data is likely to be fairly large and is difficult to be interpreted and reduced.

- Values collected from survey data cannot represent individual samples accurately. Variations of individual samples such as uncommon types of appliances, uncommon materials used for doors and windows, and different indoor preferences such as temperatures cannot be avoided. Sometimes, the variations can be huge.

- The prior parameters of compound load are required and have to be inputted to the system. However, it is very difficult for ordinary users to collect such parameters. Sometimes, the collecting of some parameters even requires special expertise.

- The advantage is that this solution is cheap: no additional hardware installation is needed for this method.

1.4 Review of measurement based methods

The measurement based methods require additional sensing devices connected to individual sub loads of interest. Generally, the two different streams are direct sensing, which senses the current drawn by the load and indirect sensing, which senses the other physical quantities related to the acoustic, light or magnetic field.

In residential houses, three methods are usually used to implement direct measurement: "Smart plugs" [36]-[37] are devices connected between appliances and electricity outlets. They have measurement circuits inside which can acquire appliance's currents flowing via and the voltage from the outlet. Thus, smart plugs can measure the power of connected appliances in real-time. The data from different appliances can be collected either through wireless or power-line based communication; some devices [39]-[40] are installed inside house electricity panels. Pocket-size current sensors are connected to individual circuit branches in panels. These branches are either wired to specific appliances such as a dryer or to specific rooms such as a kitchen; "smart appliances"[41]-[43] are appliances that have their own measurement circuits inside and are also able to communicate with a smart meter or other devices via a Home Automation Network [44]-[48].

However, such appliances are not commonly seen in the market yet at present. In addition, the direct measurement based method can also be applied to specific loads such as HVAC and VFD in commercial and industry buildings.

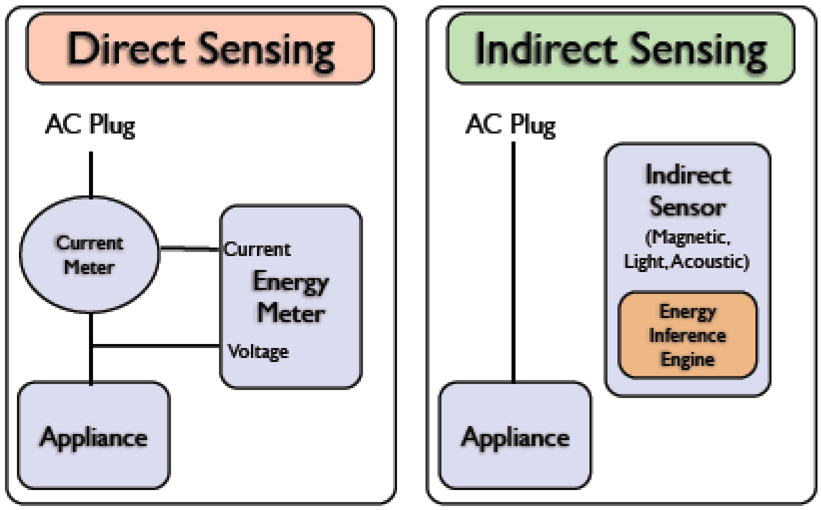


Figure 1.4: Comparison of direct and indirect sensing from [49]

Sometimes direct current sensing is not available. For example, the wiring of residential lights is often through the wall and current meter cannot be connected without cutting out the wire. Thus, [49] proposed and developed the “indirect” sensors (acoustic, light and magnetic sensors) and put them near the appliances to estimate their power consumptions. No in-line sensor such as current-plug is needed. Also, each sensor has a calibration process so that the sensors can establish a specific relationship between the measured physical quantities and the power. Direct and indirect sensing are compared shown in Figure 1.4.

[50]-[51] are also “indirect” sensors but they measure only the state transitions or ON/OFF changes of appliances. To estimate the energy, the state recordings need to be associated with the meter-side power data. The working principle can be explained by using formula (1.2). [3] uses a clamp based magnetic sensor ,which can be clamped around the supply cable of an appliance. Although the summation of the cable’s flow-in and flow-out currents is equal to zero, after signal amplification, a minor magnetic difference can still be observed. In [51],

the concept “binary sensor” is proposed and it can be used to detect ON/OFF states of appliances.

Overall, the characteristics of measurement based methods can be summarized as below:

- They require additional hardware measurement devices.
- The cost of a system can be very high in terms of installation and maintenance.
- In some cases, a communication system is also necessary. Moreover, the customer acceptance of Home Automation Networks is still low, in spite of the government and media efforts.
- The accuracy of the consumption data collected through measurement is much higher than that of the statistic data based method, especially when using direct sensing devices. This is because the consumption of individual load is directly measured.

1.5 Review of non-intrusive load monitoring based methods

The research on non-intrusive load monitoring (NILM) originates from MIT [53]: The goal of the work was to develop a method for power companies to study the residential load characteristics without having to enter the residences.

Unlike the measurement based method, a NILM system uses only the information acquired from the main breaker level or meter-side. This system is a viable alternative to the Home Automation Network.

Recently, with the fast development and vast deployment of smart meters, more research attention has been drawn into this area. Smart meter’s high-resolution data acquisition capability, communication capability along with its computation capability [52]-[54] can provide sufficient support for the implementation of NILM in residential houses. On the other hand, NILM can add great intelligence and value to the smart meters and make smart meters a truly "smart" solution for residential energy management.

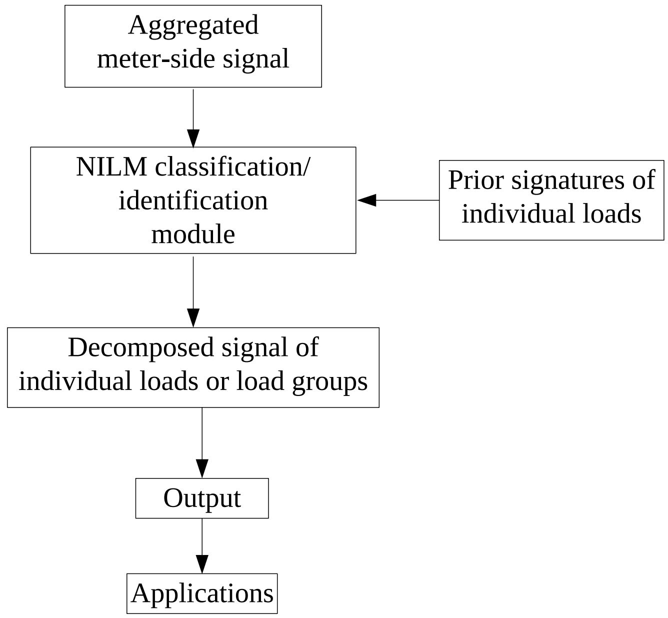


Figure 1.5: Typical NILM procedure

A typical NILM method and system comprise the steps as shown in Figure 1.5 [55]-[79]. The aggregated meter-side signal is acquired from meter-side through either smart meters or additional data acquisition devices. Also, specific appliance features or signatures are collected and mathematically characterized [94]. Afterwards, both the aggregated signal and signatures are fed into the core step: NILM classification/identification module. In this step, the aggregated signal are classified or identified, and the signal is decomposed into the individual load or load group level. Finally, the results are formatted as output, and different applications such as energy estimation, demand response, and condition monitoring can be implemented.

Generally, all the NILM studies can be divided into the categories shown in Figure 1.6. Detailed literature reviews will be presented in the overview sections of relevant chapters. Here, a rough division is given:

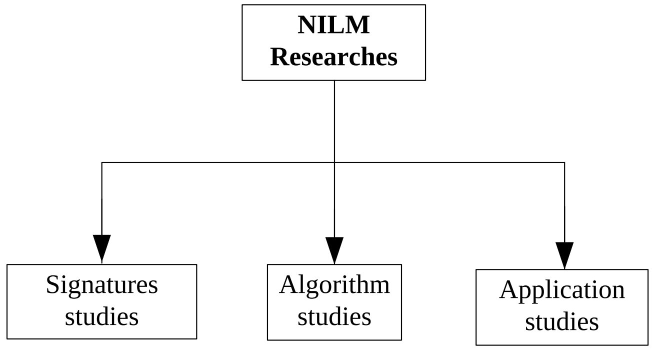


Figure 1.6: Chart of current NILM researches

- Signatures studies. In the time domain, the studies include steady-state signatures [55]-[60] and transient signatures [61]-[66]. The other studies also involve signatures after particular transforms such as Fourier transform [59]-[60],[69] and wavelet transform [68]. These studies will be reviewed further in Chapter 3.

- Algorithm studies. According to [94], two main algorithm approaches are studied: the signal-combination based method [59]-[60],[69]-[79] and the event based method [55]-[58],[61]-[66]. The algorithm studies along with the signatures studies have been the main focuses in NILM research area. It should be noted that the two are often tied together since sometimes different signatures will result in different algorithms. These studies will be reviewed further in Chapter 4.

- Application studies. [84]-[86] discusses how to use NILM to estimate the power consumption of variable-speed drives. [87]-[90] presents on how to use NILM for condition monitoring and fault diagnostics. In [91]-[92], a load-shedding strategy is proposed based on NILM. However, traditional application such as energy estimation for common residential energy components has not been fully explained before. This issue will be discussed further in Chapter 6.

Compared with the survey based method and the measurement based method, the characteristics of NILM based method are as follows:

- Low cost: no additional hardware needed except for the aggregated meter-side signal acquisition device. In many cases, the meter-side acquisition can be achieved by using an existing smart meter.

- High accuracy: the accuracy may be lower than that from the direct measurement based methods but is much higher than that from the survey based method and indirect measurement based methods.

Overall, the NILM based method provides good balance between cost and accuracy and is therefore the most promising load decomposition technique so far.

1.6 Thesis scope and outline

The purpose of this thesis research is to solve the remaining but critical challenges and limitations related to the existing NILM based methods:

1. Unable to deal with complex loads effectively. Complex loads such as continuous-varying loads and multi-state loads have not been sufficiently researched. However, in fact, they are an important portion of residential loads and need to be addressed.

2. Time-consuming training process. Many NILM methods require a time-consuming training/learning process to establish the map between specific activated appliances and their aggregated signal. Moreover, after the load inventory is changed, training process has to be redone. This requirement is a critical obstacle that prevents NILM from being applied to the ordinary households.

3. Lack of research on signature extraction. This is an important research area that has been neglected by previous NILM researchers. The existing measurement based signature extraction methods are actually intrusive. This is another critical obstacle that prevents NILM from being practically applied.

4. Insufficient research on event detection. The previous event detection methods are very simple and can lead to detection error in some cases.

5. Insufficient research on energy estimation methods. How to estimate energy based on NILM results and how to estimate energy for specific energy components such as load groups and background power have not been addressed before.

The thesis is organized to present different studies to tackle the above listed problems. The outline is below:

- Chapter 1: Clarifies the basic questions of this work.
- Chapter 2: Presents a study of event detection and tackles problem 4. The study consists of a review of the existing methods, explanations of two new sophisticated methods, comparative studies based on real field data and discussions on the special issues of event detection.
- Chapter 3: Presents a study of event-window based load signatures and prepares for tackling problems 1&2. The study consists of a review of the existing load signatures, explanations of the proposed load model and discussions on different event-window signatures.
- Chapter 4: Presents a study of a proposed NILM identification algorithm and tackles problems 1&2. The study consists of a review of the existing algorithms, explanations of the proposed algorithm, a discussion on the system implementation and thorough verification and comparative studies.
- Chapter 5: Presents a study of non-intrusive signature extraction and tackles problem 3. The study consists of a review of intrusive signature extraction methods, explanations of the proposed algorithm and thorough verification studies.

- Chapter 6: Presents a study of energy estimation for residential house and tackles problem 5. The study consists of explanations of estimate methods for different energy components, an interpretation of the residential energy characteristics and their implications to the Time-of-Use price.

- Chapter 7: Presents the main conclusions from this work, and suggestions for future studies and improvements.

- Appendix: Presents a preliminary study of a load decomposition technique specified for a commercial compound load. The appendix also addresses problem 5.

Chapter 2

Event detection

A load event is defined as the transition of a load's state. Event detection is an essential pre-processing step for most of the methods and algorithms proposed in this thesis. The quality of event detection has a direct impact on the final results of event-window based load decomposition and signature extraction.

This chapter presents elaborate discussions on event detection issues. It first reviews the existing event detection method and identifies the potential challenges. To deal with these challenges, this chapter proposes two data-segmentation based methods---segment-conjunction method and slope method. The basic idea is that instead of directly seeking for status-transitions, one can find out all the continuous data segments and then the portions between two neighboring segments can be considered as the events. The two proposed methods are compared according to the tests on actual field data.

In addition, this chapter also addresses the special problems of event detection---double-phase event detection, adjacent event handling and event overlap. Some contents in this chapter have been submitted as publication [103].

2.1 Overview

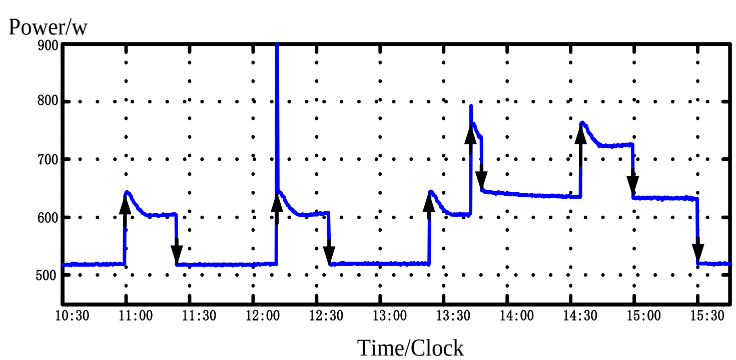


Figure 2.1: Real-time power data acquired from meter-side

According to [55], a load event is defined as the transition of a load's operation state. As can be seen from Figure 2.1, although aggregated signal acquired from meter-side contains the summation of individual signals of all currently activated appliances, an event which is relevant to a certain appliance' operation state change such as its ON/OFF usually associates with a power jump and thus can be observed as an independent "edge" from meter-side signal. To capture such edges, event detection methods are needed.

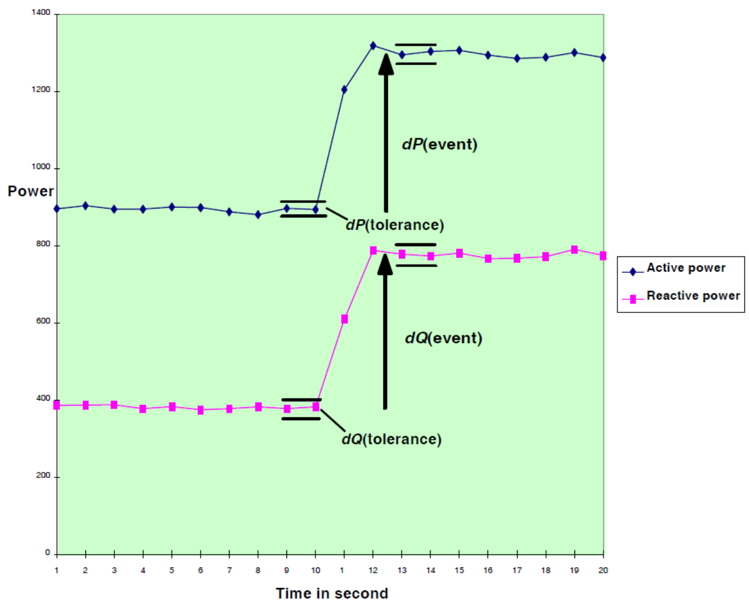
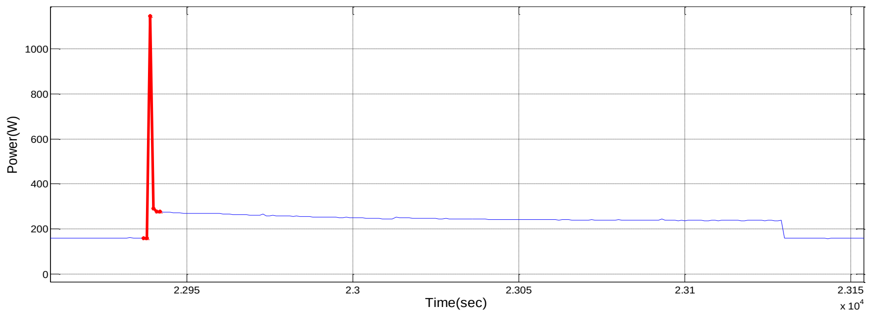


Figure 2.2: Detecting an event in sample data from [58]

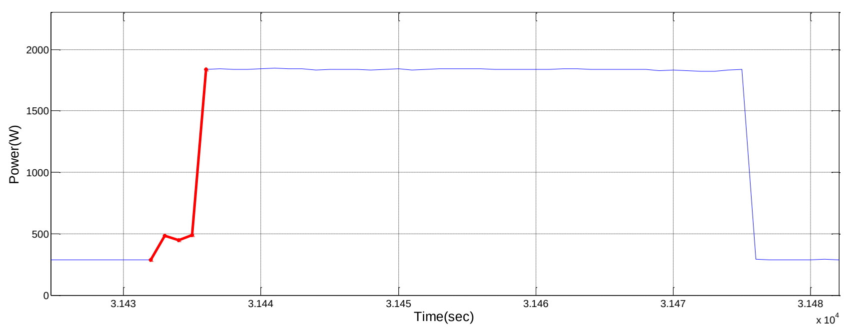
The previous researches on event dection are limited and most of them rely on a certain threshold based step-change detection [55]-[59]. In [57], a revoving memory keeps the last measured successive real power values P1 and P2. δ is the detection threshold and whether $P2-P1<\delta$ is continuously tested. Once $P2-P1>\delta$ is found, the corresponding step change is captured as an event. Based on interested load power level, different thresholds can be adopted. For example, in [59], 100W is used as the threshold. In [58], a similar method is proposed with extra points considered to judge if a steady period starts right after a potential event. Also, both real power and reactive power are considered together for determination. A period of change is detected if the site-specific thresholds dP_{tres} and dQ_{tres} are exceeded while the tolerances dP_{tol} and dQ_{tol} are not exceeded according to the following equation (P_i, P_{i+1} and Q_i, Q_{i+1} are successive samples). The process is shown in Figure 2.2.

$$\begin{cases} |(P_i + P_{i+1})/2 - P_{i+3}| > dP_{tres} \\ |(Q_i + Q_{i+1})/2 - Q_{i+3}| > dQ_{tres} \\ |P_{i+4} - P_{i+3}| < dP_{tol} \\ |Q_{i+4} - Q_{i+3}| < dQ_{tol} \end{cases} \tag{2.1}$$

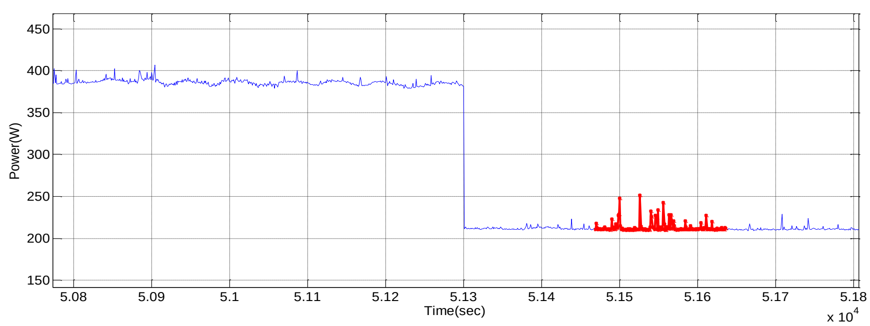
The above step-change methods can deal with simple events. However, there are some remaining challenges that need to be coped with.



(a) Spike-type event



(b) Slow event



(c) False event (signal noise)

Figure 2.3: Examples of remaining challenges for event detection

As can be seen from Figure 2.3, (a) is a spike-type event that can be often seen when a motor device is started. This event is composed of a sharp positive step change and a smaller negative step change that follows right after the positive change. The traditional step-change based methods may consider this event as two independent ON/OFF events; (b) shows a slow ON event from a microwave. However, it actually has two small step changes followed by a big step change. The traditional method such as [58] will discard this event because $|P_{i+4} - P_{i+3}| > dP_{tol}$; (c) shows the mis-detection case---these quick step changes are actually caused by signal noises and should not be considered as events.

To deal with the above challenges, this chapter proposes two novel data-segmentation based methods. As shown in Figure 2.4, the new philosophy behind is that instead of looking for step change like the previous methods do, the proposed methods look for all the continuous data segments and then the data portions between two neighboring segments are considered as the events no matter how complicated they are. Besides, the continuous data segment should be capable to include limited abnormal points as long as the continuity of the data segment is not significantly broken by them.

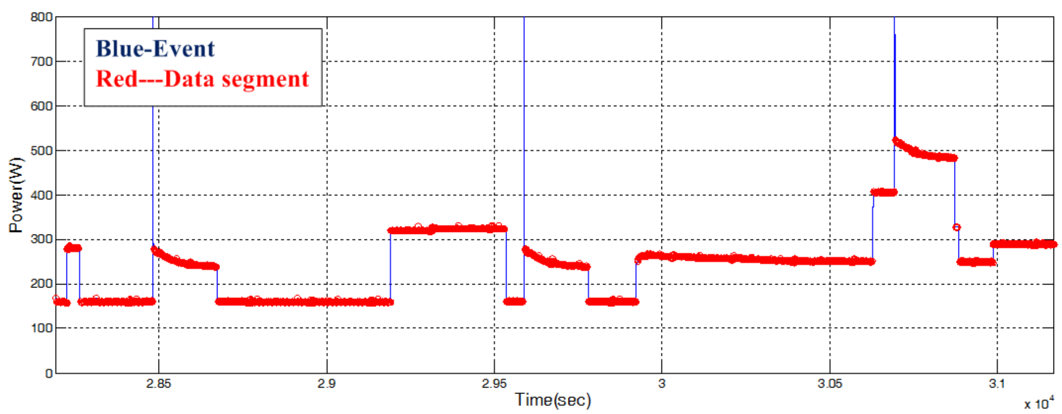


Figure 2.4: Example of data segmentation based event detection

In addition, other special issues about event detection such as double-phase event detection, adjacent event handling and event overlap will also be discussed in the end of this chapter.

2.2 Data segmentation based methods

As shown in Figure 2.4, data segment is defined as a segment of which the continuity is good. It should include three scenarios: 1) a steady segment that has a constant power level and does not change over time. It can be seen as a straight line; 2) a continuous varying segment that has inconstant but slowly varying power level over time. It can be seen as a curve; 3) a segment with acceptable noises like the one shown in Figure 2.3 (c).

Two methods that can detect the above defined data segments are proposed. The data portions between two neighboring segments are considered as events. They are explained as follows.

2.2.1 Segment-conjunction method

As shown in the flowchart below, the segment-conjunction method has 4 steps.

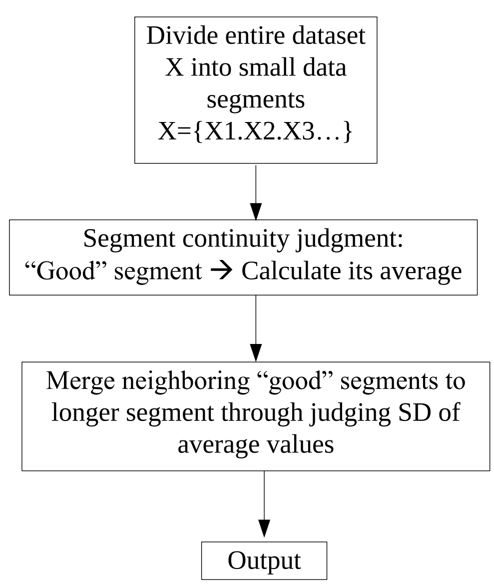


Figure 2.5: Flowchart of segment-conjunction method

Step 1: Divide the entire dataset X into small continuous data segments {X1, X2, X3…} with a preset minimal data length such as 5 points per segment. Since the data is acquired every 1 second, 5 points represents 5 seconds.

Thus, we have X= {X1, X2, X3…..} X: aggregate of magnitudes. X1, X2, X3… follows the original sequence in the dataset.

Step 2: Judge each segment to see whether it is a "good" segment with acceptable variation.

If the variation among each segment is beyond a certain threshold, it is marked as a "bad" segment; if the variation is below the threshold, it is marked as "good" and its average value is calculated. Standard deviation (SD) is used as an index for variation evaluation.

Examples of good and bad segments are shown in Figure 2.6. Each circle represents a data point. Compared with (a), in (b) the variation of five points is too big or in other words, its SD exceeds the predefined threshold, hence this segment is not considered acceptable.

Figure 2.6: Example of good and bad segment

After marking all these segments, calculate the average value of each "good" segment. E.g. If X3, X4, are good segments, Ave3=Average(X3), Ave4=Average(X3)…

Step 3: Try to merge neighboring "good" segments: Calculate the standard deviation (SD) of the average values of neighboring segments. If it is less than a threshold we set, like 2.0, join them together as longer segment.

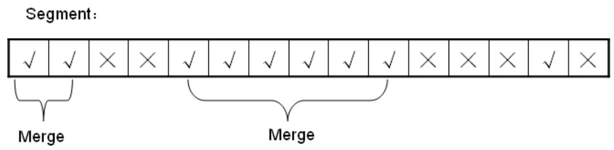


Figure 2.7: Merge good segments together based on SD and range calculation

For example, assume X3,X4,X5,X6, X7 segments are "good" segments and they are continuous. If SD (Ave3, Ave4, Ave5) <2.0 then {X3,X4,X5}=LX1. LX1 is the merged longer segment. Similarly, if SD(Ave3, Ave4, Ave5, Ave6)>2.0, X6 is excluded from LX1 and LX1 is finalized as a whole segment that includes X3,X4 and X5.

Step4: Output. Label data segments and events in different colors

2.2.2 Slope method

Slope algorithm focuses on a group of data points' slopes, which is considered as an effective index to judge the continuity of data segment.

As is known to us, slope dx/dt can be used to evaluate the velocity of variation. Points in the segments that have good continuity should have small slopes. In reality, the approximation below is used:

$$\frac{dx}{dt} = \frac{\Delta x}{\Delta t} \tag{2.2}$$

Here, Δx is the difference between two successive acquired data points. Δt is the time difference between two sampling points, which is actually the acquisition interval. Since Δt is fixed, only Δx needs to be taken into account. In other words, Δx of each point can represent the point's slope or velocity of variation.

Figure 2.8 below shows the slope Δx_n for a specific data period. Here $\Delta x_n = x_{n+1} - x_n$ .Power curve is shown in dash line and the slope line below it shows the corresponding the slope values of each point.

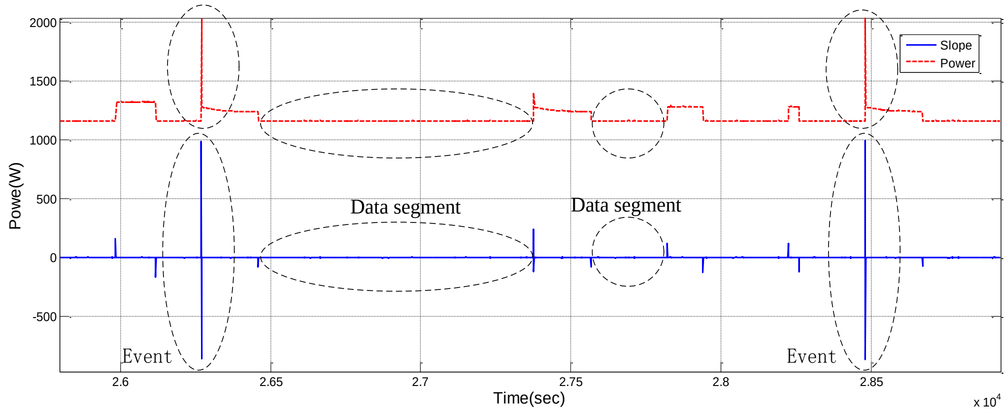


Figure 2.8: Slopes of a data period

From this figure, it can be seen that many slope values are close to zero, which indicate small variations. Some points, however, are very large, which can reach up to 50 or even 1000. Those points express large variations, which could be an event or part of an event. The basic idea is to join continuous points that all have slope values close to zero as a data segment.

The flow chart of the proposed slope algorithm is shown in Figure 2.9.

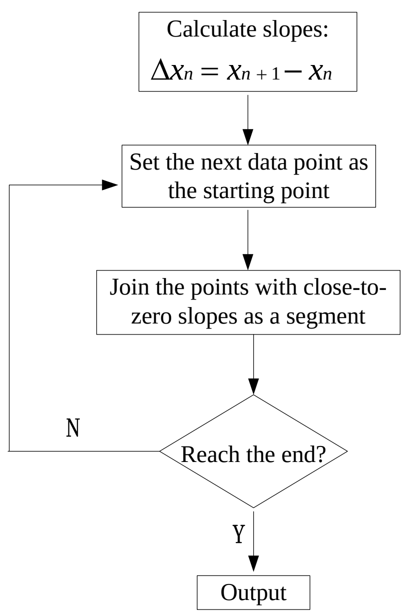


Figure 2.9: Flowchart of slope algorithm

Step 1: Subtract each data point from the next data point following it. $\Delta x_n = x_{n+1} - x_n$. In other words, the slope values of all data points are calculated.

Step 2: From a starting location, try to join the following points with small slopes. A special concern here is the occasional noisy points shown in Figure 2.10.

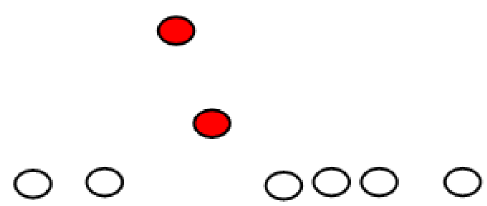


Figure 2.10: Noisy points in a data segment.

For this situation, since slopes of noisy points are big, they will not be automatically connected with the points in front. However, in reality, one long segment with few noisy points included is still acceptable. In order to prevent cutting off a segment too early due to the noisy points, the method will make an additional check on the steady points behind the noisy peak at the same time, to see if they are close to the values in front of the short peak. If they are close, these noisy points are "bypassed" and the connection with the following points will continue.

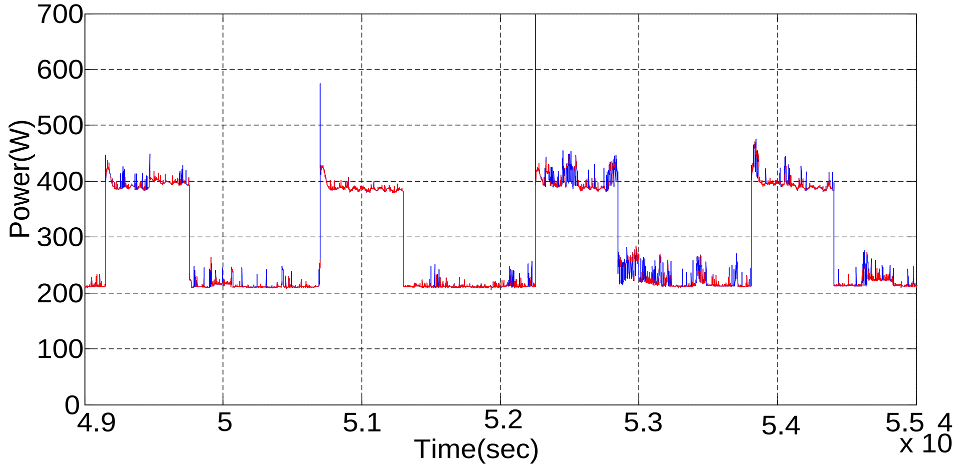
The connection stops when both of the following two conditions are met:

- The slope value of the point is beyond the threshold;
- The difference of the average of the steady points after this point and before the point is also beyond the threshold

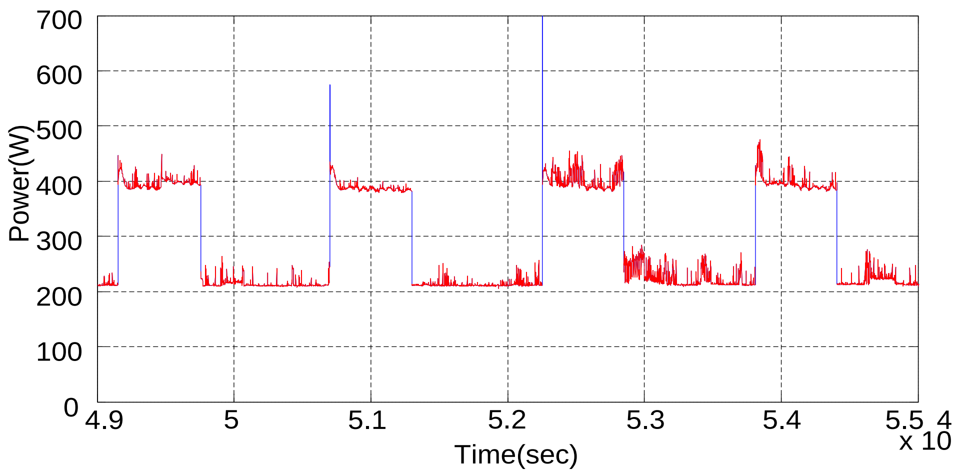
Step 3: After the interruption point, start the processing of a new segment from the first point that has a close-to-zero slope value after the interruption point. Redo step 2. The whole iteration ends until all points of the dataset have been processed.

Step 4: Output.

2.2.3 Comparison of the two proposed methods



(a) Segment-conjunction method

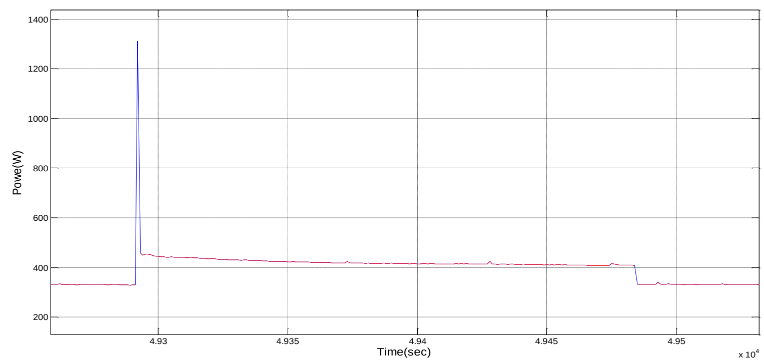


(b) Slope method

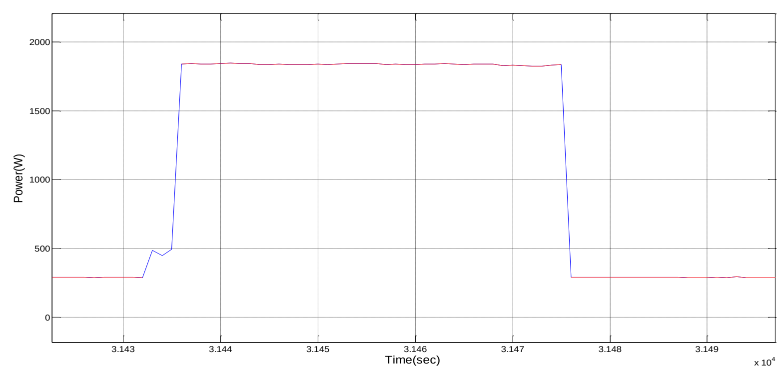
Figure 2.11: Comparison of the two methods on dealing with noises

First of all, the performance of the two proposed methods on noisy point handling is compared. Figure 2.12 shows four consecutive fridge's operations. They are all polluted by signal noises from the meter-side. The two proposed methods are applied to the four cycles respectively. As can be seen from (a), the data segments (in red color) captured by using the segment-conjunction method are interrupted many times by noises; while in (b), there are only 8 big data segments that represent the ON and OFF states of fridge. Accordingly, 8 events are captured (in blue color) between neighboring data segments. Hence, the slope method seems much more powerful when dealing with noisy segments.

Referring to the other challenges mentioned in 2.1, both of the two methods can effectively deal with the spike-type event and slow event that contains more data snapshots. Examples are shown in Figure 2.12.



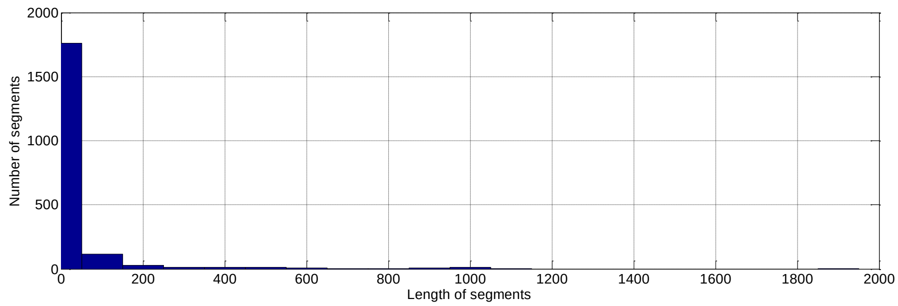
(a) Spike-type event



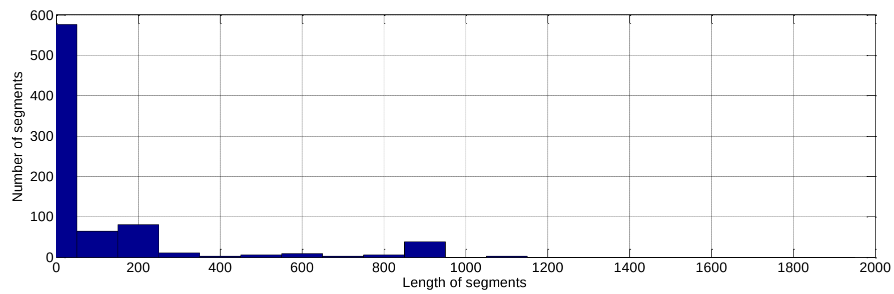
(b) Slow event

Figure 2.12: The two methods on dealing with spike-type event and slow event

The two methods have also been applied to the data acquired from a single hot phase of a local residential house on a typical day. The histograms of captured data segments using the two proposed methods are shown in Figure 2.13. As can be seen, the number of segments captured from the slope method is smaller but the average length of data segment is longer. Again, it proves that the slope method can catch longer segments containing noises. Accordingly, the number of events captured by the slope method will also be smaller but much more reasonable.



(a) Segment-conjunction method



(b) Slope method

Figure 2.13: Histograms of captured data segments using the two proposed methods

2.3 Special issues with event detection

2.3.1 Detect double-phase events

As shown in Figure 2.14, some events happen simultaneously on both of the two hot phases in a residential house. It is because some appliances are connected between the two opposite hot phases to gain a higher voltage. Therefore, there is a need to recognize these double-phase events and label them. The following formula can be used to effectively make judgment on double-phase events. At a given instant when simultaneous events are found at Phase A and B, if

$$\left|dP_A / dP_B\right| > Thres_{\min} \;\; and \;\; \left|dP_A / dP_B\right| < Thres_{\max} \tag{2.3}$$

then the event is labeled as a double-phase events. The power differentials between dP_A and dP_B should be very close since they have almost identical currents. Typical values of $Thres_{\min}$ and $Thres_{\max}$ are 0.9 and 1.1.

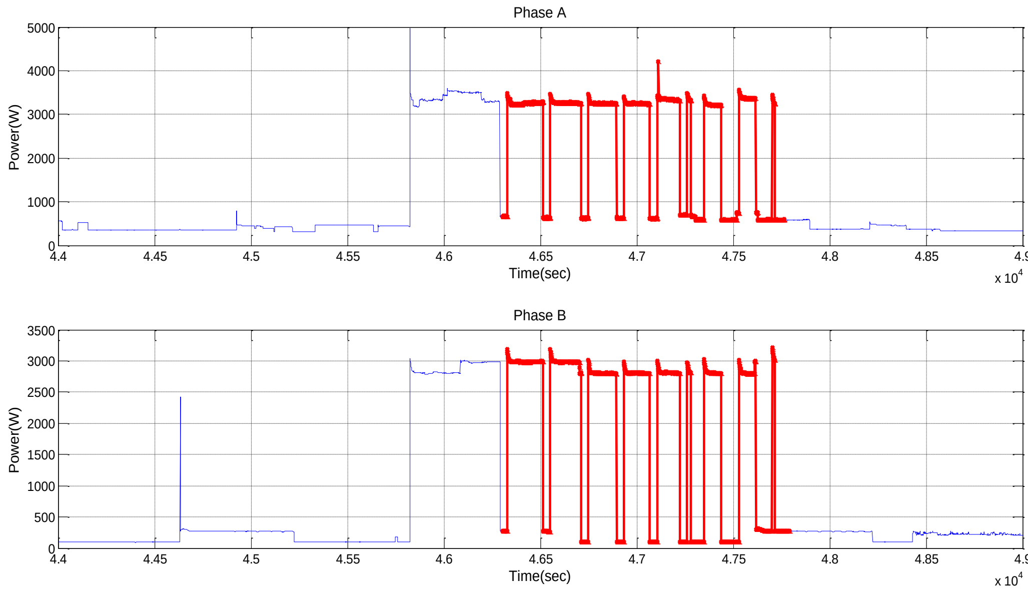


Figure 2.14: Examples of double-phase events

2.3.2 Adjacent events

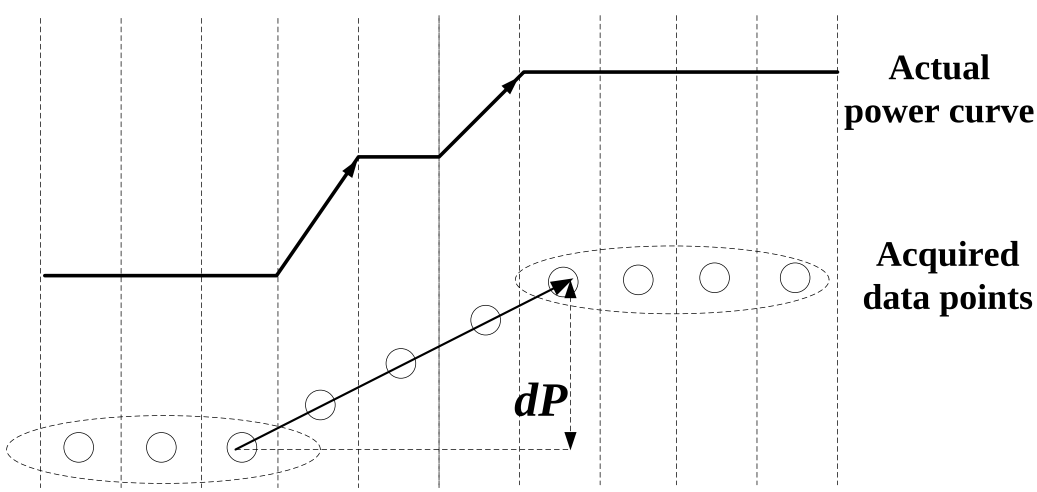


Figure 2.15: Example of adjacent events

Sometimes adjacent events may lead to event misdetection due to low sampling rate. As can be seen from Figure 2.15, the actual power curve has two adjacent events with a short interval in between. The acquired data points (circles) are shown under the actual curve as comparison. For each interval, one data point is acquired. Based on the proposed event detection methods, during this period, the first 3 points and the last 4 points will be considered as two segments and the 3 points in between will be detected as a slow event. This obviously violates the truth. However, when the sampling rate is low this problem is evitable.

One solution is to increase the sampling rate. For example, doubling the sampling the rate will result in different results as shown in Figure 2.16. In this case, three segments can be captured and thus two independent events dP_1 and dP_2 can be found.

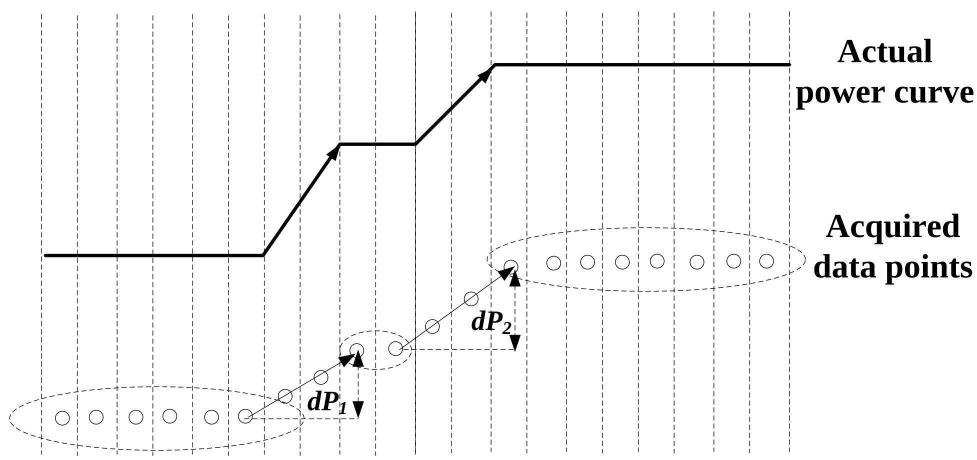


Figure 2.16: Result when the sampling rate is doubled

However, when the sampling rate cannot be increased due to hardware limitations, the suspected slow event that could comprise multiple events can be broken into pieces. This can often help solve the misdetection of events without adding cost on higher sampling rate. Still taking the period in Figure 2.15 as example, if an event is found too long (4 intervals in this example), a method is that one manually divide it in half such as the steps shown in Figure 2.17. With respect with Figure 2.15 and Figure 2.15, the two events dP_1 and dP_2 can still be obtained even with original data acquisition capability.

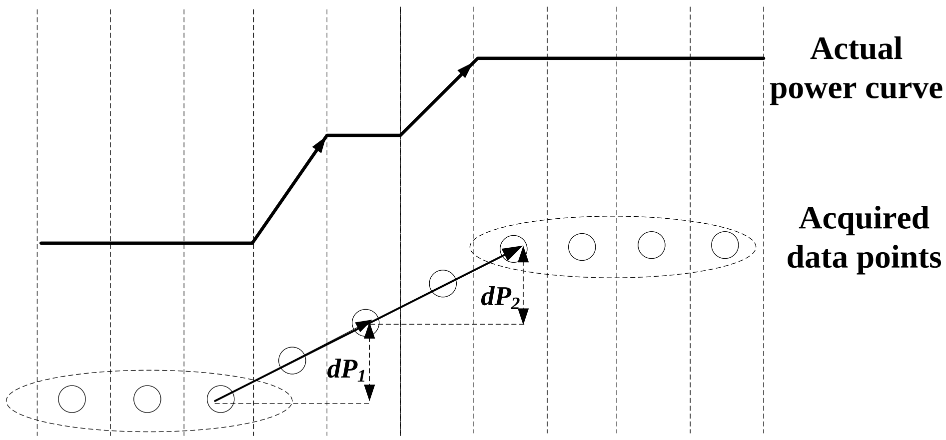


Figure 2.17: Solving adjacent event problems by separating a slow event

2.3.3 *Event overlap*

Also, in some cases, simultaneous occurrence of more than one appliance events can be encountered. This does not happen frequently but does exist. Theoretically, if two events fall in the same time interval, any event detection method will result a misdetection---only a single event can be captured. One example is shown in the Figure 2.18. The ON event in overlapping is composed of two single events belonging to the fridge and furnace. Its power magnitude is roughly the summation of the two.

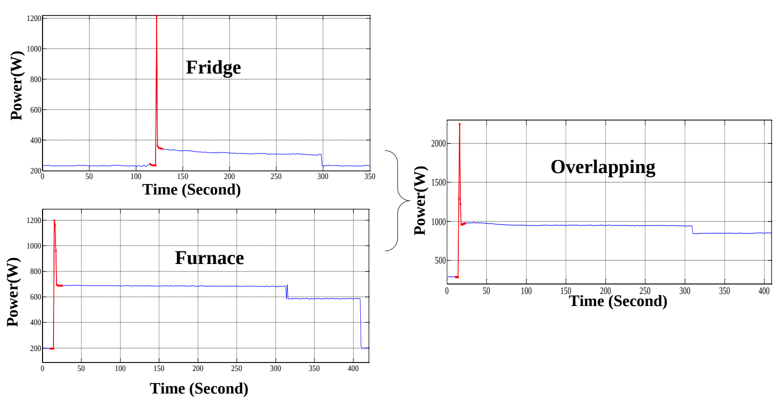


Figure 2.18: Example of event overlap

This problem is mathematically analyzed as a probability problem. Now assume a period of M seconds (intervals) is studied and N appliances that each show up K times on average are included in this period. Therefore, for each time, N different appliance events can overlap with each other. The total number of possible combinations for events occurring at different seconds is M^N and the number of non-overlapping combinations is $A_M^N \,.\, A_M^N$ can be calculated based on factorial. Assuming all the events are randomly distributed, the probability of overlap occurrence is:

$$\begin{cases} P = 1 - (\dfrac{A_M^N}{M^N})^K \\ A_M^N = \dfrac{M!}{(M-N)!} \end{cases} \tag{2.4}$$

In practice, for a specific hour, assume 3600 points are captured, 5 appliances are involved and each appliance shows up 3 times on average. According to (2.4), the probability of event-overlap occurrence in this hour is

$$P_h = 1 - (\frac{A_{3600}^5}{3600^5})^3 = 0.84\% \tag{2.5}$$

Based on P_h, for an entire day, the probability of event-overlap occurrence is

$$P_{day} = 1 - (1 - P_h)^{24} = 18.3\% \tag{2.6}$$

The above results may be higher or lower than the ones in reality because appliances events are not uniformly distributed along the time line. In some hours, different appliance operations are more expected to be seen while in some hours are not. Also the event frequency of some particular appliances such as stove and washer within certain hours could be extremely high.

It is also found that increasing sampling rate can also help solve the problem. For example, now assuming the sampling rate is increased to 0.5sec/snapshot, there will be 7200 snapshots in an hour. Therefore, we have:

$$\begin{aligned} P_h &= 1-\left(\frac{A_{7200}^5}{7200^5}\right)^3 = 0.42\% \\ P_{day} &= 1-(1-P_h)^{24} = 9.5\% \end{aligned} \tag{2.7}$$

The event overlap probability is decreased to approximately half of the original value.

2.4 Summary

To summarize, this chapter firstly reviewed the existing event-detection methods and identified several challenges that need to be solved. They are spike-type event, slow event and signal noises. To cope with these problems, two data segmentation based methods were proposed. Instead of looking for state transitions directly, the proposed methods look for stable data segments in which a certain level of noise is also acceptable. The first method bases on the conjunction of small segments while the second method studies the slope pattern of data points. Detailed comparison of the two methods was also presented and overall speaking, the slope method has a better performance, especially when dealing with noisy signals. In other words, it is more suitable for practical applications. It was also found both methods can effectively detect spike-type event and slow event.

In addition, specific common issues with event detection were also discussed: a simple method to detect double-phase event was proposed; misidentification due to adjacent events was also discussed. Both hardware and software solutions were proposed and explained; in the end, the problem of event overlap was also introduced and the method to calculate its probability was addressed. Based on a hypothetical calculation, it was also found higher sampling rate can effectively reduce the probability of overlap.

Chapter 3

Event-Window Load Model and Load Signatures

This chapter presents detailed discussions on a specific load modeling approach---event-window modeling and different types of load signatures associated with this model. With this model, loads are treated as a time window that consists of all the events such as turn ON/OFF events and other middle changes during its operation process. The signatures associated with the event-window are studied. These signatures include both electrical and non-electrical ones that can accurately and completely describe the operation process of any particular type of load.

The aim of this chapter is to establish a technical foundation for the proposed event-window based load identification method which will be further discussed afterwards. Some contents in this chapter have been submitted as publication [104].

3.1 Overview

3.1.1 Review of single-state load model and its signatures

Most of the existing non-intrusive load monitoring methods treat all the appliances as a single-state model which has a pair of identical ON/OFF edges and a constant power demand between them [55]-[60]. In other words, it assumes the operation of appliance has a constant steady point which does not change its electric status such as real power, reactive power and harmonic content dramatically. This assumption simplifies the load models but unfortunately, it cannot accurately reflex the reality--- many complex loads cannot be represented by this model and thus large errors could occur for the future identification steps.

Representative signature studies of single-state model can been seen in [59] and [60]. New single-state signatures such as “instantaneous power waveform” and “instantaneous admittance waveform” were discussed. Its assumption is still that each load only has a constant set of signatures that is not changing with time. Examples are shown in Figure 3.1.

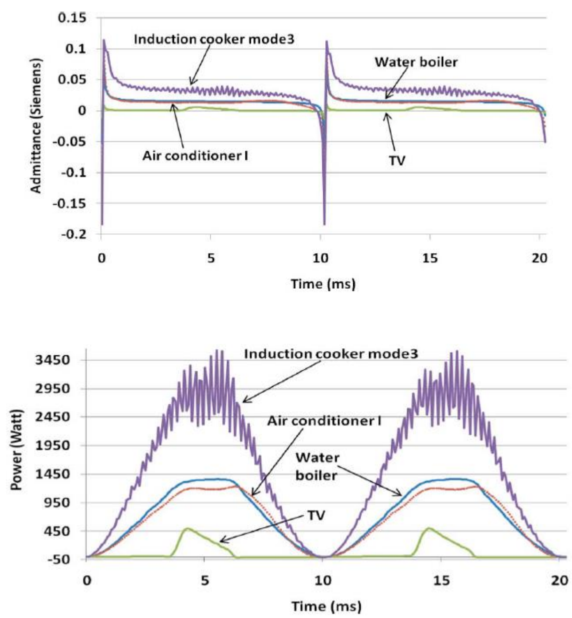


Figure 3.1: Example of single-state signatures from [59]

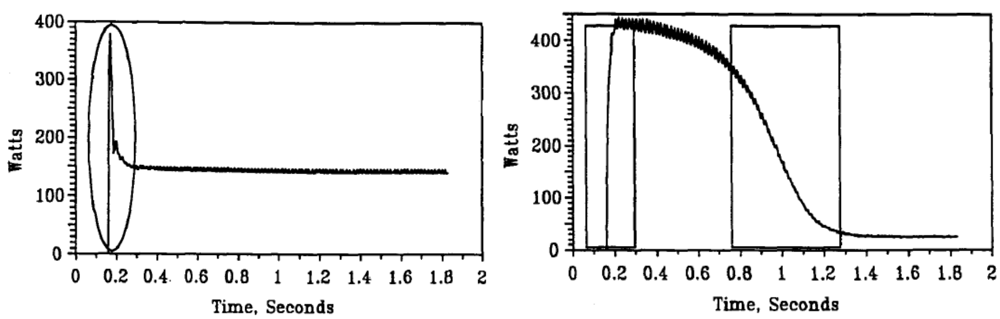
As can be seen, the waveforms of all loads look almost the same for the period 0-10 ms and 10-20ms. However, this is only true for the following scenarios:

1. When signatures are only compared in a short time scale such as within a second. In this case, 10 ms is only half cycle (50Hz system).

2. When the studied loads have only single operation state. In Figure 2-1, only water boiler can fall in this category. Air conditioner, TV and induction cooker are all complex loads which have multiple or continuous varying states. They should not to be treated as single-state appliances.

3.1.2 Review of transient load model and its signatures

Some other researches [61]-[66] focus on the transient signatures of loads at the moments when the studied load is turned ON. It is believed that different loads create consistent observable turn-on transient profiles suitable for identifying specific load classes. The turn-on transients associated with a fluorescent lamp and an induction motor, for example, are distinct because the physical tasks of igniting an illuminating arc and accelerating a motor are fundamentally different. Examples are shown as below:



(a) Florescent Lamp (b) Measured induction motor

Figure 3.2: Examples of real power turn-on transients from [62]

The drawbacks of modeling a load using only its turn-on transients are listed as below:

1. Some loads do not have significant turn-on transients such as resistive loads. The loads with observable turn-on transients are limited to florescent lights, motors and some electronic appliances. Some major loads such as stove and incandescent lights may not contain observable transients.

2. The NILM based on turn-on transient models can only determine the operating schedule of loads but cannot track the whole process of loads. It is because a large portion of load information is not considered except for the turn-on transient. For example, turn-off events usually do not contain sufficient transient information as turn-on events and cannot be easily identified in this way.

Also, similar to single-state load model, complex operation process of loads cannot be identified. NILM application such as load energy cannot be accurately calculated.

3. In order to acquire and process transient signatures, both high-resolution data-acquisition system and high-speed processing units are needed. Existing smart meters may have to be significantly modified to support this kind of application. To meet this requirement, particular hardware solution such as using multi-processing unit is proposed in [63]. However, it also increase the total cost significantly.

3.1.3 Proposed event-window model and its signatures

To solve the drawbacks of steady-state load model and transient load model, event-window load model is proposed and its signatures are studied.

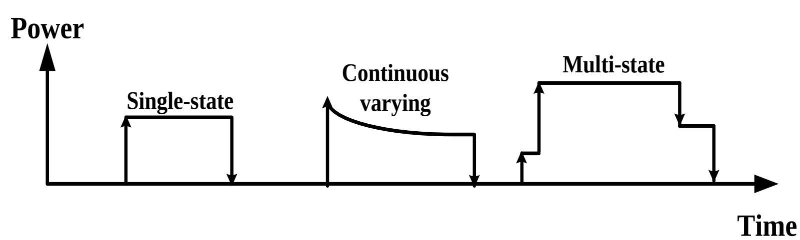


Figure 3.3:Power curves of three types of loads

Generally, the loads in the residential houses can be divided into three types as shown in Figure 3.3: single state appliance has an identical pair of ON/OFF events. During its whole operation process, the electric signatures of this load stay almost constant; continuous-varying appliance usually has a pair of different ON/OFF edges and a gradual varying power demand in the middle. Multi-state is more commonly seen as heavy or complex loads such as furnace and washer. Furnace has more than one working stage according to current environmental temperatures and washer has many steps like rinse and drainage that follows a certain operation pattern. The examples and characteristics of house loads are listed in Table 3-1.

Table 3-1 Load type and examples

Load type	Examples	Event	Power demand
Single-state	Light bulb; Toaster	ON=OFF	Flat
Continuous varying	Fridge; Freezer	ON≠OFF	Varying
Multi-state	Furnace; Washer	Multiple events	Varying or flat

Hence, the approximation of assuming all appliances have single states can lead to failure or significant error of identification for continuous-varying appliances and multi-state appliances. Besides, for combination based approach, the error on complex loads can further affect the identification of single-state appliances heavily. For accurate energy monitoring purpose, real operation processes of such complex loads need to be captured and treated specifically.

Therefore, a more generic model---event-window is proposed. An event-window is defined as the collection of all signatures between any pair of rising/falling step-changes (events) of the power demand as measured by the smart meter. Sample load windows are shown in Figure 3.4. Window 1 contains one ON and one OFF event associated with one appliance. There is no activation of other appliances in between. This is called a non-overlapping window. Non-overlapping window contains complete signature information about an appliance. Window 2 is called an overlapping window as it contains an ON event belonging to another appliance. In reality, only short duration (toaster) or always-on appliances (fridge) have more chances to present themselves in the form of non-overlapping windows. Many of them will overlap with others. The main idea of the proposed load identification method in Chapter 4 is to identify and pick out the right windows that represent the interested appliances. This is accomplished with the help from window signatures or characteristics that will be discussed later in this chapter.

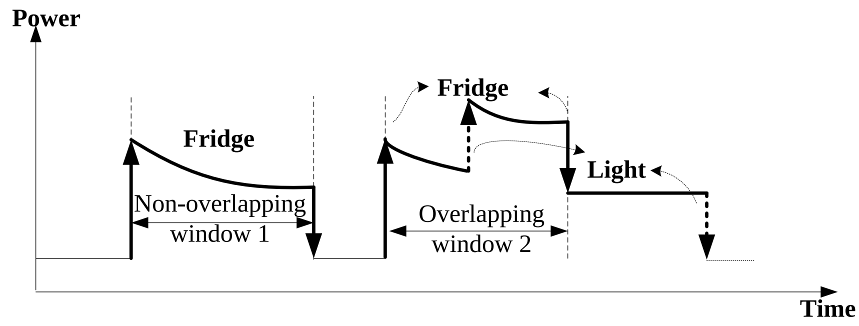


Figure 3.4: Non-overlapping window and overlapping window

Each window contains five types of signatures that are listed as below:

- Event signatures
- Event pattern signatures
- Power trend signatures
- Time/Duration signatures
- Phase signature

Comprehensively, these signatures are able to describe the load operation from a more complete and accurate perspective. They allow the load model to contain multiple steady states because different events' signatures and event pattern signatures are now included. In the meanwhile, power variation which can be often seen in continuous-varying loads is also described. Besides, time/duration signatures are proposed because they can reflect the usage behaviors of appliances. Finally, phase signature is also explained to describe the electric "location" of the load in a residential house. These signatures are respectively discussed as below.

3.2 Event Signatures

An event refers to the change of the operating state of an appliance, which can be often seen as a step change or edge in its power, reactive power or harmonic content versus time plot. The edge can be either rising or falling. Each edge can be characterized by the changes in power (P), reactive power (Q) and current waveform (W) as shown in Figure 3.5. For a certain appliance, its attributes stay generally fixed for each time operation if the system voltage does not change sharply. In single-state model, only one set of P-Q-W is considered since the OFF event is assumed to be the identical reverse of the ON event [55]. In event-window model, however, the number of P-Q-W is equal to the events that really happen inside this event-window.

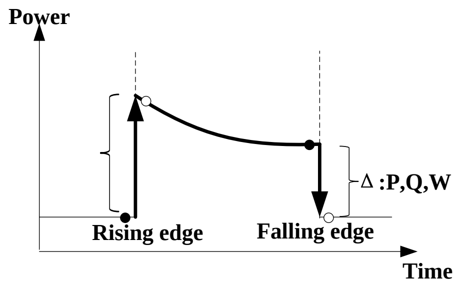


Figure 3.5: An illustration of event signatures

3.2.1 Real Power signatures

As shown in Figure 3.6, a load is connected to the power system and it can be represented by using a two-port network [98].

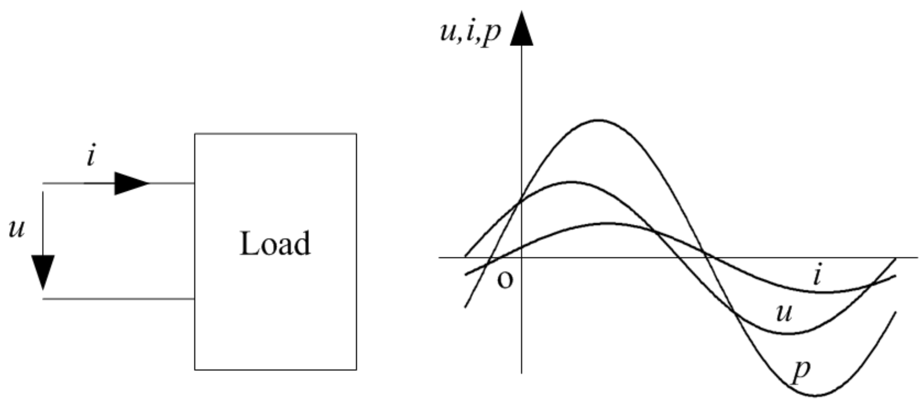


Figure 3.6: Two-port network representation of load and its voltage, current and power

For residential loads, the port voltage $u=\sqrt{2}U\sin(\omega t+\varphi_u)$ .U is the magnitude of residential system voltage and is usually around 120V in North America. ω is the fundamental frequency of power system and is usually around 60Hz in North America. φ_u is the phase angle of voltage and its value is dependent on the reference such as the swing bus voltage in system.

The current flowing into the load is $i = \sqrt{2} I \sin(\omega t + \varphi_i)$. I is the magnitude of load current and is dependent on u and the equivalent impedance of load. φ_u is the phase angle of current and its value is also dependent on the a certain reference in the system.

Real power is defined as the average power absorbed by the above network within a cycle. Its value is:

$$P = \frac{1}{T} \int_0^T p dt = UI \cos(\varphi_u - \varphi_i) = UI \cos \varphi \quad (3.1)$$

Heavily affected by the designed functions, different loads can have very distinct real powers. For example, a stove usually has a large power since it is a cooking device and a lot of heat is converted from electricity; a LED light on the other hand normally consumes a little power since it only generates a little light. Typical values of real powers of different residential load types are shown in Table 3-2.

Table 3-2 Typical values of real power of residential loads

Appliance Type	Power (W)
Incandescent light	120
Fluorescent light	40
TV	500
Stove	2000
Oven	3000
Kettle	1500
Washer	500
Microwave	1500
Fridge	200
Freezer	100

3.2.2 Reactive Power signatures

In order to introduce the concept of reactive power of loads, the equation (3.1) can be rewritten as [98]:

$$\begin{aligned} P &= UI\cos(\varphi_u - \varphi_i) - UI\cos(2\omega t + \varphi_u + \varphi_i) \\ &= UI\cos(\varphi_u - \varphi_i) - UI[\cos(\varphi_u - \varphi_i)\cos(2\omega t + 2\varphi_i) \\ &- \sin(\varphi_u - \varphi_i)\sin(2\omega t + 2\varphi_i) \\ &= UI\cos\varphi[1 - \cos((2\omega t + 2\varphi_i)] + UI\sin\varphi\sin(2\omega t + 2\varphi_i) \\ &= p_1 + p_2 \end{aligned} \tag{3.2}$$

In the above equation, p_1 does not change its direction of power transferring and thus represent the actual power consumed by the two-port network. On the other hand, p_2's frequency is 2ω and in a cycle under the actual frequency , its average value is zero. In reality, this part of power is transferring back and forth between the network source and network load. Its magnitude $UI\sin\varphi$ is defined as the reactive power of load in power system and its unit is “var”:

$$Q = UI\sin\varphi \tag{3.3}$$

As can be seen from (3.3), when φ >0, $Q = UI\sin\varphi$ >0 and it means the load is inductive and can absorb reactive power; when φ <0, $Q = UI\sin\varphi$ <0 and it means the load is capacitive and can generate reactive power. $\cos\varphi$ is defined as the power factor of load.

Compared (3.3) with (3.2), *UI* can be viewed independently and it is defined as the apparent power of load. Thus we have:

$$\begin{cases} S = UI = \sqrt{P^2 + Q^2} \\ P = S\cos\varphi \\ Q = S\sin\varphi \end{cases} \tag{3.4}$$

Their relations can be graphically shown as the triangle below:

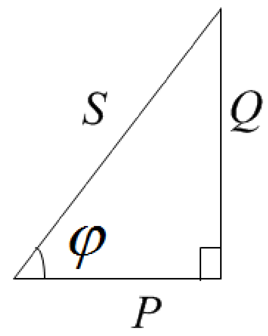


Figure 3.7: Relations of P,Q,S and φ

In power system, many resistive loads such as incandescent light, stove, oven and kettle has a power factor which is close to 1. It also means they have negligible Q or p_2 . For motor based loads such as fridge, freezer, washer, dishwasher and dryer, their equivalent circuits are similar to inductors. Thus they have a lower power factor and have a notable Q or p_2. Hence, reactive power levels can be very useful signatures when distinguishing between active loads and reactive loads.

3.2.3 Harmonic signatures

Harmonic contents are sinusoidal components of a periodic waveform with a frequency that is an integral multiple of the fundamental power frequency [99]. When harmonics are combined with the fundamental frequency component, sinusoidal waveform distortions are caused. In power system, current waveform distortion often accompanies with the use of power electronic loads, which includes high frequency switching circuits such as AC-DC, DC-DC and DC-AC conversion circuits. An example of a power electronic circuit is shown in Figure 3.8.

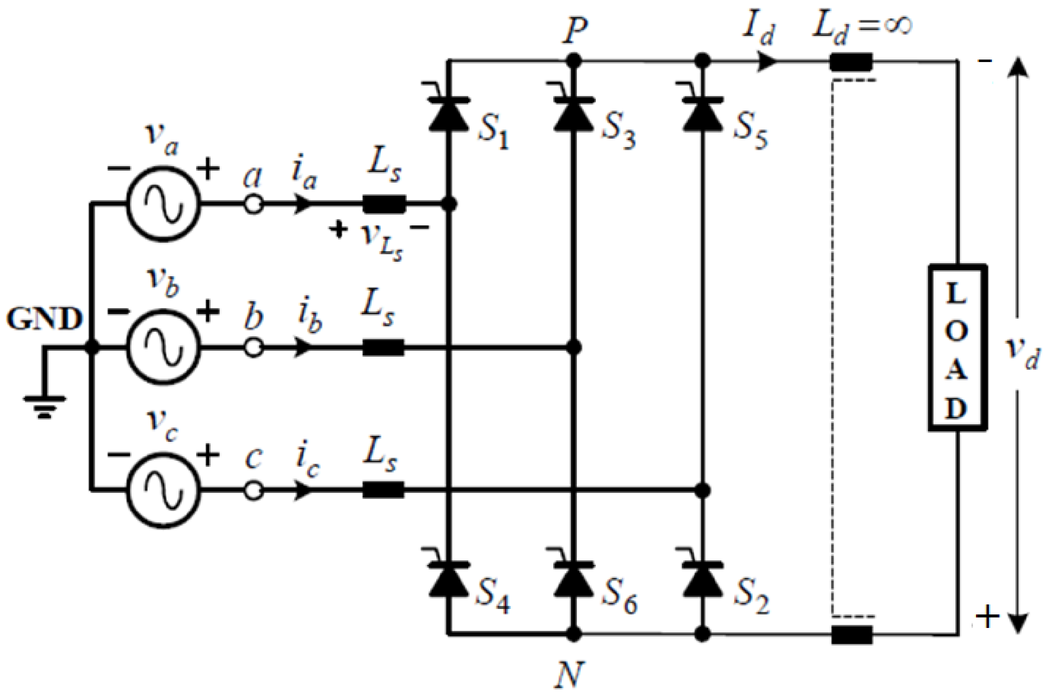


Figure 3.8: Example of a power electronic circuit--- Three phase SCR rectifier

Circuits like the above can generate distorted current waveforms which can be described by a function of *f(t)* with period *T* and fundamental frequency of $f_0=1/T$. *f(t)* can be expressed by a Fourier series:

$$f(t) = c_0 + \sum_{h=1}^{\infty} c_k \cos(2\pi h f_0 t + \phi_h) \tag{3.5}$$

where $c_k \angle \phi_h = \dfrac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} f(t) e^{-j2\pi h f_0 t} dt$.

The Fourier series decomposes the original waveforms into a series of sinusoidal components with different frequencies. The component of f_0 is called the fundamental component, and the hf_0 component is called the h-th harmonic of the periodic function. As an example, Figure 3.9 shows the distorted harmonic current of a typical DC power supply load, and Figure 3.10 presents the harmonic spectrum of this waveform [101].

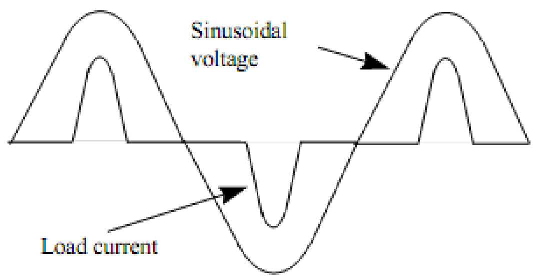


Figure 3.9: Distorted harmonic current waveform of a typical rectifier

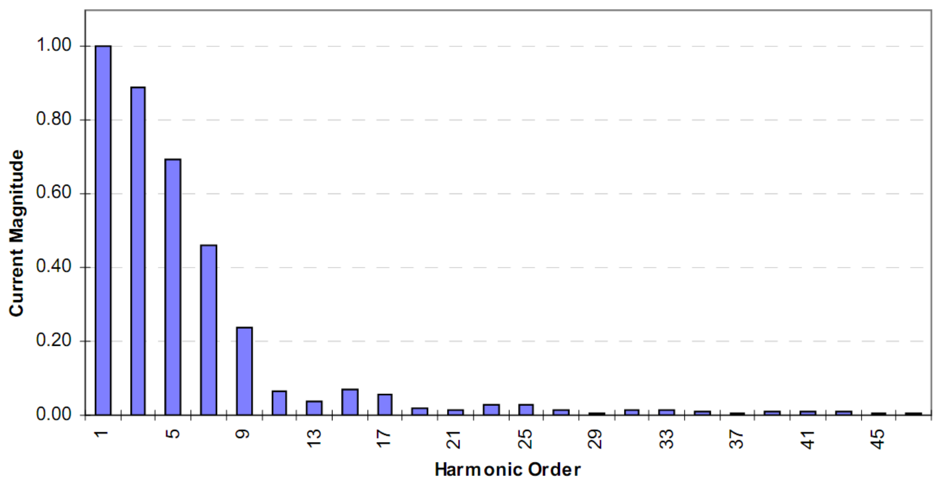
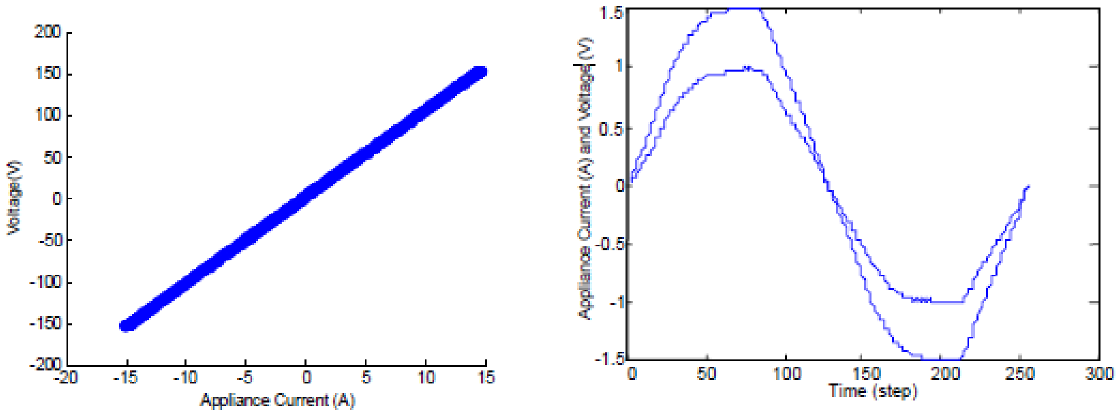


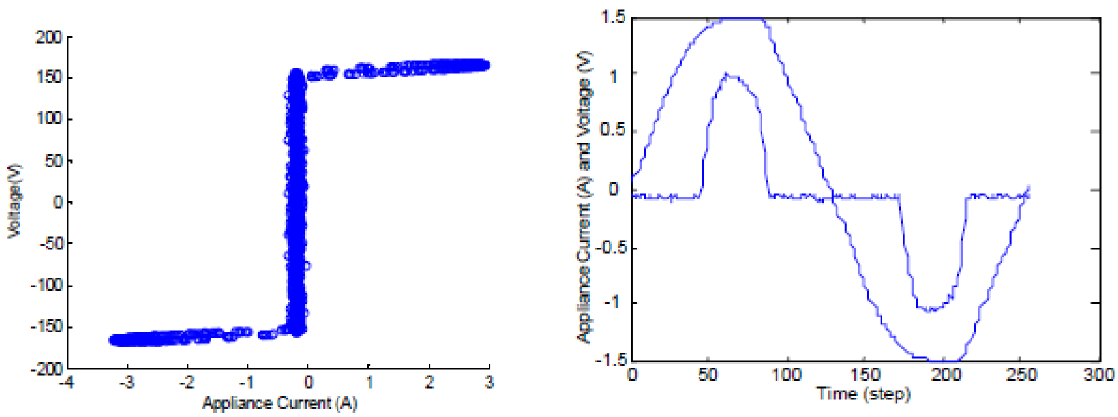
Figure 3.10: Harmonic spectrum of a typical rectifier

Residential loads can be divided into two types --- linear loads and non-linear loads. Linear load barely causes any distortion on its current waveform and thus has no harmonic contents on its spectrum after applying Fourier transform to examine their current waveforms. Examples of linear loads are incandescent light, fans, stove, oven and dryer; non-linear load can cause significant distortion on its current waveforms and large harmonic contents can be seen after applying Fourier transform to its current waveforms.

Another equivalent way to analyze harmonics is using V vs. I plot. Examples can be seen in Figure 3.11. As can be seen, linear load has a “linear” relation between its voltage and current (a) while non-linear load has “non-linear” relation between its voltage and current (b).



(a) Linear-load



(b) Non-linear load

Figure 3.11: Examples of V vs. I plots of linear and non-linear loads

3.3 Event Pattern Signatures

Event pattern signature describes the logical sequence of operation events of a load. In other words, it represents the sequence of appearances of edges. For example, a washer usually follows the below operating modes: water-fill, immerse, rinse, drainage and spin-dry. In a cycle, a fixed pattern such as +50W,-50W,+100W,-80W,+480W,-500W will be seen. This power pattern or the event pattern signature is very unique and is very important for identification of multi-state appliances. There are three types of basic event sequences: repetitive sequence, fixed sequence and the combination of the two.

Stoves, dryers or some coffee makers are typical multi-state appliance with repetitive sequence due to their internal integer-cycle controllers [96]. Figure 3.12 shows an operation process of stove.

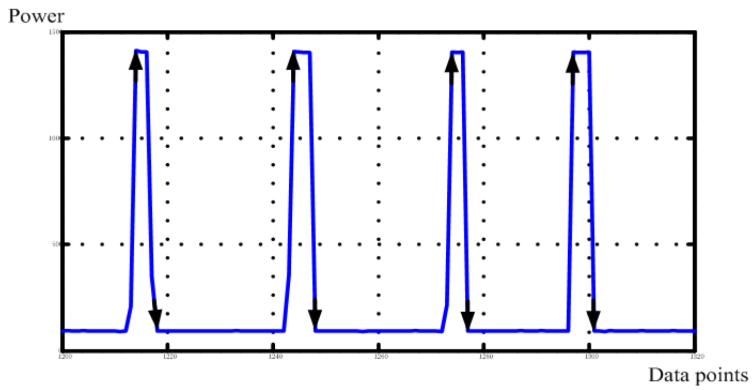


Figure 3.12: Repetitive sequence

An example of fixed sequence is washer. It is shown in Figure 3.13.

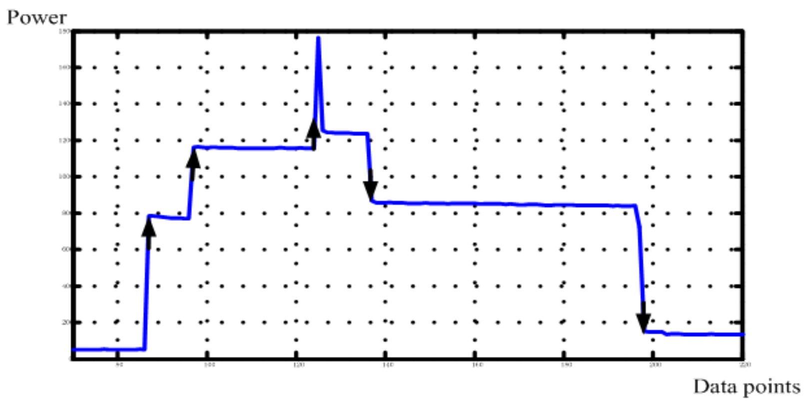


Figure 3.13: Fixed sequence

Sometimes, a combination of repetitive and fixed sequence occurs. Figure 3.14 shows a furnace. It has repetitive heating cycles. Besides, each heating cycle includes a fixed sequence pattern. According to the environment temperature, the heating cycle may show up 2-5 times closely to each other.

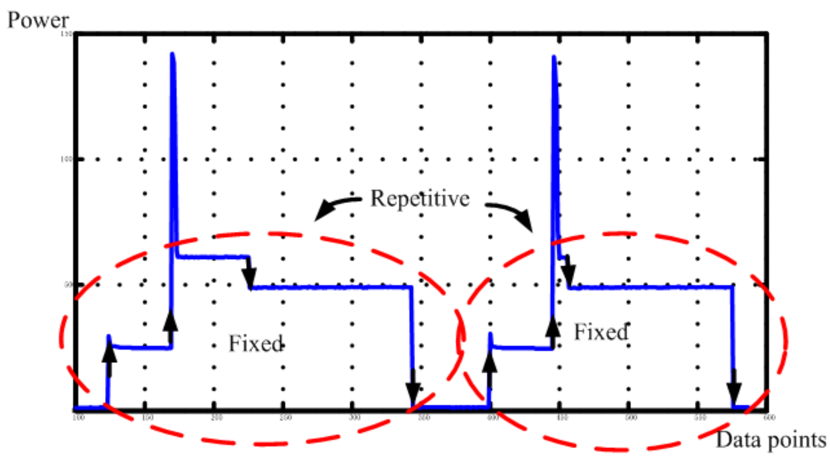


Figure 3.14.Combination sequence

Table 3-3 shows some examples of appliances with the sequence patterns as discussed above measured through experiment.

Table 3-3 Sequence pattern and examples

Load type	Examples
Repetitive sequence	Dryer; Stove; Some coffee makers
Fixed sequence	Incandescent light bulb; Fluorescent light bulb; Kettle; Microwave; Toaster; Oven; Fridge; Freezer; Computer
Combination	Furnace, Some dishwashers

3.4 Power Trend Signatures

A trend signature refers to the variation of power demand between two neighboring events. For example, an inductive motor often accompanies with a rising spike right after being turned on due to its large inrush current. This kind of transient can be seen from Figure 3.15.

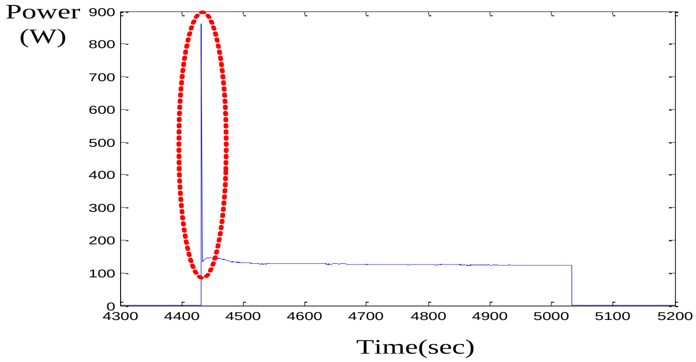


Figure 3.15: Trend signature 1---rising spike (Fridge)

After start, as shown in in Figure 3.16, with the motor speed gradually increases, its current drawn may sometimes decrease and form a gradual falling curve.

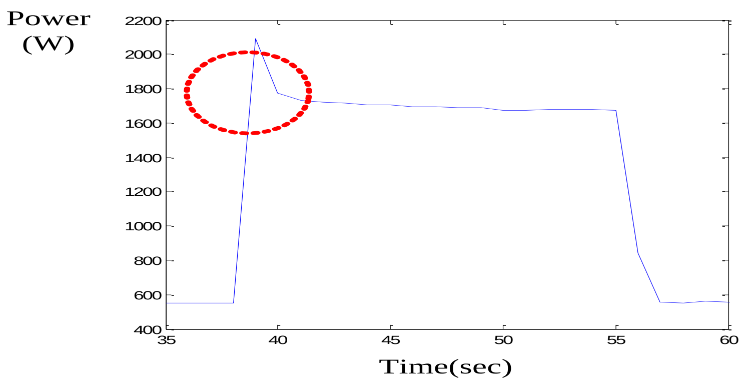


Figure 3.16: Trend signature 2---gradual falling (Dryer)

Some electronic devices may experience an instant interruption. A TV set may experience a falling spike at moments of switching channels. An example is shown in Figure 3.17.

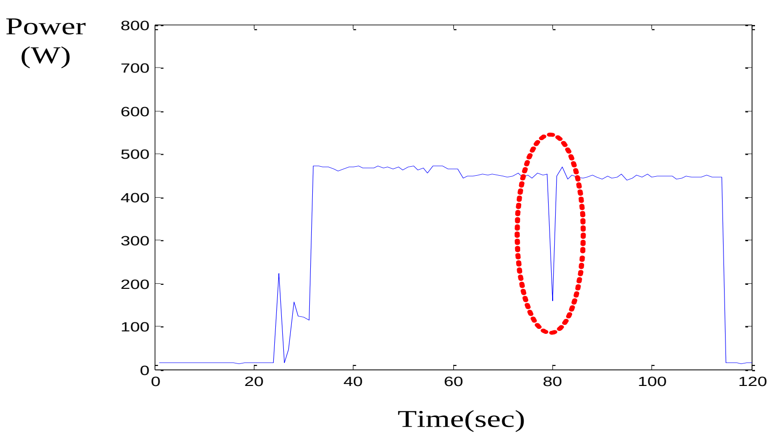


Figure 3.17: Trend signature 3---Falling spike (TV)

Pulses are usually caused by electronic switches. A lot of stoves have pulses because they have integer-cycle controllers inside [96]. It prevents itself from overheating. Another example is an inverter based motor device that adjusts its frequency/speed all the time. An example can be seen from Figure 3.18.

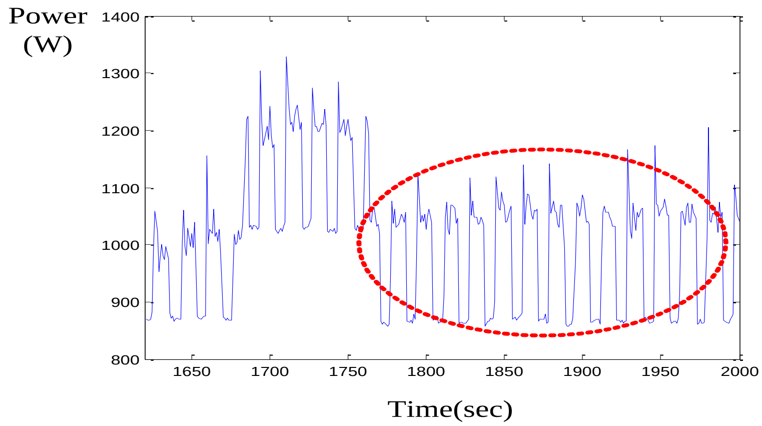


Figure 3.18: Trend signature 4---Pulses (Washer)

A lot of appliances have negligible transient characteristics and present almost flat power curves during operation. Actually, this is the only type of appliance which can be ideally represented by the old single-state load model used in previous researches. An example can be seen as below:

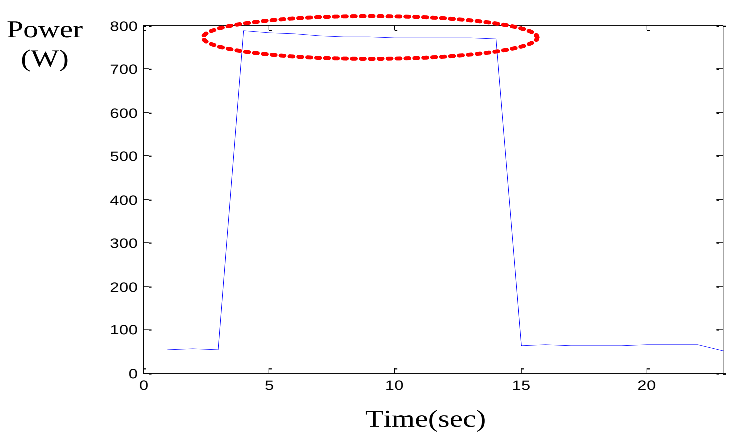


Figure 3.19: Trend signature 5---Flat (Kettle)

In contrast, as can be seen from below, some appliances may have continuous low-frequency fluctuations all the time instead of a steady state. An example can be seen from Figure 3.20.

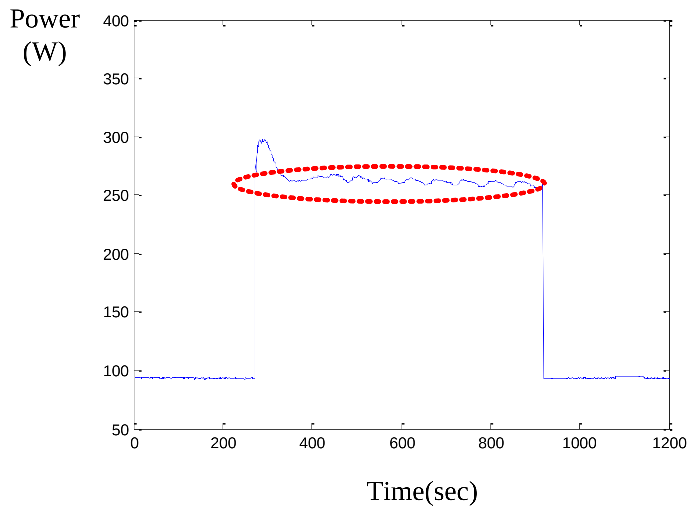


Figure 3.20: Trend signature 6---Fluctuation (Freezer)

In addition, some audio and video electronic devices such as audio box, laptop and desktop PC may have high-frequency noises which are shown in Figure 3.21. It is because their power consumptions constantly vary with the sound, visual display or other tasks their internal analogue or digital circuits are processing.

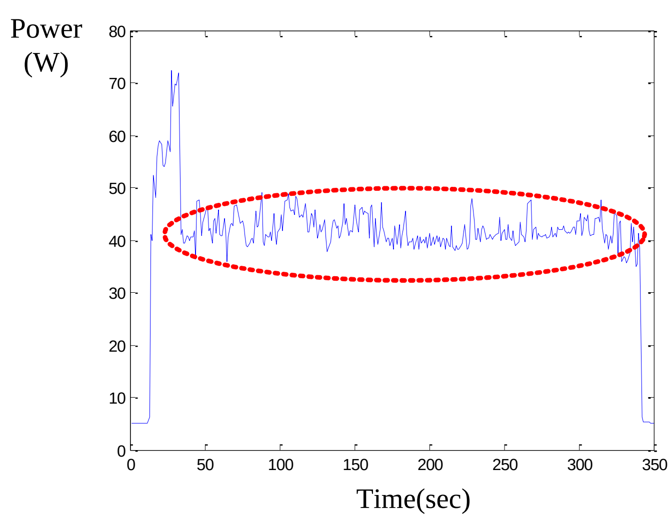


Figure 3.21: Trend signature 7---High frequency noise (Laptop)

It should be noted some appliances such as fridge may have more than one type of trend signatures. Also, some trends are not only found in one type of appliance but usually in several types of appliances due to their common electrical characteristics.

Since continuous power points are being measured from smart meter, trend signatures can be further represented and defined by several slope $(\Delta P / \Delta t)$ variation modes described in Table 3-4. Also, these slope modes can be used as a scanning method for identification purpose.

Table 3-4 Trend signatures and slope characteristics

Trend signatures	Slope characteristics
Rising spike	A large negative slope following a larger positive slope
Gradual falling	Continuous small negative slopes
Falling spike	A large positive slope following a large negative slope
Pulses	Continuous pairs of large slopes
Flat	Continuous small slopes
Fluctuation	Continuous small slopes; signs of slopes slowly change
High frequency noise	Continuous small slopes; signs of slopes quickly change

3.5 Time/Duration Signatures

The time of load window appearance relates closely to its function. There are some statistical studies on residential load modeling which present typical load on-hours such as shown in Figure 3.22 [31],[97].

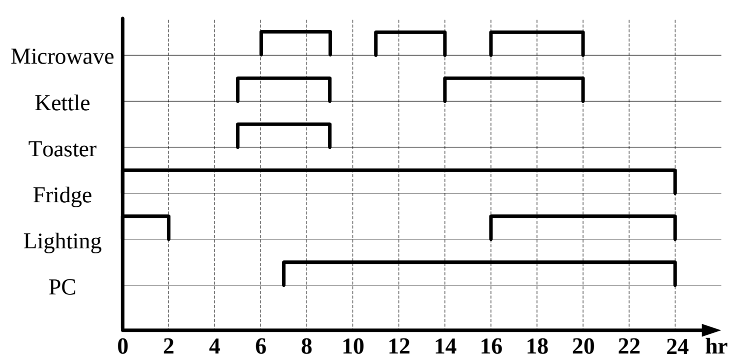


Figure 3.22: Typical appliances on-hours for weekends

As can be seen, microwaves are more expected to be seen before breakfast, lunch and supper; lights are usually turned on in the early morning or after dark; fridge and furnace are likely to run throughout 24 hours.

Duration of load window is also determined by its function characteristics. No one keeps microwave on for more than 30 mins at a time. One working cycle of fridge is barely longer than 40 mins. As for lights, depending on each use, it might be on from minutes to hours. Based on statistical survey, some typical load window lengths are given in Table 3-5.

Table 3-5 Typical load window Lengths

Load name	Min length	Max length
Fridge(cycle)	>10 mins	<40 mins
Freezer(cycle)	>10 mins	<40 mins
Furnace(cycle)	>5 mins	<30 mins
Stove	>5 mins	<45 mins
Kettle	>3 mins	<15mins
Washer	>20 mins	<90 mins
Dryer	> 20 mins	<75 mins
Bedroom light	>0 min	<5 hrs
Living room light	>0 min	<8.5 hrs
TV	>0 min	<10 hrs

3.6 Phase connection Signatures

There are two 120V hot wires installed in a typical North American residential house as shown in Figure 3.23. Hereby, the two wires can be named as A and B. Most appliances are connected between A or B and neutral. However, some heavy appliances such as stove and dryer are connected between A and B to gain a 240V voltage. Inside a meter, two CTs are connected to A and B individually. As a result, from aggregated signals of CTs, one can tell if one appliance is phase-A, phase-B or phase A-B type. It should be noted phase-AB appliance has symmetrical edges that can be detected by both CTs. For most appliances, once they are placed or installed in a house, they will never be moved. Examples are stove, fridge, microwave, furnace, lights and even large TVs. Only very a few of them have uncertain phase signatures such as laptop.

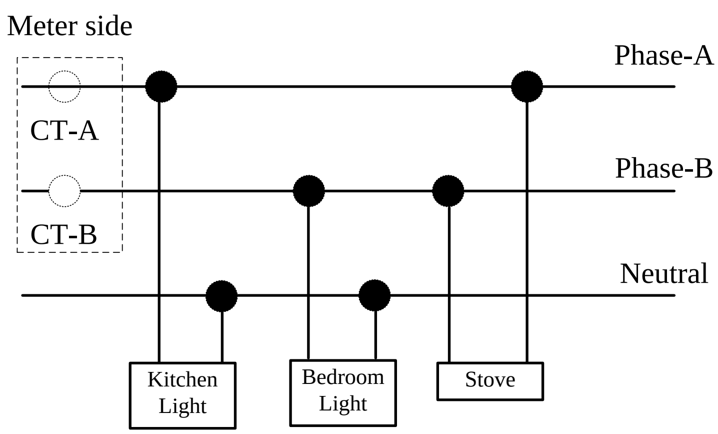


Figure 3.23: North America residential wiring

3.7 Summary

This chapter proposed a novel load model--- event-window model. Compared with the existing single-state and transient load models, the proposed event-window model can accurately depict the process of complex loads such as continuous varying loads and multi-stage loads which in fact take up a large portion of end-user loads in power system.

Based on the proposed load models, five categories of signatures were discussed in detail. Different events' signatures and event pattern signatures were included so that multi-stage appliance operation can be accurately described. In the meanwhile, power variation which can be often seen in continuous-varying loads was explained. Besides, based on the behavior statistics of different types of appliances, time/duration signatures were presented. Finally, phase signature was introduced to label the electric "location" of the load in a North American residential house.

The proposed event-window model along with the above load signature studies established a solid foundation for proposed event-window based load identification method that will be discussed in Chapter 4.

Chapter 4

Event-window based Load Identification

This chapter presents the details of event-window based load identification method. Chapter 3 has discussed all types of event signatures. How to achieve load identification based on these signatures is therefore the main problem this chapter attempts to solve.

Firstly, an overview of the existing NILM algorithms is given. Secondly, theoretical details such as the proposed event-window load identification procedure and calculation of individual signature scores are elaborately explained. Thirdly, the implementation of a practical NILM system in a real residential house is presented. The details related to the practical side are such as data acquisition and preprocessing also addressed. Fourthly, the proposed algorithm is verified extensively based on the real data from several local houses and a public dataset. Finally, to further clarify the advantages of the proposed algorithm, a previous signal-combination based method is also implemented using neural network technique. With a bottom-up based simulation, comparative studies of the proposed method against the signal-combination based method are conducted. The findings are discussed and concluded in the end. Some contents in this chapter have been submitted as publication in [104].

4.1 Overview

Load identification is the most essential task for NILM algorithm. Researchers have attempted to solve the problem using different algorithms. In [55], the authors suggest dividing these algorithms into two categories---signal-combination based algorithm and change of appliance state or event based algorithm. This chapter intends to introduce a new category---event-window based algorithm. The structure of NILM algorithms is shown in Figure 4.1. The existing signal-combination based and event based algorithms are respectively reviewed. Based on the reviews and identified problems, event-window based algorithm is proposed eventually.

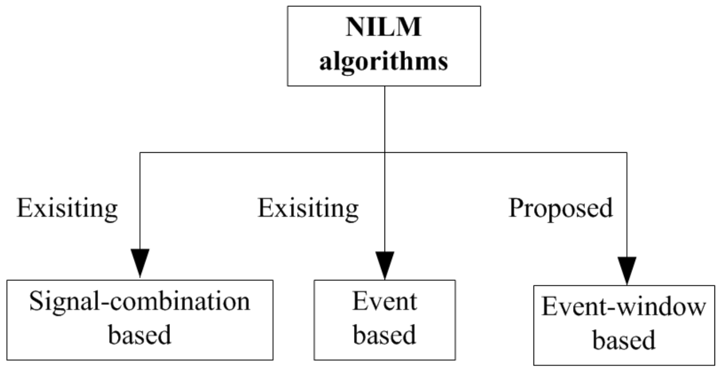


Figure 4.1: The structure of NILM algorithms based on [94]

4.1.1 Review of signal-combination based algorithms

In [94], the first major research stream is signal-combination based algorithms [59]-[60],[67]-[79]. This stream looks for a combination of appliances that the resultant aggregate signal is as close to the observed signal as possible. In other words, different appliance combinations are matched simultaneously to the disaggregated signal and the best match is selected as the desired combination. Typically, it is achieved through the prior training process on a limited number of combinations first. Afterwards, it is believed the trained model can also support future classification on untrained combinations.

One example is shown in Figure 4.2, in an enclosed system with limited and fixed number of appliances, specific appliances are turned on and their aggregated signals at the meter-side are respectively recorded and labeled. In other words, training sets can be generated this way and each training set is relevant to a specific combination. Since each appliance has an ON and OFF state (if not considering the middle states), for a system that consists of N appliances, there are in total 2^N combinations. It is therefore impossible to try all the possible of combinations when N is big. However, limited number of tries may be enough already to set up the classification model. Then the classification model should be able to predict future combinations.

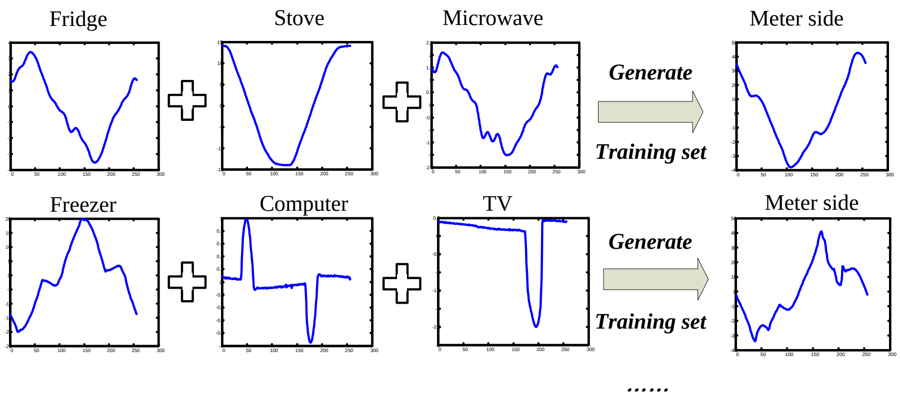


Figure 4.2: Example of the training process for signal-combination based algorithm

Typical pattern recognition approach such as neural network and support vector machine can be used. [69] uses the real and imaginary parts of harmonic contents shown as below as training inputs and a single-hidden-layer MLP model for training.

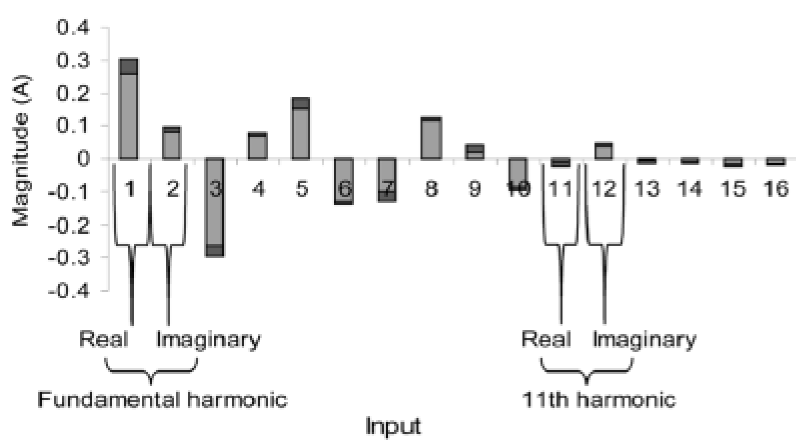


Figure 4.3: Training inputs from [69]

To enhance the accuracy, [78] further combines other features such as surrounding conditions (temperature, humidity and so on) with various electric features together into training process. The entire training process becomes very complicated. The features are shown in Figure 4.4 below:

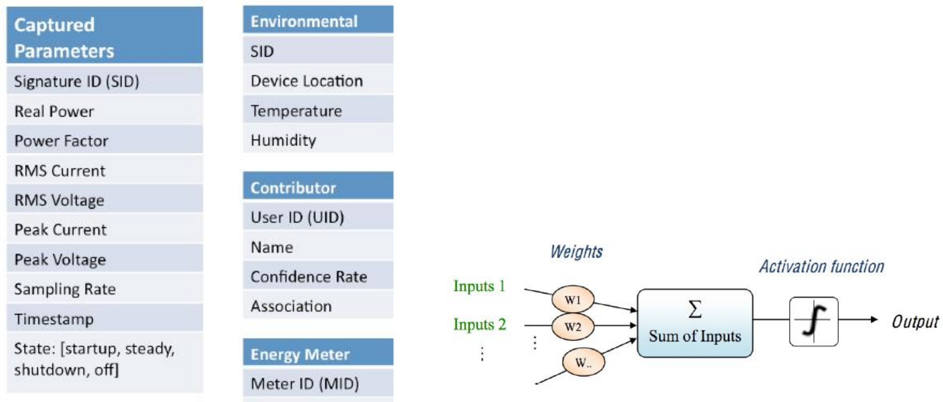


Figure 4.4: Training inputs from [78]

Some optimization based algorithms such as least residue, integer programming [72], and genetic algorithms [73] can also be used together with classifiers such as neural networks. An example is shown in [59]. As can be seen, different features such as current waveform (CW), Eigenvalues (EIG) are fed in using either least residue or neural networks into the "Committee Decision Mechanism" (CDM).

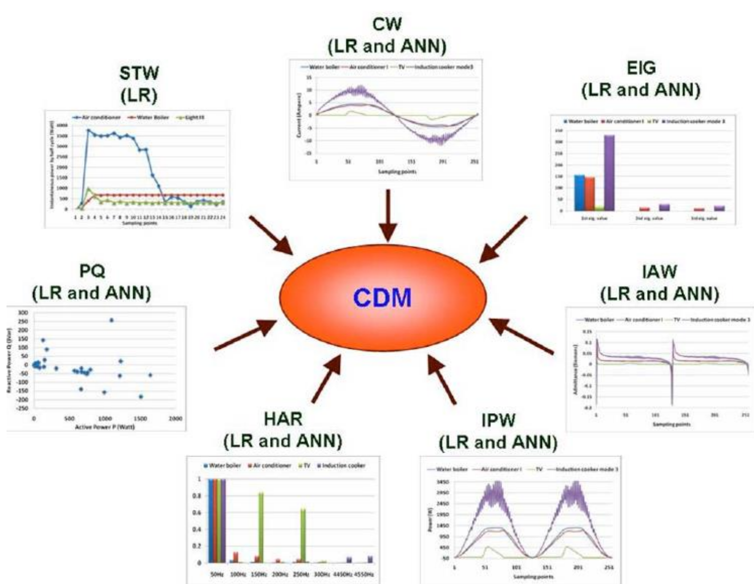


Figure 4.5: Training inputs from [59]

In practice, there are several fatal drawbacks for this signal-combination based stream:

1. Since this stream directly deals with aggregated or combined signal, all loads in the system have to be covered. For example, in a 20-appliance house, even if only 3 appliances are interested to be identified, the other 17 appliances still need to participate in the training process because they can significantly affect the aggregated signal when they are turned on together with the interested 3.

2. Due to the number of appliances and the number of possible combinations, to obtain a reliable classification model, very extensive training process is needed. For example in [70], the basic training set includes 300 combinations for only 9 loads. This is not very practical for ordinary household owners to perform.

3. The obtained model is very vulnerable to inventory change. If the user changes the load inventory, for example after he purchases a new appliance or replaces his old appliance, the previous trained model becomes not reliable and the training process has to be redone. It further makes the algorithm less practical.

4. The algorithm is not likely to be able to deal with multi-state or continuous varying loads. If different states of appliances are all taken into consideration, the training process will be over-complicated due to the large number of combinations. For most of the existing researches, only a single state is considered for training. This will lead to large error when dealing with a system that owns complex loads.

Generally speaking, the signal-combination based methods only work well for system consisting of ON/OFF appliances. In the meanwhile, the studied system should have a fixed load inventory. Besides, the users have to be knowledgeable and patient enough to go through such a long initial process. In order to more clearly reveal the above drawbacks, a signal-combination based algorithm based on [69] is implemented and compared with the proposed method in Section 4.5.

4.1.2 Review of event based algorithm

According to [94], the other algorithm stream is based on the change of appliance operation state or event. The first NILM publication done by MIT [55] is based on this approach. As shown in Figure 4.6, only real power and reactive power are considered as the event signatures due to the hardware limit and processing speed at that time. When an event is detected, it is compared to the P-Q map and identified as the name of the closest "circle" in the map.

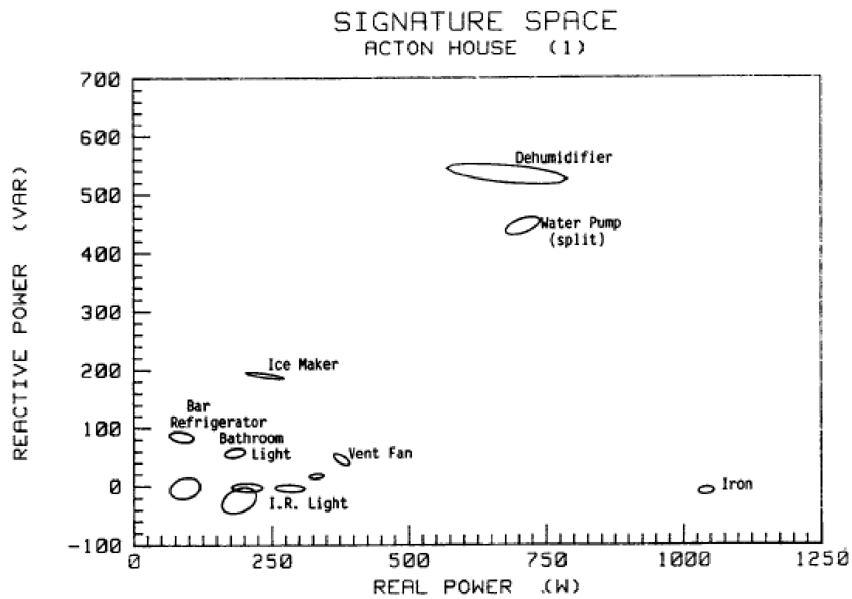


Figure 4.6: Event based algorithm from [55]

As can be seen from Figure 4.6, some appliances have very close signatures and overlap "circles" on the P-Q map. [61]-[64],[87] have attempted to enhance the separation by including transient characteristics into consideration. As shown in Figure 4.7, different sections of turn-on transient are enveloped and used as signatures for event match (identification).

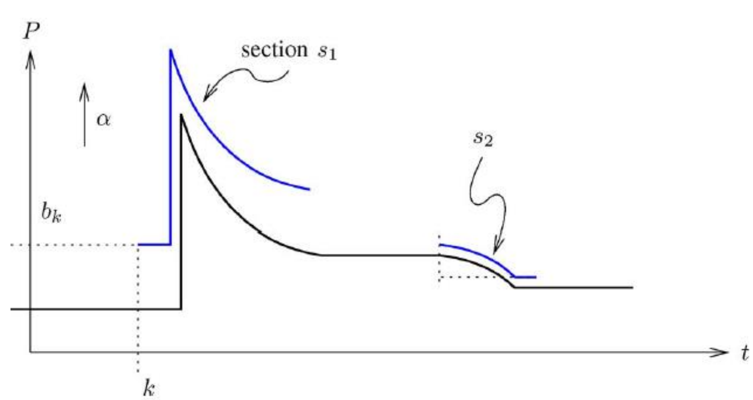


Figure 4.7: Event based algorithm from [87]

Agreed with the comments given in [94], this research stream is more robust and practical. The event based algorithms only focus on interested target loads and do not need to cover all the loads in the system. It can also resist the change of load inventory. However, it has the following drawbacks that need to be improved:

1. The existing event-based algorithm looks into “isolated” events at different time points but not all the events of a load as a whole. When the number of events increases, with only isolated pieces of information, errors or mis-identification can easily occur.

2. Since events are separately identified, the whole operation process and energy of load cannot be re-organized and tracked. In [62], it is claimed that only operating schedule of loads can be obtained. Also, it cannot effectively deal with multi-state and continuous varying loads.

4.1.3 Proposed event-window based algorithm

To solve the problems of the above algorithms, the event-window based algorithm is proposed as a new attempt. Instead of looking into the aggregated signal or separated events, it looks into the related events of a certain load as a whole. In other words, it takes the entire operation process of a load into consideration.

Its procedure is shown in Figure 4.8. The events of interested loads are detected, identified and finally re-organized as output. Firstly, meter-side current and voltage are acquired and transferred to event detector. Related data acquisition will be explained in Section 4.3.1. Secondly, as explained in Chapter 2, event detector detects all the events of all the loads. Thirdly, through specific data preprocessing, real power, active power, harmonic of all events are retrieved and organized. Also trend signatures are scanned along the power curves. This will be addressed in Section 4.3.2. The next steps are the core of load identification procedure--- through signal split, window candidate selection and evaluation, decision of identification can finally be made with respect to the prior collected appliance signatures. These signatures are used as benchmarks and they are named as appliance candidates. The details of this core step will be discussed in Section 4.2. Finally, output results are organized and displayed on user interface. This will also be presented in Section 4.3.3.

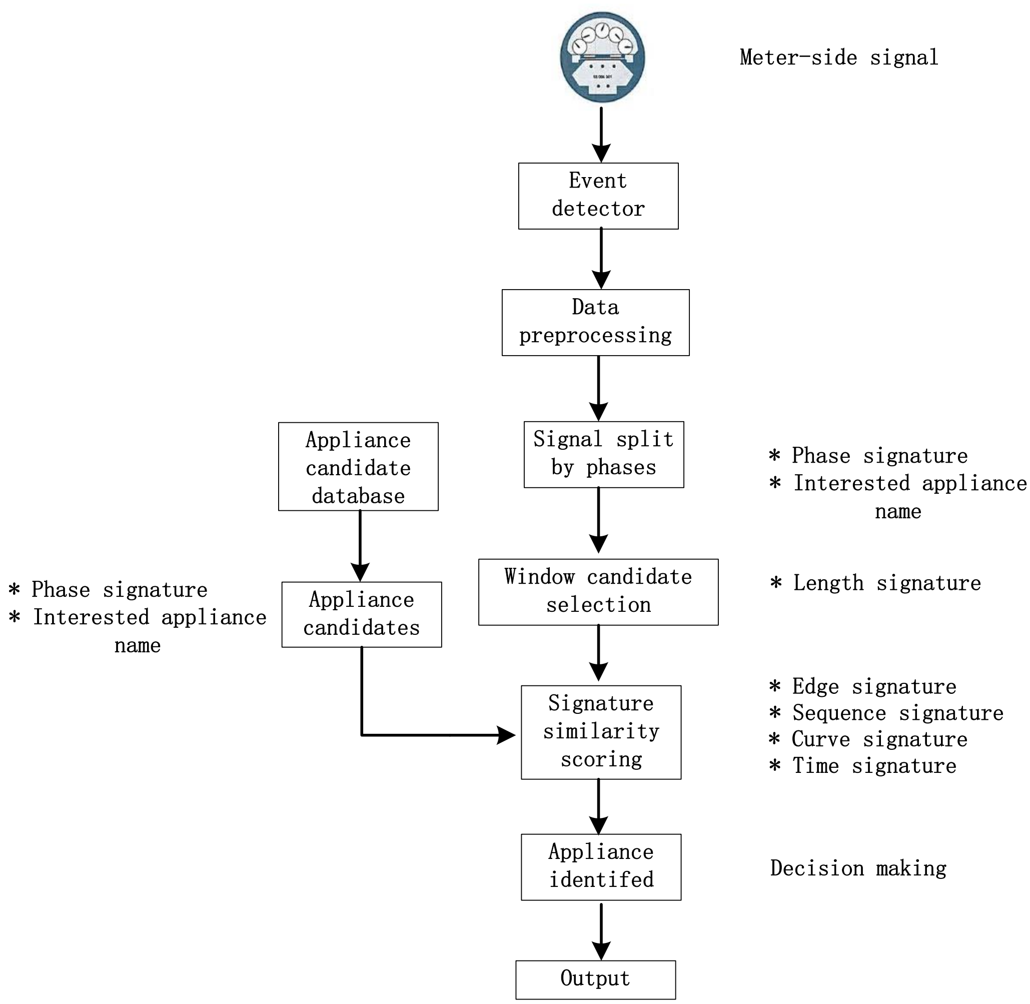


Figure 4.8: General Identification procedure

4.2 Event-window based algorithm

4.2.1 Load identification procedure

A. Split signal by phases

Two CTs inside meter naturally separate the aggregated signal acquired by smart meter into signals of two phases: phase-A signal and phase-B signal. Accordingly, to deal with phase-A signal, only phase-A loads will remain as candidates. So is for phase-B signal. Normally, a light bulb connected to Phase B will never be seen from CT-A. This is a very important step because many candidates can be ruled out very easily.

Two exceptions should be addressed: for phase A-B appliance, since any of its edges shows up simultaneously at both phases, it will be treated as appliance candidates only if two CTs can detect two identical edges at the same time. Since the processes at both phases are the same, any phase signal can be chosen for identification purpose; for portable appliance, on the other hand, since it has an uncertain phase signature, it will be left as candidates for both phase signals.

B. Select window candidates

After signal split, suppose a section of aggregated signal from CT-A is measured as shown in Figure 4.9.

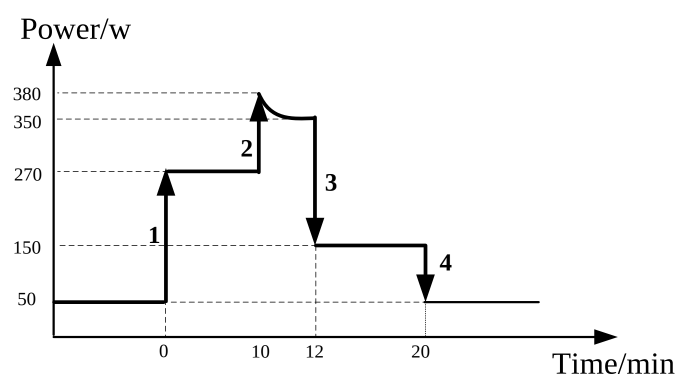


Figure 4.9: A section of meter signal collected from CT-A

Table 4-1Window candidates vs. appliance candidates (1)

Appliance candidate	Window candidate 1-3	Window candidate 2-4	Window candidate 1-4	Window candidate 2-3
Kettle				
Fridge				

Light				
Furnace				

Firstly, applying power slope based event detection to the data, 2 rising edges and 2 falling edges can be located (two large positive slopes and two large negative slopes) and labeled. As discussed in section II, signal collection between any pair of rising and falling edges is considered as a window candidate. In Fig.9, there are in total 4 window candidates: 1-3, 2-4, 1-4, 2-3. Those window candidates are waited to be compared with appliance candidates one by one.

Typically, a residential house may have more than 500 rising and falling edges per day. It indicates the potential number of window candidates per day can be 250,000 in maximum. This will cause too much computing burden. One way to reduce the window number is to discard some window candidates by using appliances' possible window lengths.

According to Table 4-1, window candidate 2-3 is too short to be possible for kettle, fridge and furnace. Candidate 1-4 is also too long for kettle. Those windows are firstly thrown away even before they enter into next evaluation step. In fact, this window length limit has stronger effect on refining longer period data. For a day period, only 120-200 window candidates will be left based on multi-case studies.

C. Evaluate window candidate with respect to appliance candidate

This is the core step of identification. In this step, the collected signatures of appliance candidates in prior are set as initial benchmarks. Then each of the left window candidates will be compared to these benchmarks and obtain their individual scores on event, event pattern, trend and time signatures ($S_{evt}, S_{ptn}, S_{trd}, S_{time}$). Then those 4 scores will be synthetically considered to get an overall score. Afterwards, the overall score is used to judge if this window candidate is matching the benchmark---the appliance candidate. The mathematical judgment is completed by using the following linear equation.

$$g(x) = \omega^T x - \gamma \tag{4.1}$$

with

$$x = \begin{bmatrix} S_{evt} \\ S_{ptn} \\ S_{trd} \\ S_{time} \end{bmatrix}, \omega = \begin{bmatrix} \omega_{evt} \\ \omega_{ptn} \\ \omega_{trd} \\ \omega_{time} \end{bmatrix} \tag{4.2}$$

x includes the scores of each signature with respect to the prior benchmarks. The determination on individual scores will be elaborated in Section 4.2.2. ω is the weight vector since for different types of appliances, different signature sub-scores are not equally important. $\omega^T x$ is therefore the overall score. γ is the qualification threshold. The result *g(x)* has two scenarios: when $g(x) \geq 0$, this window candidate is determined as this appliance; otherwise not.

Overall, the above process is equivalent to a linear discrimination classifier. This classifier is a two-class classifier. Each specific load has its independent weights and classifier. Its classifier judges if a certain event window matches itself or not. Table 4-2 lists up some typical ω and γ values of several appliances.

Table 4-2 Examples of load ω and γ

Load name	Distinctive signatures	ω^T	γ
Fridge	event, trend	[0.53 0.17 0.22 0.08]	0.85
Microwave	event, time	[0.65 0.19 0 0.16]	0.85
Furnace	event, pattern	[0.53 0.47 0 0]	0.85
Stove	event, pattern, time	[0.51 0.3 0 0.19]	0.85
Washer	event, pattern	[0.55 0.45 0 0]	0.8
Kettle	event	[0.86 0.14 0 0]	0.85
Laptop	event, trend	[0.5 0.25 0.25 0]	0.8
Average	---	[0.59 0.28 0.07 0.06]	0.85

The weights ω are firstly estimated based on the observation and analysis for different appliance types. For example, knowing furnace and washer have unique event pattern signatures, S_{ptn} will be emphasized; knowing microwave is often used before meals, S_{time} is emphasized. Generally, event signature is always important since it determines the electric characteristics of a window. Event pattern signature is important too, especially for multi-state appliances. Trend signature is important for motor related and some electronic appliances which contain transient characteristics. Time signature functions accessorily and is more effective for time-oriented loads such as kitchen appliances. After weights are pre-defined, their values will be optimized and verified through a simulation program. This program generates numerous testing windows based on existing load signatures and then it adjust values of ω and γ to ensure that maximum number of correct identification can be made for each type of load. This is elaborated in Section 4.2.3.

Usually weight vector ω does not change much for the same type of loads from a house to another. This is because the same type of loads normally share very similar working principles and follow a certain standard in a particular region. The row “Average” in Table 4-2 gives a rough setting when the load type is not known. This can be used to cope with unfamiliar loads whose weights have not be studied and optimized. The advantage of the weight vector is that when comparing, there is no absolutely dominant signature---various signatures are bonded together to ensure fairness and accuracy.

Identification threshold γ is normally set as 0.8-0.85 for most cases. It should be noted all the weights and thresholds can still be adjusted manually and locally for a specific house in a small scale such as ±0.15 to achieve the optimum accuracy. For example, γ can be lowered if imposed signal noise is significant. Weights can also be tested and adjusted accordingly. This is also explained in the end of Section 4.2.3.

D. Make decision

In the end, Table 4-3 is calculated according to equation (4.1) and the $g(x)$ values are filled in Table 4-3.

Table 4-3 Windows candidate vs. Appliances candidate (2)

Appliance candidate	Window candidate 1-3	Window candidate 2-4	Window candidate 1-4	Window candidate 2-3
Kettle	0.15	-0.45		
Fridge	-0.3	0.15		
Light	-0.85	-0.85	-0.85	
Furnace	-0.75	-0.75	-0.75	

From the signs of classifier values, window 1-3 is determined as a kettle while window 2-4 is a fridge since their values are greater than 0. If no positive value is found, it means the events could be due to mis-match of events or an unknown appliance not registered in the database yet (maybe not interested by users either).

This linear classifier can also be substituted by more sophisticated classifiers such as neural networks or decision tree. Those variations are not discussed here.

4.2.2 Individual signature scoring

This section addresses on how to calculate individual signature scores $S_{evt}, S_{ptn}, S_{trd}, S_{time}$ according to the signature benchmarks from the appliance candidates.

A. Event signature score S_{evt}

From the signature database, an appliance candidate only includes its own P-Q-W event sets. In contrast, a window candidate may include other events caused by overlapped appliances. The comparison is trying to answer if this window candidate includes all of the appliance candidate's events. Thus the process is like this: each of registered events will be compared throughout all the events in window candidate one by one. Then:

$$S_{evt} = \frac{N_e^{'}}{N_e} \tag{4.3}$$

where N_e is number of event types defined in appliance candidate and $N_e^{'}$ is recognized number of appliance event types in window candidate.

As shown in Figure 4.10, both of the two registered events B-D are found in the window ($N_e = N_e^{'} = 2$). However, if only B exists, it is very likely the window candidate is only one part of the appliance process and its $S_{evt} = 0.5$ ($N_e = 2, N_e^{'} = 1$).

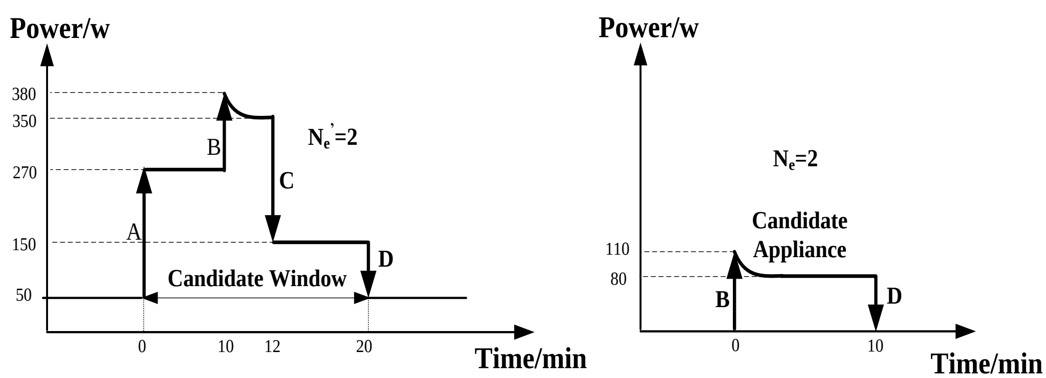


Figure 4.10: Event signature scoring

P, Q can be easily compared since they are quantitative values. As for current waveform W, one can conduct comparison in either time-domain or frequency domain. Selecting proper harmonic orders can also eliminate the impact from noises and dc offset.

Since an event is determined by three sub-attributes, again, different weights can be set to those attributes: for linear and active load such as stove, P should be emphasized; for non-linear load such as microwave, W should be emphasized; for reactive load such as fridge, Q should be emphasized. Those weights can be pre-defined for different appliance types. Synthesizing them together, two events can be determined as identical or non-identical.

Overall, S_{evt} indicates the existence of events of appliance candidate in window candidate.

B. Sequence signature score S_{seq}

For ON/OFF type appliance, it has a fixed sequence of events; for multi-state appliance, as discussed in section II, fixed sequence and repetitive sequence may either be found.

For fixed sequence events, they always follow a certain order pattern. For a window candidate, its event order should comply with the order pattern defined in appliance candidate. For example, as shown in Figure 4.11, a space heater has 5 events in the order of A-B-C-D-E. It is expected to find A-B-C-D-E in the window. On the other hand, an A-B-D-C-E sequence may imply a different appliance process and B-C-A-D-E is even more different.

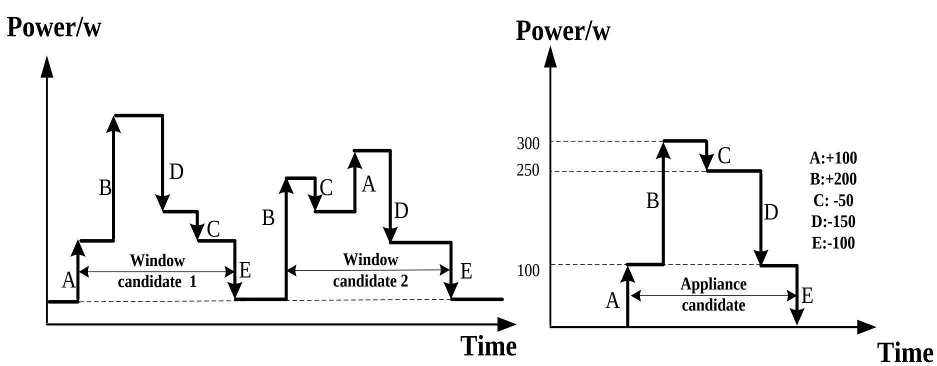


Figure 4.11: Sequences of two window candidates compared to the appliance candidate

To quantify the difference of two sequences, a simple method based on calculating the position changes of letters is proposed. Suppose the appliance candidate above has a sequence labeled using letters A-B-C-D-E. Window candidate 1 has A-B-D-C-E; window candidate 2 has B-C-A-D-E; window candidate 3 has C-B-A-D-E. Then we have the Table 4-4.

Table 4-4 Example of position change

Window	Position	Length of

candidate	change of letters	changed position
A-B-D-C-E	C:3→4 D:4→3	|4-3|+|3-4|=2
B-C-A-D-E	A: 1→3 B: 2→1 C: 3→2	|3-1|+|1-2|+|3-2|=4
C-B-A-D-E	A: 1→3 C: 3→1	|3-1|+|1-3|=4

It is easily known that B-C-A-D-E and C-B-A-D-E are more disordered than A-B-D-C-E compared to the original sequence A-B-C-D-E based on their lengths of changed positions. For a given sequence composed of *n* letters/events, the maximum possible length of changed position is:

$$M = \sum_{k=0}^{L}[n-(2k+1)]\ ,\ L < \frac{n-1}{2} \tag{4.4}$$

From (5), it can be calculated that:

for ON/OFF appliance, *n =2,M=2* (AB→BA);

for three-event appliance, *n =3, M=4* (ABC→CBA);

for four-event appliance, *n =4,M=8* (ABCD→DCBA) ;

for five-event appliance, *n =5,M=12*(ABCDE→EDCBA).

Based on the discussion above, S_{trd} for appliance with fixed sequence can be quantified as

$$S_{trd} = 1 - \frac{N_f}{M} \tag{4.5}$$

where N_f is the length of changed position of a window candidate as shown in Table 4-4.

For example, sequence C-B-A-D-E's $S_{trd} = 0.67$ ($N_f = 4$) while sequence E-D-C-B-A's $S_{trd} = 0$ since it is completely opposite to the original sequence A-B-C-D-E ($N_f = 12$).

One exception is if S_{evt} is already found smaller than 1, S_{seqf} will be automatically set to zero due to mismatch in the number of relevant events.

The appearances of repetitive events are also counted in the step of determining S_{evt} and only if its number is more than one, it is recognized as an repetitive event.

$$S_{trdr} = \frac{N_r^{'}}{N_r} \tag{4.6}$$

where N_r is number of repetitive event types defined in appliance candidate and $N_r^{'}$ is recognized number of repetitive event types in window candidate.

As for a combination sequence load such as furnace, S_{trd} can be decided based on its two sub-indices fixed pattern score S_{trdf} and repetitive pattern score S_{trdr} .

C. Trend signature score S_{trd}

As discussed in Section 3.4, power slope based scanning can effectively scan the window candidate and further determine the existences of trend signatures with respect to appliance candidate.

$$S_{trd} = \frac{N_t^{'}}{N_t} \tag{4.7}$$

where N_t is the number of trend signature types defined in appliance candidate and $N_t^{'}$ is the recognized number of trend signature types in window candidate.

D. Time signature score S_{time}

In the end, the moment of appearance of window candidate t is also compared with time signature defined for appliance candidate. As shown in Fig.6, the time signature of appliance candidate is defined as one or several hour ranges T such as {17-23},{6-8,11-13,16-18}. The judgment is made by equation (4.8).

$$S_{time} = \begin{cases} 1, t \in T \\ 0, t \notin T \end{cases} \tag{4.8}$$

4.2.3 Optimization of weights

As explained in Section 4.2.1, for different loads, their weights on event, event pattern, trend and time are different. In addition, their events also have different weights on real power, reactive power and harmonic contents. Rough ranges of weights are pre-defined based on the experience and understanding of the electrical/physical attributes of different loads, but their values should be further fine-tuned and verified through a specific program. Once the weights are determined, they do not need to be changed from one house to another. This advantage also implies no local training process is necessary when the proposed NILM is applied to a new system.

To better explain, the above process is compared to the "face recognition" feature supported by many digital cameras. The "face recognition" feature can automatically recognize human faces from different photo backgrounds. Similar to the identification of a certain type of load, face recognition also have internal weights on different facial attributes such as color/shape of eyes, nose, mouth and so on. However, when taking a picture, no matter the photographed person is standing with whatever objects such as animals, cars and other people, the weights of facial attributes do not need to be re-adjusted. The reason behind this is that the "face recognition" weights have been already optimized by the manufacturer with different photo backgrounds and thus the optimized weights are hardcoded into the cameras before they are sold to customers.

Similar to the "face recognition" process, for a specific load, this optimization program generates numerous pre-labeled event-windows based on the signatures of this load and also other types of loads that can mimic different "backgrounds" (other appliances). Besides, certain amount of signal noises (<10%) are added onto these windows. By doing this, the large number of pre-labeled testing event-windows includes not only the actual objective loads but also the other loads.

According to equation (4.9), different set of ω and γ values will result different judgments ---correction identification and false identification. If the most appropriate values of ω and γ for this specific load can be found, maximum correct identification rate can be achieved. The optimization goal can be described as below:

$$\begin{cases} \max\ f(\omega,\gamma) = \sum_{correct-match}(\omega^T x - \gamma) - \sum_{mis-match}(\omega^T x - \gamma) \\ s.t. 0 \le \omega \le 1 \\ \quad 0 \le \gamma \le 1 \end{cases} \tag{4.9}$$

The process of the above optimization is based on empirical initial ranges and method of exhaustive search. Since the accuracy needed for ω and γ is not extremely high, for each try, the step change is set as 0.02. By using this simple optimization method, optimal ω and γ values can be found for each load type.

On top of the above automatic process, another kind of optimization that is based on the local data collected from the target house can be further applied to fine-tune the values. This can be understood as a local calibration process. For a specific appliance, if its activities have been labeled or manually identified for a short period such as a single day, its weights can be further increased or decreased by values within a range of ±0.15 according to the labels, until a maximum identification rate is achieved. This testing procedure is convenient because for most of the cases, weights will not leave far apart from the results obtained from (4.9).

4.3 System implementation

To validate the proposed event-window based algorithm, practical system has been implemented and taken into field test. This section will discuss the key components of the system and the feasible solutions. The data flowchart of this NILM system is shown as below. Current and voltage signals at different phases are acquired; after data preprocessing, event matrix and trend scan table will be generated and fed into the identification module; finally, reports are generated and pushed up to the specific user interface for display.

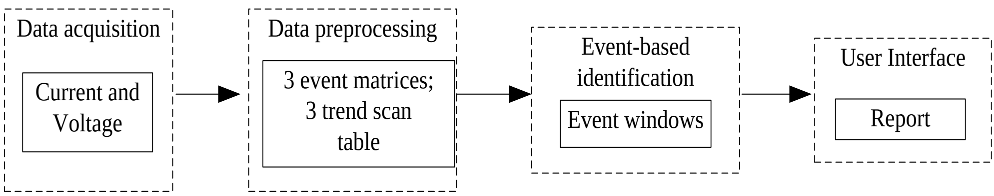


Figure 4.12: Data flow chart of the NILM system

4.3.1 Data acquisition

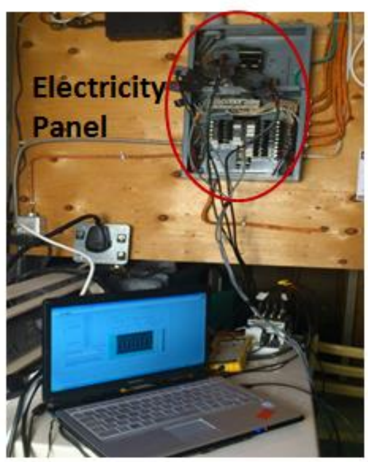


Figure 4.13: Data acquisition system at the meter-side

The setup of the data acquisition system is shown in the above picture: Electricity panel of a residential house is first opened; two current transducers are respectively connected to two phases. For each phase, the current transducer is clamped around its live wire; also a voltage transducer is connected between one of the live wires and the neutral wire; currents and voltages are being transferred to the National Instrument (NI)-DAQ system. The NI-DAQ system supports simultaneous inputs from multiple channels and has a high-resolution A/D converter in it; finally, digital signals of voltage and current are sent to the connected laptop via a USB port. Generally speaking, this data-acquisition system behaves like a smart meter.

The data is acquired at per second basis. For each second, a snapshot composed of 6 cycles of synchronous currents and voltage are recorded. For each cycle, 256 points are acquired. Example of a snapshot is shown in Figure 4.14.

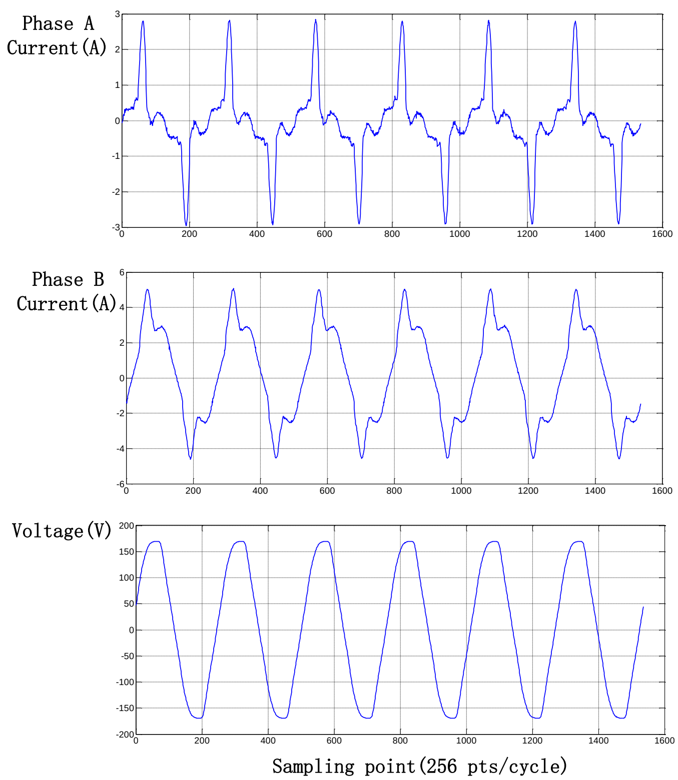


Figure 4.14: Example of a data snapshot

Besides, similar NI-DAQ based acquisition system is applied to acquire load signatures. The setup is shown as below: only one current transducer is used and connected to the live or neutral wire of the load supply cable. The voltage transducer is connected to a vacant electric outlet adjacent to the outlet the load is plugged into. The data format and sampling rate stay the same as the meter-side.

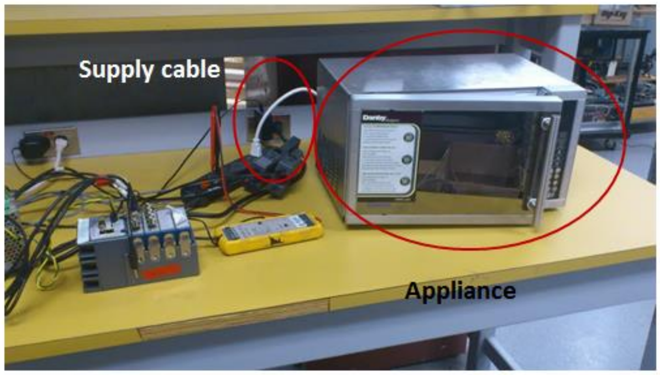


Figure 4.15: Data acquisition of load signatures

4.3.2 Data preprocessing

The amount of data acquired from the previous step is usually very large and not convenient for processing by the identification module. Hence, a data preprocessing step that can compress the data but keep the important information becomes necessary. The purpose of this step is to generate event matrix and trend scan table.

A. Generation of event matrix

Table 4-5 Example of event matrix

Event No.	Start sec	End sec	Rising/Falling	Real Power (W)	Reactive Power (Var)	Fundamental Current	…	3rd harmonic current	5th Harmonic Current	7th Harmonic current
1	6	7	2	28.7	0.9	0.3	…	0.1	0.0	0.1
2	38	40	…	…	…	…	…	…	…	…

An example of event matrix is shown in Table 4-5: event No., time points, rising or falling type and electric signatures of all the events are calculated and recorded for the data of a given period such as a day. As discussed in Chapter 2, events can be divided into 3 groups: phase-A events, phase-B events and phase A-B events or double phase events. Accordingly, 3 event matrices are generated respectively for the above three groups. Later on, in the identification step, different event matrices can be selected according to the phase signature of the objective load.

To improve the signature's accuracy, electric signatures are calculated based on the subtraction of the average of three consecutive points after a certain event and the average of three consecutive points before this event. For example, a data series have t_1-t_{10} ten data points and a load state change or an event happens at t_4-t_5. In this case, the electric signatures of this event are calculated using the following equations:

$$\begin{cases} \Delta P = ave(P_{t5}, P_{t6}, P_{t7}) - ave(P_{t2}, P_{t3}, P_{t4}) \\ \Delta Q = ave(Q_{t5}, Q_{t6}, Q_{t7}) - ave(Q_{t2}, Q_{t3}, Q_{t4}) \\ \Delta I^h = ave(I^h_{t5}, I^h_{t6}, I^h_{t7}) - ave(I^h_{t2}, I^h_{t3}, I^h_{t4}) \end{cases} \tag{4.10}$$

It should also be especially noted that since only 6-cycles snapshots are acquired within each second, P,Q and I^h from different snapshots have to be aligned up before subtraction. For alignment, voltage can be used as reference since the phase angle of voltage stays constant before and after a load state transition. Two approaches are proposed for the calculations of P,Q and I^h. The snapshot shown in Figure 4.14 is used as an example.

$$\begin{cases} I^1 = \left|I^1\right| \angle(\theta_I^1 - \theta_V^1) \\ I^h = \left|I^h\right| \angle(\theta_I^h - h \cdot \theta_V^1) \\ V^1 = \left|V^1\right| \angle \theta_V^1 \\ P = \left|I^1\right|\left|V^1\right| \cos(\theta_V^1 - \theta_I^1) \\ Q = \left|I^1\right|\left|V^1\right| \sin(\theta_V^1 - \theta_I^1) \end{cases} \tag{4.11}$$

Approach 1: Apply Fourier transform to the 6 cycles of two-channel current and one channel-voltage. Then use equation (4.11) to calculate P,Q and I^h.

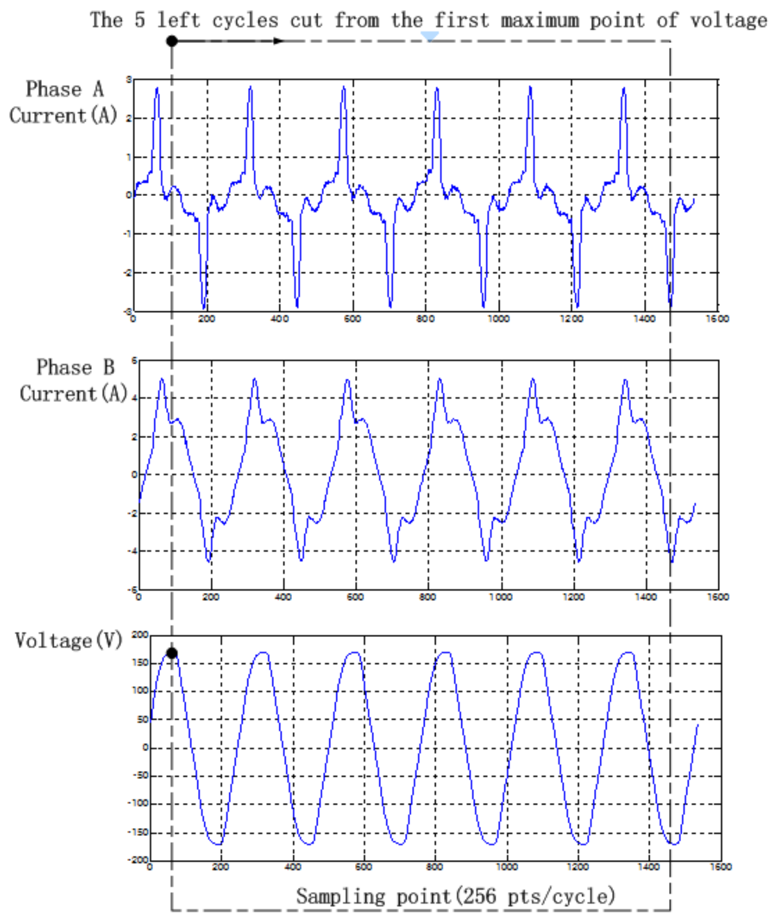


Figure 4.16: Approach 2 for P,Q and I^h calculation

Approach 2: Voltage is assumed to be completely sinusoidal. The first step is to locate the first maximum point of voltage from its 6 cycles. Then only the five left cycles after this point are considered for calculation. This is shown in Figure 4.16. By doing this, currents from different snapshots are aligned with respect to the voltage. In other words, θ_V^1 in equation (4.11) is set to 0 degree. Therefore, the above equation can be simplified as (4.12) :

Compared with Approach 1, Approach 2 is much faster since there is only one time Fourier transform on 5-cycle currents. However, it is less accurate since not all cycles are fully utilized and the assumption that voltage has no harmonic content is not always true.

$$\begin{cases} I^1 = \left|I^1\right| \angle \theta_I^1 \\ I^h = \left|I^h\right| \angle \theta_I^h \\ P = \left|I^1\right| V \cos(\theta_I^1) \\ Q = \left|I^1\right| V \sin(\theta_I^1) \end{cases} \tag{4.12}$$

B. Generation of trend scan table

The above 3-phase event matrices have included event, event pattern, time/duration and phase connection signatures. However, as discussed in Chapter 3, trend signatures should also be fed into load identification module for judgment. The other task in the data preprocessing stage is to generate a trend scan table shown as below:

Table 4-6 Example of trend matrix

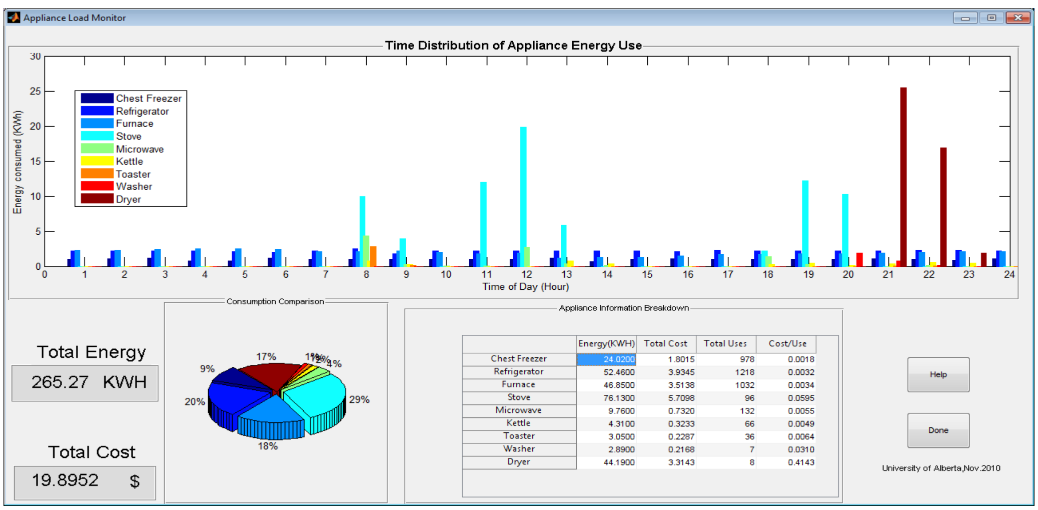
Window No.	Rising spike	Falling spike	Pulses	Fluctuation	Quick vibrate	Gradual Falling	Flat
3-4	T	F	F	T	F	T	F
4-5	F	F	F	F	F	F	T
…-…	…	…	…	…	…	…	…

As can be seen from Table 4-6, for a given window, different trend slope signatures are recognized according to the slope characteristics listed in Table 3-4. The above process is conducted throughout all types of phase connections for the studied period. Finally, 3 trend scan tables that each represents one type of phase connection can be generated.

4.3.3 User interface

After identification decisions are making from the event-based identification module, activity report along with energy estimation results are generated and displayed on a graphic interface shown as below.

The results are updated every half an hour and displayed in the interface shown in Figure 4.17. As can be seen, appliance electricity consumption information is formatted into the table and charts. The table summarizes the total energy counted from a certain date and converted expenses with respect to local electricity rates (say,7.4¢/KWH). The pie chart presents the percentage composition of individual appliance so users can be aware of the significance of reducing a certain appliance's consumption. Finally, from the time distribution of energy use chart, user can understand his identified appliance activities statistically with respect to hours. This information is quite essential for residential house owners to adopt proper demand response strategies such as load shifting according to utility's TOU rates. More discussion can be found in Chapter 5.



	Energy(KWH)	Total Cost	Total Uses	Cost/Use
Chest Freezer	24.0200	1.8015	978	0.0018
Refrigerator	52.4600	3.9345	1218	0.0032
Furnace	46.8500	3.5138	1032	0.0034
Stove	76.1300	5.7098	96	0.0595
Microwave	9.7600	0.7320	132	0.0055
Kettle	4.3100	0.3233	66	0.0049
Toaster	3.0500	0.2287	36	0.0064
Washer	2.8900	0.2168	7	0.0310
Dryer	44.1900	3.3143	8	0.4143

Figure 4.17: Appliance energy decomposer software

4.4 Verification using real house data

The above algorithms and devices were tested in several real residential houses in Edmonton, Canada for several weeks. A laptop based data acquisition system was hooked to the electricity panel and behaved like a smart meter. A Zig-bee transceiver was connected to its USB port to bridge the communication with the appliance register. After registration was finished, a computer program based on the proposed algorithms was launched and started to process. Interested appliance activities were identified. To verify the identification rates, appliance activities under track were either recorded manually or labeled through human experts' inference for comparison.

In addition, verification was conducted based on a public dataset provided by MIT [95]. In this dataset, several appliances were labeled using extra data-logging devices that were directly connected with objective appliances. The results and discussions are respectively shown in the following sections.

4.4.1 Verification based on House #1

Table 4-7 Identification rate accuracy for house #1(7 days)

Appliance Name	Actual operation times	Correctly identification times	False identification times	Identification accuracy(%)
Fridge	312	253	0	81
Microwave	54	50	0	93
Washer	5	5	0	100
Cooktop	1	1	0	100
Stove elements (low power)	16	16	0	100
Stove elements (high power)	18	18	0	100
Kettle	20	17	0	85
Dryer	5	5	0	100
Heater	248	223	0	90
Waffle iron	13	13	1	92
Average	---	---	---	94

The results for house #1 are listed in Table 4-7 (this house has no furnace). The definitions of all the table columns are explained as below:

- Actual operation times (AOT): the actual observed and inferred operation times by manual identification.

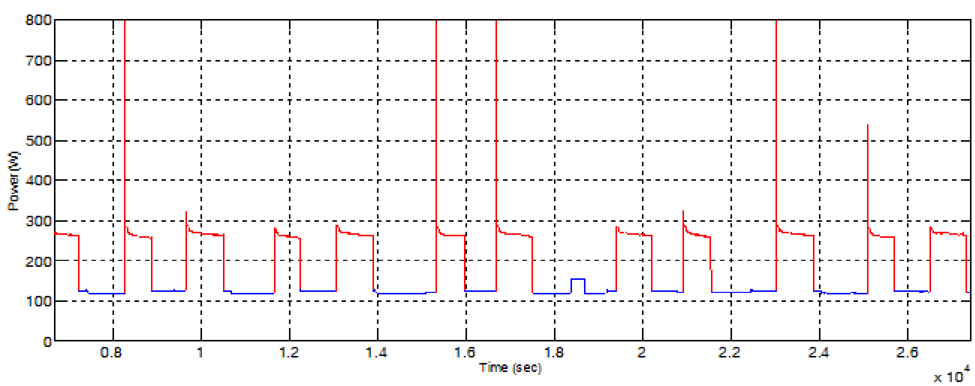
- Correctly identification times (CIT): the correctly detected times by the algorithm that matches the actual operation times. It should be noted, for some action-intensive appliances such as stove, washer and dryer, if their detected events are close to each other (say shorter than 30 mins), there should be combined together as one event. It is because those appliances usually have repetitive events that actually come from single time operation.

- False identification times (FIT): Times misidentified by the algorithm which cannot match the actual window observed or inferred manually.
- Identification accuracy: Defined as

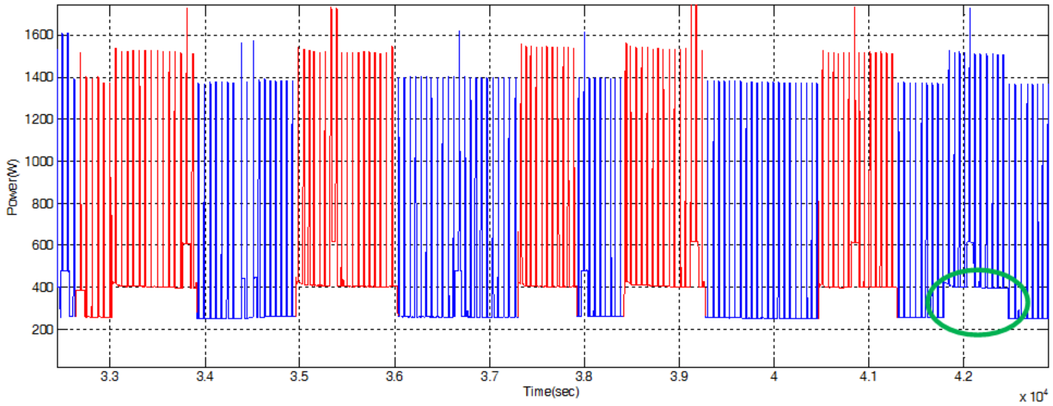
$$Accuracy = \frac{CIT - FIT}{AOT} \times 100\% \tag{4.13}$$

Therefore, using equation (4.13), false identified time has a negative impact on identification accuracy and thus the index can be more objective. In order to more clearly present the results and discussions, examples of identified event-windows are automatically marked in red by the algorithm and some failures are circled in green. These examples are shown as below:

1. Fridge (Identification accuracy: 81%)



(a) Success identification

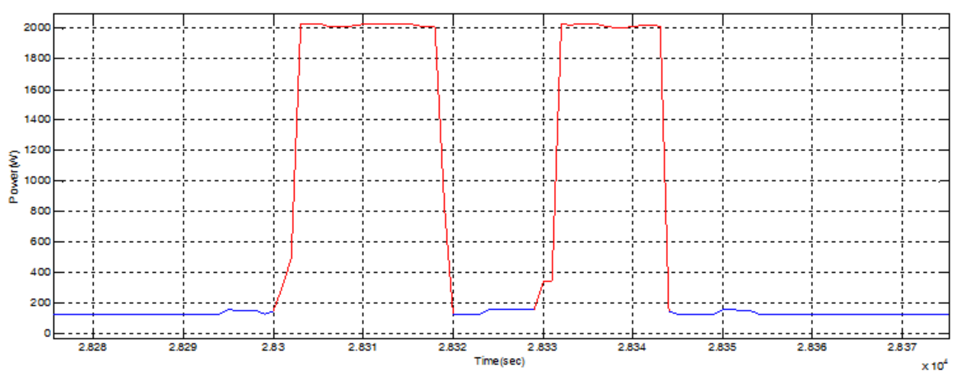


(b) Failure (green circle)

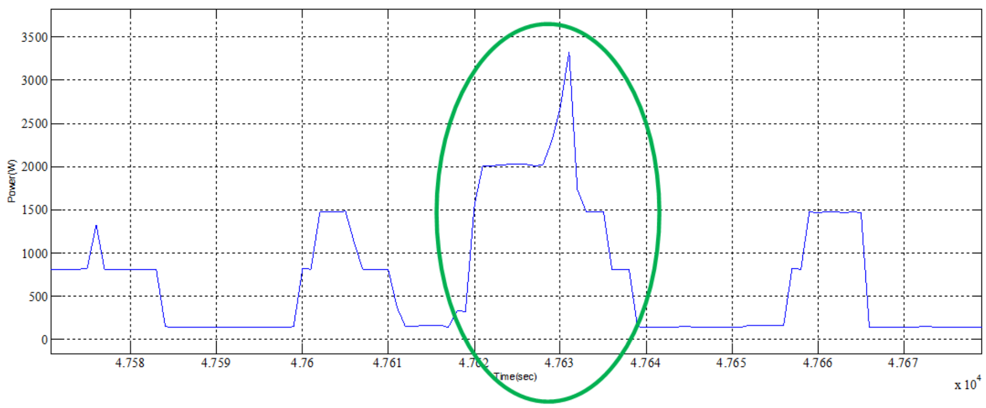
Figure 4.18: Examples of identification for fridge

Most events of fridge can be easily identified. Some failures are due to the overlapping with pulse-based appliance such as a stove. One example is shown as the green circle part in Figure 4.18 (b). Its ON event is ruined by the "needles" of stove which cannot be detected by the event detector any more. However based on human inference, this could highly possible be a real fridge event. As shown in Figure 4.18 (b), under this circumstance, many fridges can still be correctly identified, which shows the robustness of the proposed algorithm.

2. Microwave (Identification accuracy: 93 %)



(a) Success identification



(b) Failure

Figure 4.19:Examples of identification for microwave

As can be seen from Figure 4.19, (a) shows the example of successful identifications; (b) shows a failed case. The reason for failure is that the OFF event of this microwave operation overlaps with another appliance.

3. Washer (Identification accuracy: 100%)

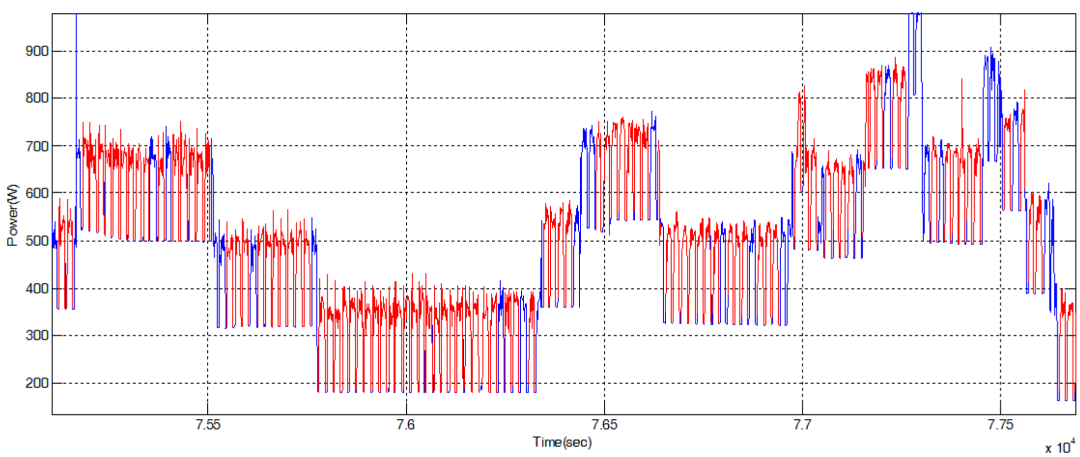


Figure 4.20: Examples of identification for washer

4. Dryer (Identification accuracy: 100%)

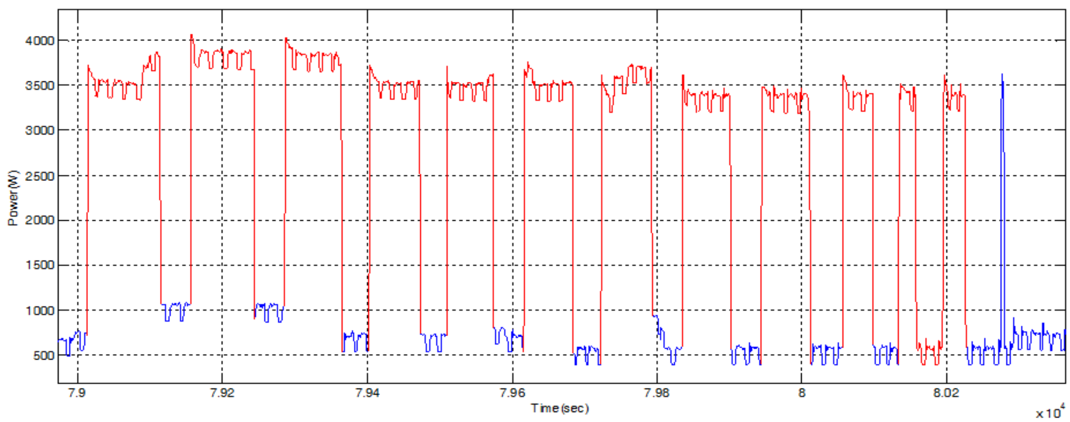


Figure 4.21: Examples of identification for dryer

5. Stove elements using low power (Identification accuracy: 100%)

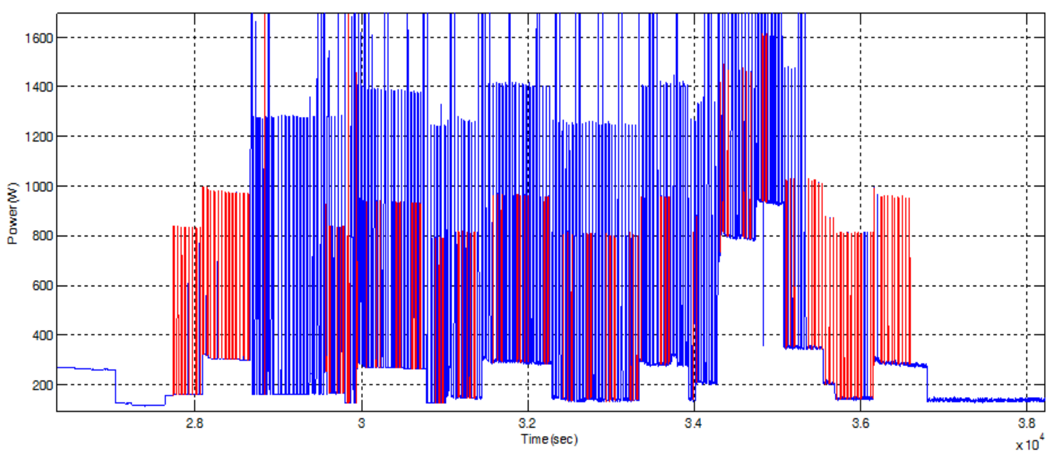


Figure 4.22: Examples of identification for stove elements using low power

6. Stove elements using high power (Identification accuracy: 100%)

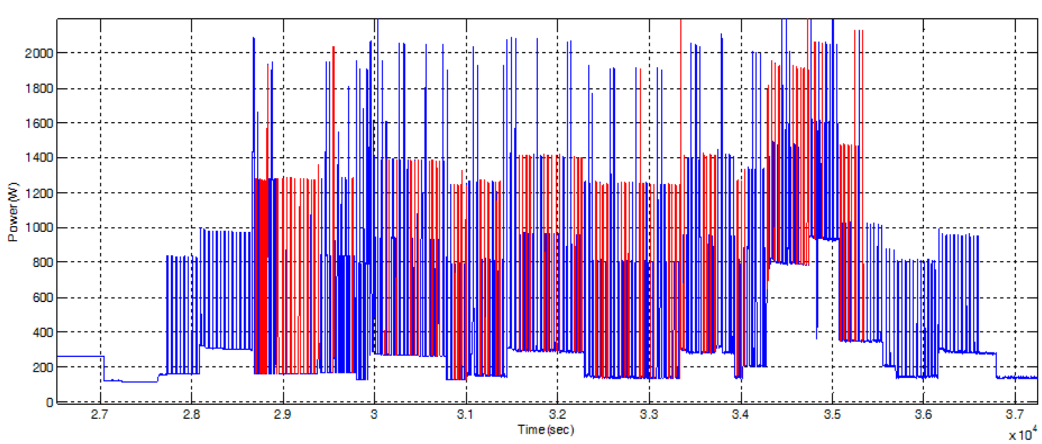


Figure 4.23: Examples of identification for stove elements using high power

7. Cooktop (Identification accuracy: 100%)

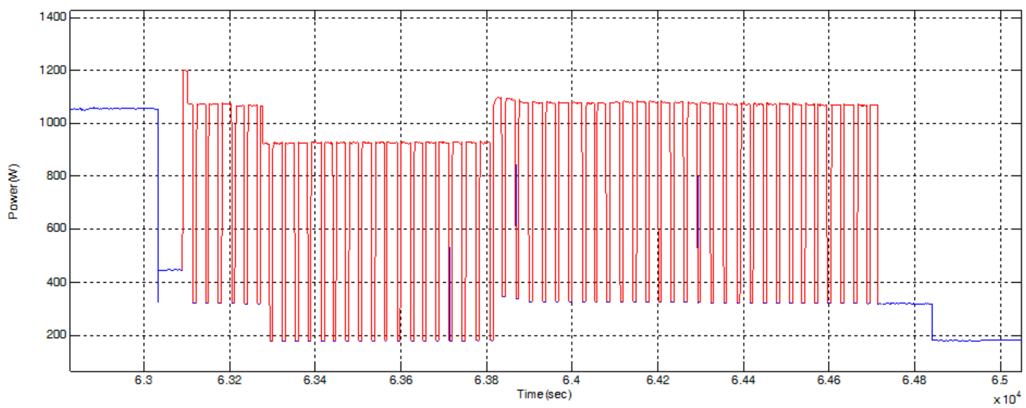
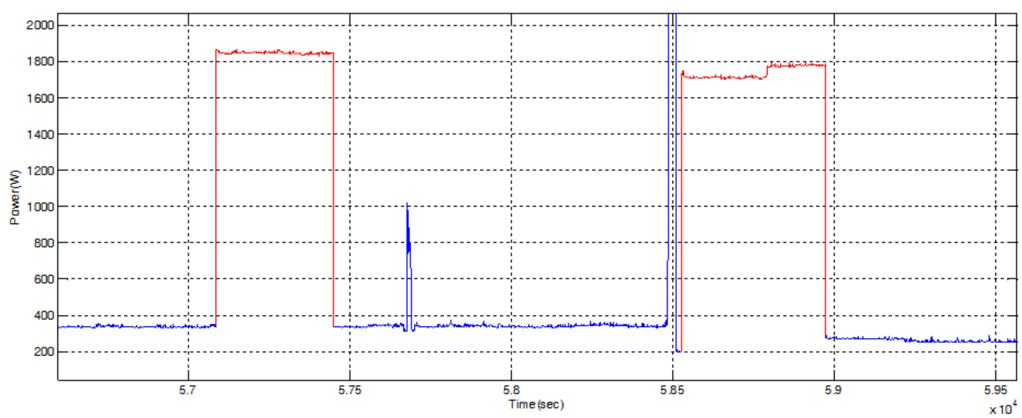
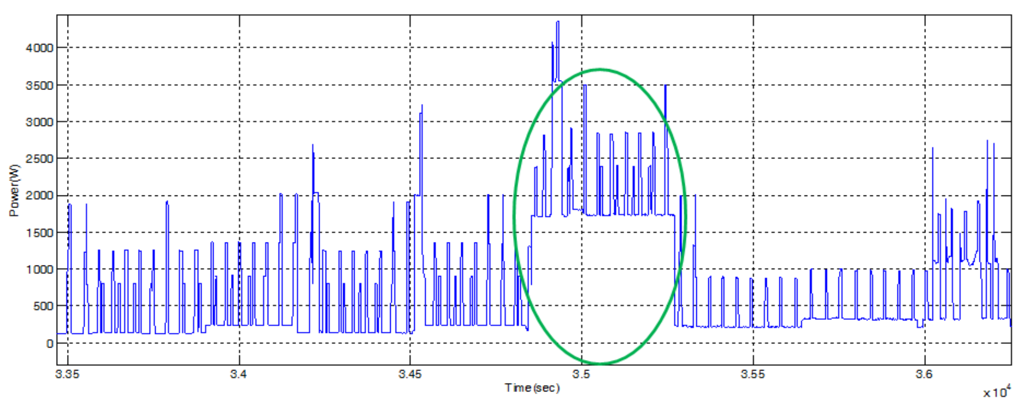


Figure 4.24: Examples of identification for coffee maker

8. Kettle (Identification accuracy: 85%)



(a) Success identification



(b) Failure

Figure 4.25: Examples of identification for kettle

Similar to the fridge, sometimes when the events of kettle overlap with other frequent appliances, identification failure may occur due to event detection failure.

9. Heater (Identification accuracy: 90%)

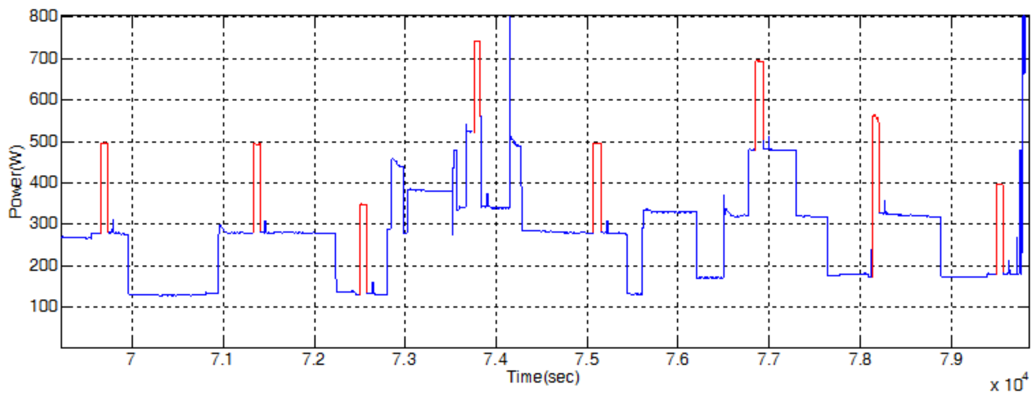


Figure 4.26: Examples of identification for heater

10. Waffle iron (Identification accuracy: 92%)

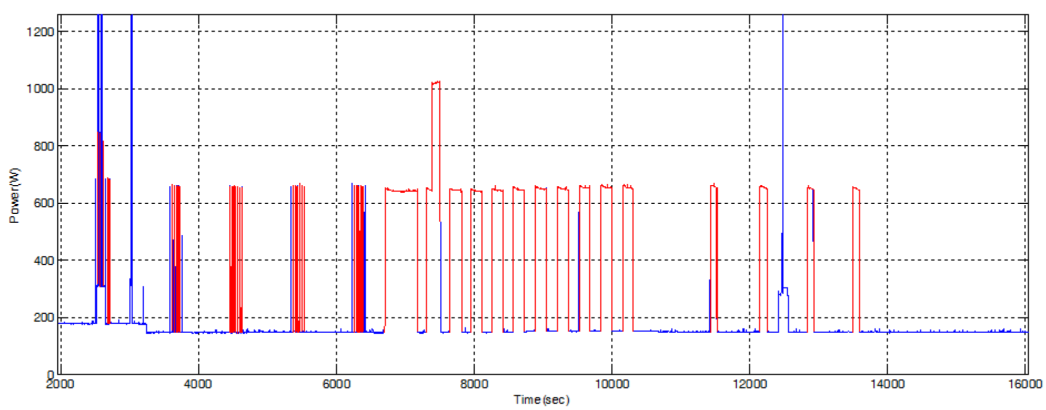


Figure 4.27: Examples of identification for waffle iron

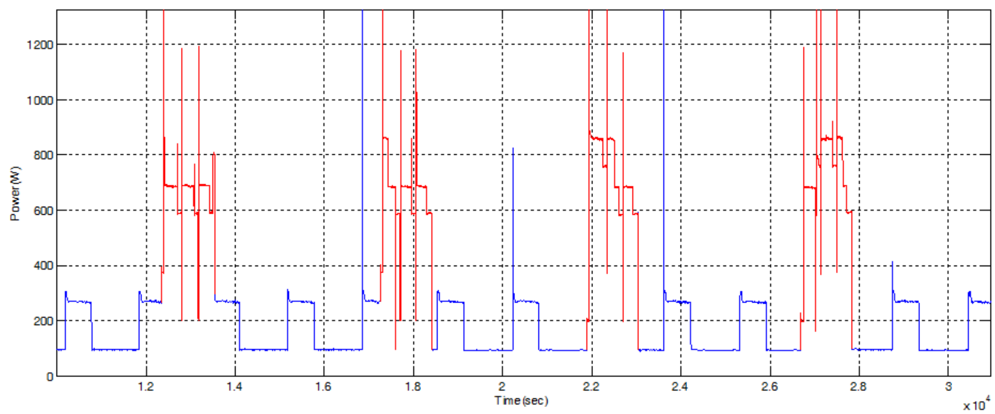
4.4.2 Verification based on House #2

Table 4-8 Identification rate accuracy for house #2 (8 days)

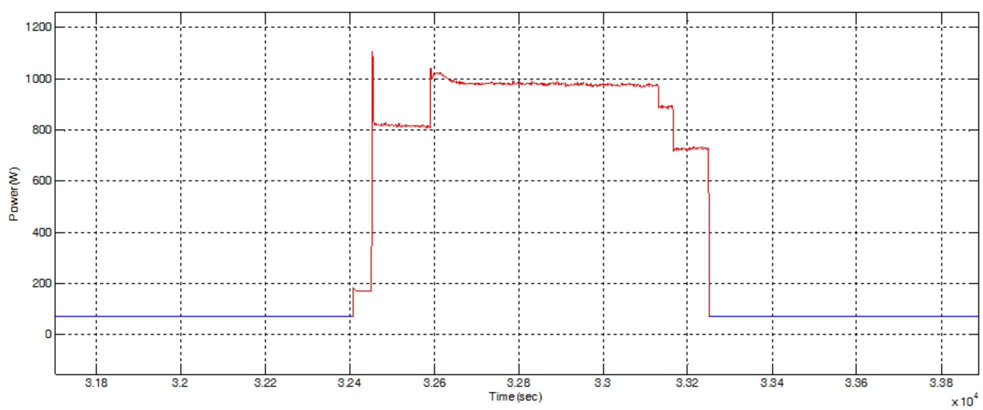
Appliance Name	Actual operation times	Correctly identified operation times	False identified operation times	Identification accuracy(%)
Freezer	576	523	0	91
Fridge	417	371	0	89
Furnace	28	27	0	96

Microwave	72	71	0	99
Kettle	14	14	1	93
Washer	4	4	0	100
Dryer	5	5	0	100
Stove elements (low power)	8	8	0	100
Stove elements (high power)	5	5	0	100
Average	---	---	---	96

The results are shown in Table 4-8. As a multi-state appliance, the identification example of furnace is shown as below:



(a) Zommed-out view of furnace identfication



(b) Zommed-in view of furnace identification

Figure 4.28: Examples of identification for furnace

As can be seen, even if the furnace's operations sometimes overlap with the fridge's operations, identification is almost not affected. This is because fridge is different from stove---although it has frequent operations throughout a day, it does not generate high-frequency pulses and it has a much lower possibility to cause an event overlap problem.

Also, it is found the proposed algorithm can withstand certain level of signal noises such as the figure shown as below. As can be seen, even if the steady state of two later fridge operations have significant noises, they can still be correctly identified as long as their events ON/OFF are not heavily polluted.

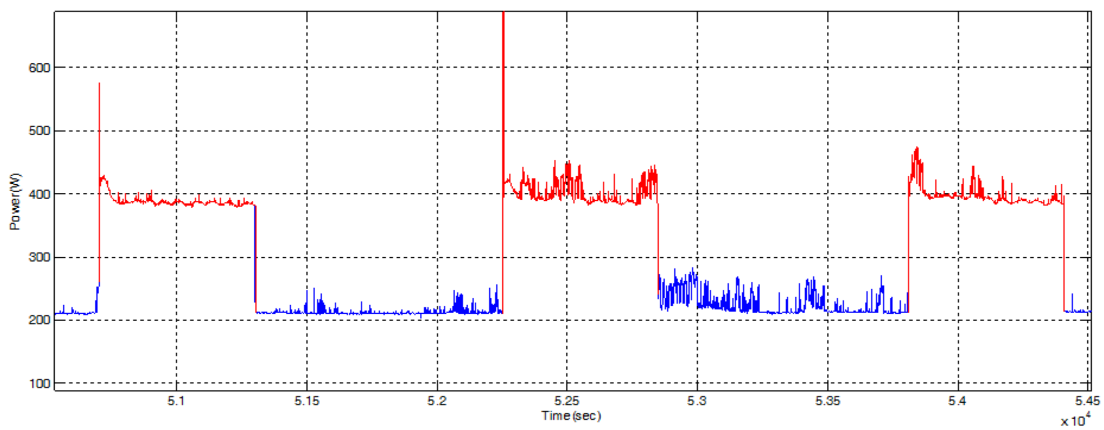


Figure 4.29: Examples of fridge identification under noisy condition

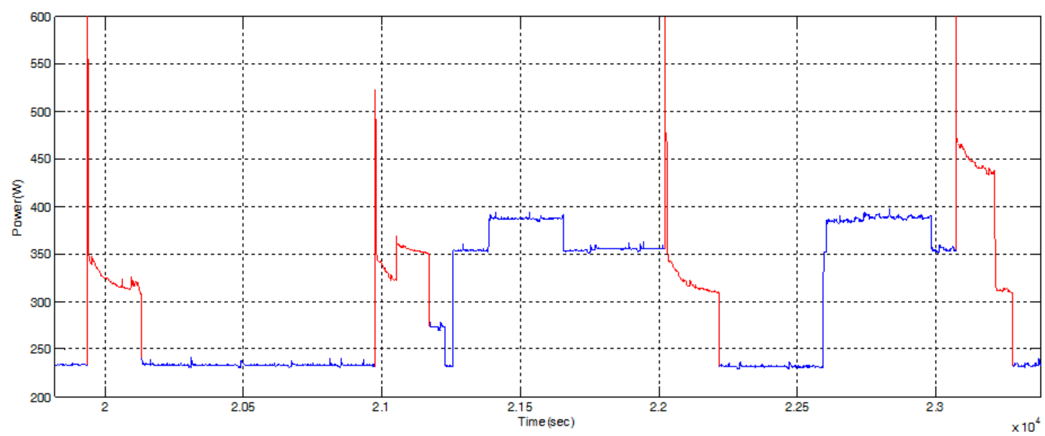


Figure 4.30: Examples of freezer identification when operations overlap with other appliances

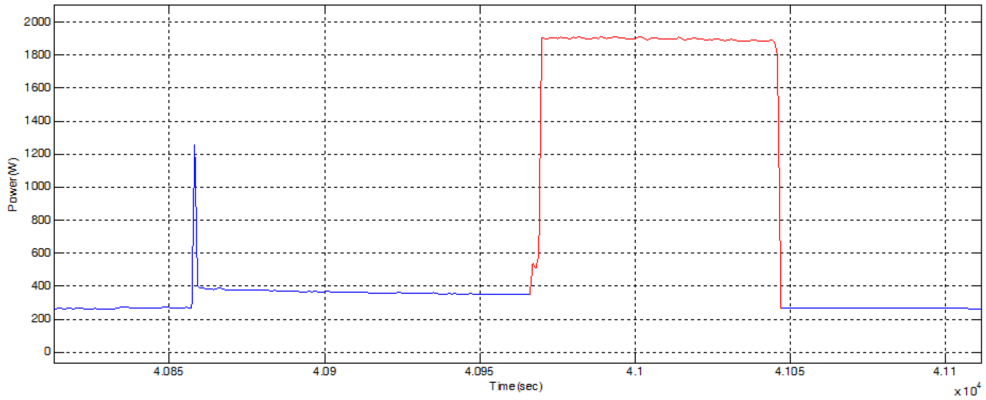


Figure 4.31: Example of microwave identification when it overlaps with fridge

In addition, it is also found that the proposed algorithm is resistant to the overlap of different types of appliances. In Figure 4.30, the first freezer's operation does not overlap with other appliances. However, the three freezer's operations behind it all overlap with some other appliances. But they are still correctly identified just like the first one. Another example is shown in Figure 4.31. As can be seen, the microwave's operation can still be identified even if it overlaps with the fridge's operation.

4.4.3 Verification based on House #3

Table 4-9 Identification rate accuracy for house #3 (7 days)

Appliance Name	Actual operation times	Correctly identification times	False identification times	Identification accuracy(%)
Fridge	25	22	0	88
Furnace	60	55	0	92
Microwave	40	36	1	90
Washer	15	15	0	100
Dryer	8	8	0	100
Stove	14	13	0	93
TV	8	7	1	75
Kettle	23	23	1	96
Average	---	---	---	92

The results are shown in Table 4-9. It is found TV has a lower identification rate. This is because TV has a flexible duration. The duration can be very long such as hours but can also be shorter than half an hour. The identification rate is therefore reduced since many more window candidates will be included which lead to potential interference.

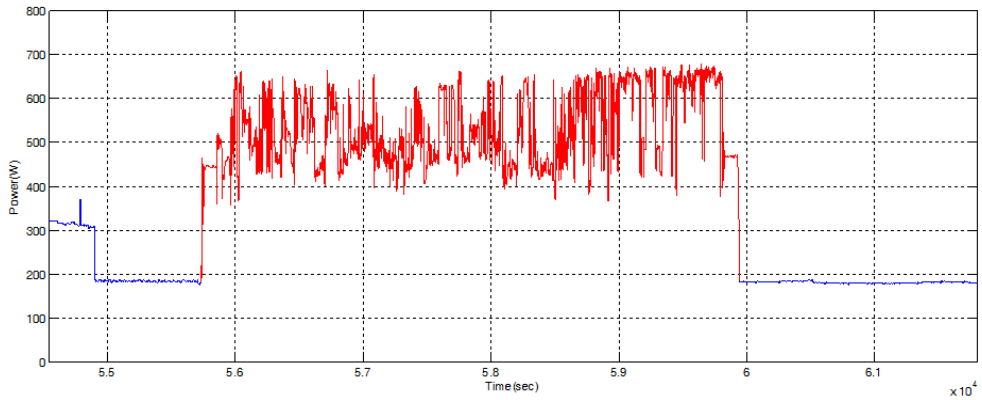


Figure 4.32: Examples of identification for TV

4.4.4 *Verification based on public dataset*

REDD, a public dataset available for NILM research was released by MIT in 2011 [95]. The data was acquired from the greater Boston area in US. The dataset can be used to validate the NILM algorithm since several appliances have been labeled using extra data-logging devices that are directly connected to the target appliances. The No.3 house from the dataset is selected for the validation purpose. In general, the dataset contains 1427284 seconds or roughly 16 days and the dataset is not continuous. Many appliances are pre-labeled. For some unknown reason, the washer is not pre-labeled by MIT, however, it is labeled by us through realistic inference with respect to dryer activities. Several major appliances have been identified and the results have been listed in Table 4-10.

Table 4-10 Identification rate accuracy for house #4 (7 days)

Appliance Name	Actual operation times	Correctly identification times	False identification times	Identification accuracy(%)
Fridge	661	628	5	94
Furnace	34	32	1	91
Microwave	95	95	4	96
Washer	7	6	0	86
Dryer	13	13	0	100
Dishwasher	5	5	0	100
Electronics	10	10	0	100
Average	---	---	---	95

4.4.5 Observations and findings

Overall, the algorithm has excellent performance in terms of identification rate. It is capable for the following cases:

- Dealing with all types of appliances including single-state appliances (microwave, kettle and waffle iron etc.), continuous-varying appliances (fridge and freezer) and multi-state appliances (furnace, washer and dryer). The average identification rates in the four houses are all above 90%;
- When the load window duration is not changing significantly or not too long;
- Even when the operations of different appliances overlap;
- Even when there is a certain level of signal noise.

However, the performance can be affected on the following cases:

- If the appliance has a very long and flexible duration, it can be troublesome such as the TV in house #3;

- If a certain appliance can generate high-frequency pulses, it might affect the identification of other appliances because the pulses may cause the event overlap problems.

4.5 Comparative studies with neural networks based method

4.5.1 Implementation of neural networks based method

This chapter also presents a detailed comparison between the proposed solution and previous signal combination based solutions such as the one discussed in [69]-[70].According to [69], a two-layer feed-forward network is adopted here for comparison.

Firstly, specific appliances were measured in the lab and their harmonic signatures were collected. Not like the proposed approach, no process related signatures is considered by neural networks. Harmonic contents of aggregated signal are used as input layer while appliance composition list as output layer. Since both magnitude and phase of a certain harmonic order are considered, the input layer has 16 nodes of up to 15th harmonic content (only odd ones). Hidden nodes are set to be 20. As shown in Figure 4.2, according to [69], numerous training sets are generated mathematically by adding up harmonic contents (waveforms) of designated individual appliances. Also, to make training more reliable, a less than 10% deviation is added to original magnitude as noises.

For test stage, a bottom-up based aggregating program turns on/off each load according to a certain operation probability. The aggregated meter-side signal is formed this way. Then both of the two approaches were tested to decode the overall meter signal and their performances are discussed as below.

4.5.2 Simulation based verification

Table 4-11 Comparison for only ON/OFF type loads

Loads	Identification accuracy(%)	
	NN based approach	Proposed approach
Microwave	97.9	99.9
Monitor	98.3	98.3
TV	99.2	98.1
Vacuum	97.6	98.6
Monitor	99.9	98.5
Incandescent light bulb	98.9	98.5
Fluorescent light bulb	99.0	99.2

Table 4-12 Comparison with complex loads

Loads	Identification accuracy(%)	
	NN based approach	Proposed approach
Microwave	92.3	99.9
Monitor	93.6	94.9
TV	84.4	94.2
Vacuum	85.0	96.3
Monitor	79.8	97.7
Incandescent light bulb	97.5	98.5
Fluorescent light bulb	95.6	99.2
Fridge	63.7	97.9
Freezer	68.5	95.3
Washer	73.4	97.1
Furnace	57.2	98.4

Comparison is firstly conducted when there are only single-state type loads. This is because single-state type loads only have a steady-state harmonic content. The results are shown in Table 4-11.

As can be seen, for a system composed of only single-state type loads, the proposed approach has a performance similar to NN based approach. This is because there is no change in each appliance's operation process. However, results are heavily changed when complex loads are brought in.

As can be seen from Table 4-12, NN based approach is significantly affected by the introduction of multi-stage loads (furnace and washer) and continuous varying loads (fridge and freezer). This drawback is actually discussed in [69] due to the lack of a steady-state harmonic content in those appliances. Their harmonic contents can vary tremendously with time. Sometimes, harmonic contents of different operational stages of the same load are not even comparable such as in furnace. To cope with this problem, NN based approach has to average the harmonic contents and use the average value for training. This will introduce not only large error to those complex loads themselves but also to those single-state loads if they are turned on at the same time. For example, for a given point, if the aggregated waveform is composed of fridge and microwave, identification of microwave may fail due to the error from fridge. In contrast, the proposed approach captures event window and utilizes process signatures to identify. In theory, the more complex the process is, the more unique its window can be and the easier it can be identified. This is the reason proposed approach has a much better performance. In the meanwhile single-state appliances will not be affected by complicated appliances since they have different events.

Table 4-13 Comparison when stove is not trained or registered.

Loads	Identification accuracy (%)	
	NN based approach	Proposed approach
Microwave	78.0	94.5
Monitor	77.8	94.3
TV	76.6	94.2
Vacuum	65.2	96.1
Monitor	95.8	95.8
Incandescent light	64.1	95.1

bulb		
Fluorescent light bulb	38.6	95.8
Fridge	51.3	96.1
Freezer	56.3	90.6
Washer	45.4	92.7
Furnace	44.3	95.6
Stove	---	---

Another obvious advantage of proposed approach is it only identifies loads that users are interested in and willing to register. In contrast, NN based approach's training process has to cover all major appliances. Also, once user purchases another heavy load such as a stove, the accuracy of identification will become not reliable at all. This is because the neural network model needs to be changed due to the newly added element. As shown in Table 4-13, any aggregated signal that has stove on at the same time will become unidentifiable (this is especially severe for other kitchen appliances). However, stove hardly has any impact to the identification of registered loads using proposed approach because the proposed approach is based on searching relevant events of registered appliances only. Those non-relevant events from stove will be excluded from the window candidates of the registered loads.

4.5.3 Observations and findings

To summarize, compared to the signal-combination based approach, the proposed approach has the following obvious advantages:

- Window based signatures make identification of complex loads possible. In contrast, signal-combination based approaches such as NN cannot efficiently identify those loads;
- Composition of appliances is not only judged by an independent point of meter-side signal but also events before and after this point. Hence, the association of load states is much more strengthened. A load's OFF

event can only be confirmed if its ON event is found within a time window;

- Training process does not need to cover all major appliances any more. Users only need to register their interested loads they want to track down;

- Nearly no additional effort if load inventory is partially changed.

4.6 Summary

Both the theoretical and practical sides of event-window based load identification method have been discussed in this chapter. The overview and classification of existing NILM algorithms, load identification procedure, signature score calculation, weight optimization, system implementation, extensive real house based validations and comparative studies with previous signal-combination based approach were all addressed in details.

Generally, the proposed event-window based algorithm has the following advantages:

- It can effectively deal with complex loads including continuous-varying and multi-state loads;

- It does not require a local training process since the weights of different signatures for different loads can be optimized in the lab before taking to the field. Only simple optimization process is needed;

- It can resist the change of load inventory;

- The average identification rates in the four tested houses are all above 90%. The overall accuracy is very satisfactory;

- It can resist a certain level of signal noises;

- It can deal with the operation overlaps of different loads.

However, the proposed algorithm does have some disadvantages for the following scenarios:

- If the appliance has a very long and flexible duration, the number of window candidates will increase dramatically which could lead to larger identification error.

- Sometimes load events can be ruined by the events of other loads, especially when the other loads generate continuous high-frequency pulses such as stove and washer. If the events are ruined, they cannot be correctly detected by the event detector and this will further also lead to identification error.

Overall speaking, the proposed algorithm makes a good balance between being effective and being practical.

Chapter 5

Non-intrusive Signature Extraction for Major Residential Loads

This chapter presents a technique to extract load signatures non-intrusively by using the smart meter data. Load signature extraction is different from load activity identification. It is a new and important problem to solve for non-intrusive load monitoring (NILM) applications.

For a target appliance whose signatures are to be extracted, the proposed technique first selects the candidate events that are likely to be associated with the appliance by using generic signatures and an event filtration step. It then applies a clustering algorithm to identify the authentic events of this appliance. In the third step, the operation cycles of appliances are estimated using association algorithms. Finally, the electric signatures are extracted from these operation cycles.

The results can have various applications. One is to create signature databases for the NILM applications. Another is for load condition monitoring. Validation results conducted on the data collected from three actual houses and a laboratory experiment have shown that the proposed method is a promising solution to the problem of non-intrusive load signature collection. Some contents in this chapter have been submitted as publication in [105].

5.1 Overview

5.1.1 Review of existing intrusive signature extraction methods

To perform load decomposition, all NILM techniques must rely on the unique signatures of individual loads. In order to obtain these signatures, the existing

NILM researches generally require measurement steps [55]-[83], which are supervised and intrusive.

For example, in [69], signatures such as harmonic contents of all appliances in the house have to be collected in advance. Only after this, training sets can be generated based on the combinations of above signature collections. In [81], it is proposed that three power meters are used to acquire the signatures of a single load at a time. The hardware devices are shown as below:

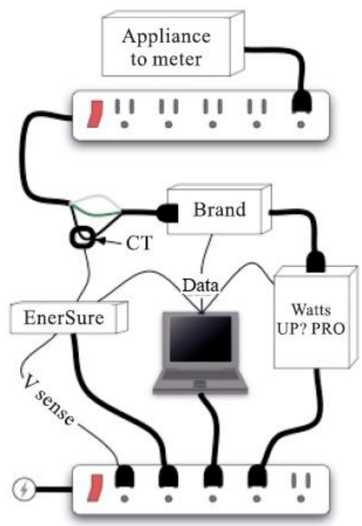


Figure 5.1: Three power meters based signature extraction

In [82] and [50], electric/magnetic field sensor based event detectors are developed to label the events of a single appliance. Then electric signatures of the appliance can be extracted by comparing its labeled events with the metering side aggregated data. One example is shown in Figure 5.2 [50] to explain its working principle: the current of an overhead fan changes when its operation state changes. And the current will cause the changes of both electric field and further magnetic field. The EMF detector is a piece of PCB board composed of signal amplifier circuits. It can process both the electric/magnetic field signals in real-time. Based on the distinction between the electric and magnetic fields, events of the overhead fan can be labeled such as b and c. When the time points of b and c are located on the aggregated meter-side signal, electric signatures of the overhead fan can be extracted for NILM applications.

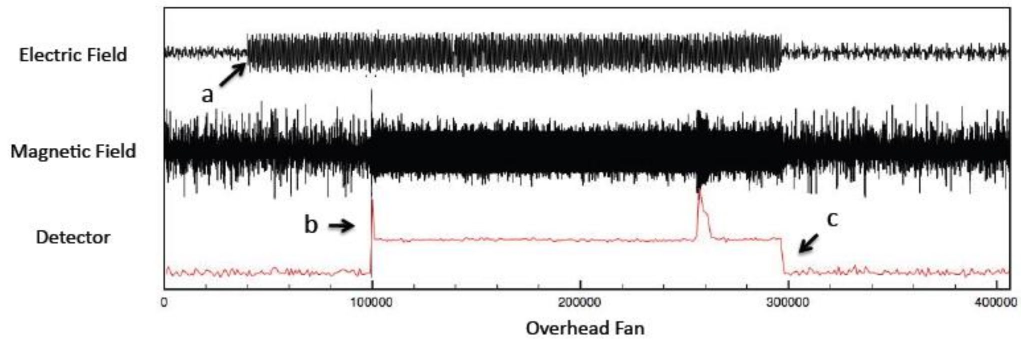


Figure 5.2: Relations of magnetic field, electric field and the EMF event detector

In [83], a smart phone based signature extraction system is developed. It requires manual confirmations from human experts or house residents to label the events of appliances. The process is shown in Figure 5.3.

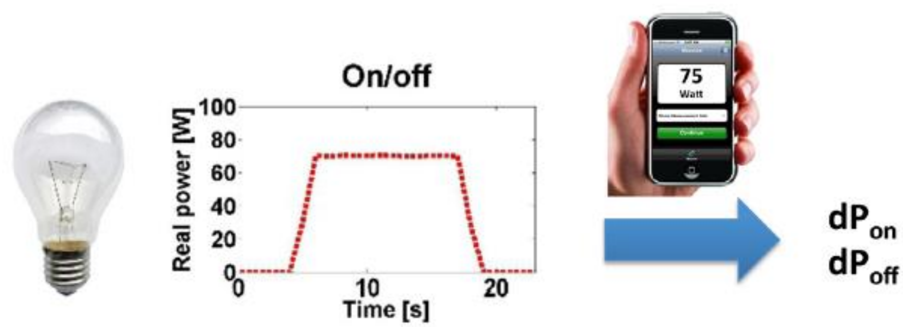


Figure 5.3: Smart phone and human confirmation based signature extraction system

5.1.2 Proposed intrusive event-window signature extraction system

As explained in Chapter 3 and Chapter 4, an event-window consists of various signatures. In order to acquire these signatures, we firstly propose to create a small signature database tailed for each home utilizing an intrusive piece of hardware called appliance register.

As shown in Figure 5.4, an appliance register is a device installed between the appliance to be registered and the electric outlet the appliance is supposed to be plugged in. The device contains a current sensor and a wireless transmitter. Once a current change is detected (an event), the device will send a signal to the smart meter (or the device which does appliance identification).Smart meter does two things: capture the event window of this appliance and determine the signatures of event window.

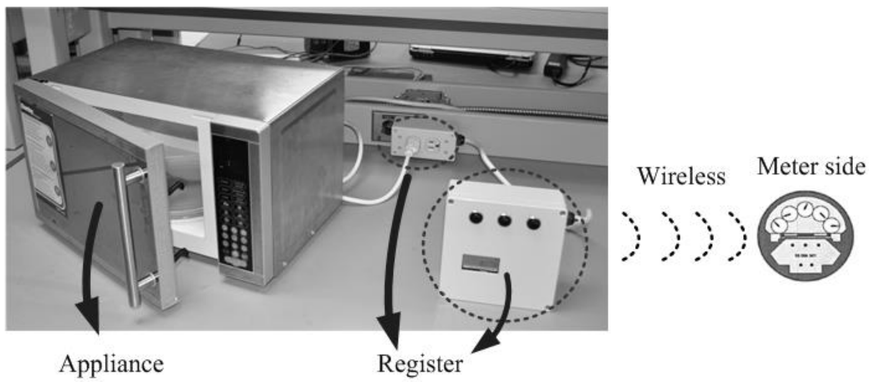


Figure 5.4: Intrusive event-window signature extraction system

Firstly, phase signature can be determined by the smart meter. Then captured event window will be scanned through and all events associated with the appliance (labeled by the register device) are picked out. Event signatures can be directly extracted. Event pattern signature can be determined based on the appearance number of event types. Trend signature can be detected based on slope modes explained in Section 3.4. In the end, time/duration signatures are automatically selected after the appliance is named by users. After waiting for one or two operating cycles of the appliance, all signatures of the appliance will have been collected. The register is then moved to another appliance interested for registration. This approach has another advantage in term of privacy: the customer can control which appliances are to be registered for identification.

5.1.3 Proposed non-intrusive signature extraction method

In fact, the above signature extraction processes discussed in 5.3.1 and 5.1.2 have the following common disadvantages:

1. They are intrusive methods. They all require additional hardware devices to either directly measure load signatures or label load events. Fundamentally, this

intrusive step is opposite to the basic principle and intention of non-intrusive load monitoring research. In other words, it makes NILM not purely non-intrusive.

2. They are supervisory methods. No matter it is based on measurement or labeling, human efforts are required to be involved. This is fairly inconvenient and less practical for ordinary household residents to perform. This disadvantage can significantly affect the wide application of NILM.

3. They are hardware based methods. Extra cost will be added into the solutions. Unfortunately, so far, there is no other pure-software based approach on signature extraction for NILM.

It is clear that the above intrusive ways of signature collection are not desirable by ordinary house owners. Therefore, there is a need for methods that can collect the appliance signatures non-intrusively. If successful, unique signatures that are specific to an appliance in a particular home can be extracted and archived. To solve the above drawbacks, this chapter proposes a novel unsupervised non-intrusive signatures extraction (NISE) approach which does not require any measurement or direct input from users and is purely software based. Given the meter-side data of a certain days, specific major appliance events can be located, associated and then their complete operation cycles can be reconstructed automatically. For the proposed NISE method in this chapter, the key is to intelligently study these events and establish knowledge of appliances automatically.

As shown in Figure 5.5, firstly, all appliance events are captured through the event detection method discussed in Chapter 2. Secondly, event filtration is adopted to identify the suspect events of interested appliances. Thirdly, event clustering is utilized as a tool to pick out the appliance’s authentic events from its suspect events. Fourthly, based on the authentic events, event association algorithm is further applied to reconstruct complete operation cycles of appliances. Finally, both electric and event pattern signatures of appliances can be extracted from the reconstructed cycles to fit any existing signal combination based NILM approach [69]-[78] and event-window based NILM approach [104]. In the following sections of chapter, event filtration, clustering and association are discussed in detail respectively. Also, the proposed approach is validated using 3 real houses and laboratory data. The accuracy of signatures extracted is compared and analyzed. Conclusions are presented in the end.

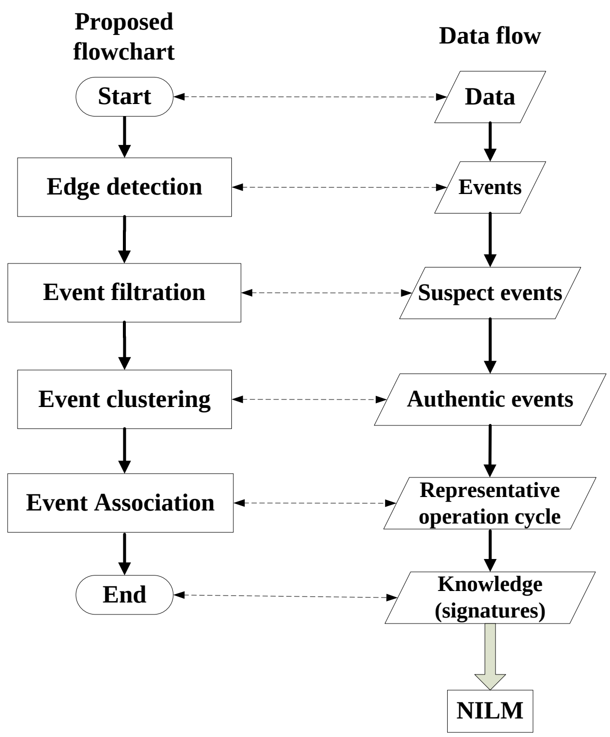


Figure 5.5: Flowchart of proposed approach versus corresponding data flow

5.2 Event Filtration

One operation cycle of each appliance may have two or more events depending on if it is an ON/OFF type appliance or a multi-state appliance. An ON/OFF type appliance such as a light bulb has only a pair of ON and OFF events while multi-state appliance such as furnace may have a series of operation state changes in the middle. An appliance’s events are heavily characterized by its function and physical electric attributes and can be roughly located. For example, a fridge usually has an ON event with real power of 70-300 W and a reactive power of 30 to 200 Var. Due to the cooling function of a fridge, the ON events can be observed during 24 hours even when users are sleeping; A microwave usually has an ON event that has an real power of 800 to 2500 W and a heavy third harmonic content. As a cooking device, a microwave is likely to be observed before mealtime.

In this chapter, the events that match the specific conditions of a certain appliance are defined as the "*suspect events*" of this appliance. The aim of event filtration is to locate the suspect events of a given appliance that may lead to the reconstruction of its entire operation cycle by carrying out future steps. To implement filtration, the conditions which can restrict the captured events are the: *real power range*, *reactive power range harmonic content range*, *with or without spike*, *single phase or double-phase* and *searching time*.

The real power, reactive power and harmonic content ranges are closely determined by the electric attributes of specific appliances. Residential loads can be roughly divided into four categories based on their linearity and reactivity: linear/active appliance, linear/reactive appliance, non-linear/active appliance and non-linear/reactive appliance. Depending on the type of category a certain appliance belongs to and its designed function, the numeric ranges of the above conditions can be quantified. For example, as a linear/active appliance, a resistive kettle has a very low reactive power and almost zero harmonic contents. Also, due to its water-heating purpose, its real power may statistically range from 1300 to 3000 W. Motor based appliances such as fridges can be viewed as inductors which lead to large reactive power. Another category is the switch-mode power supply based electronic appliances such as TVs and computers. They are neither inductive nor capacitive, but produce a large amount of harmonic contents. In addition, some appliances are both non-linear and reactive. Table 5-1 lists some examples of the above four appliance categories.

Table 5-1 Appliance categories and examples

Appliance category	Examples
linear/active	Stove, Kettle, Toaster
linear/reactive	Fridge, Freezer, Furnace, Dryer
Non-linear/active	TV, Desktop PC, Laptop
Non-linear/reactive	Washer, Microwave

A Spike can be another ancillary condition that helps locate some induction-motor based appliance events. A large inrush current occurs at the first moment of operation when the rotor is triggered from the station into movement. This unique feature is accompanied by an ON event, which can be reflected as a sharp edge on the appliance's real power curve.

The phase condition can easily separate some events from other appliances' events. In North America, some appliances are connected between two hot phases to gain a 240V voltage while others are connected between a single hot phase and a neutral to gain a 120V voltage. From identification perspective, the events of double-phase appliances can be observed simultaneously at both hot phases, whereas the events of single-phase appliances occur at only one of them. In a residential house, only a few heavy-consumption appliances are connected to double-phase electric outlets. These appliances are the stoves, ovens and clothes dryers.

The search window is another very important restriction that greatly reduces the searching space of suspect events. Some statistical studies are available on residential load behaviors which present typical appliance runtimes [97].For example, microwave's operations are more expected to be seen before breakfast, lunch and supper; lights are usually turned on in the early morning or after dark; fridges and furnaces are likely to run day and night. These occurrences show that, to locate the suspect events of specific appliances easily, it is best to search their less overlapped time ranges. For example, for a fridge, it is best to search from 2:00 AM-5:00 AM while many other appliances are generally inactive. For microwave, it's best to search the periods before meals. By doing so, the interference from other appliances can also be minimized.

In addition, to achieve the knowledge discovery purpose through massive amount of data, sufficient event samples must be obtained. Thus the data for multiple days need to be provided. According to the search window conditions, for different appliances, data pieces are cut from multiple days and re-joined together as a large data piece as shown in Figure 5.6. It is expected that the objective appliances' events will have a much higher density in the joint data pieces.

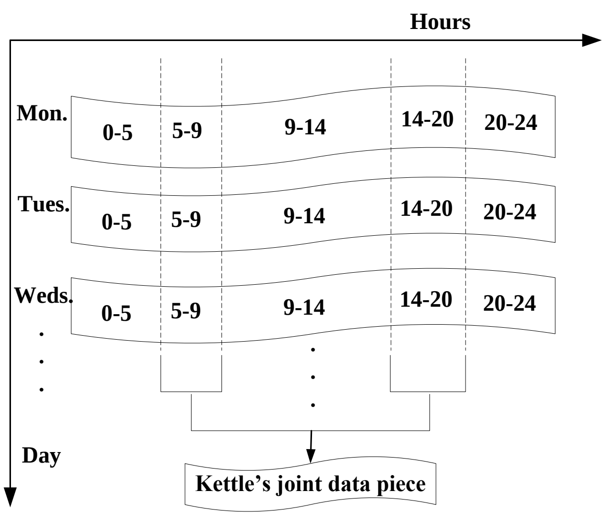


Figure 5.6: Example of data piece connection for kettle

Table 5-2 Example OF ON-Event Filtration Condition Table

Appliance	P(W)	Q(Var)	THD (%)	With spike?	Phase Condition
Fridge	70-300	30-200	0-20	Yes	Single
Furnace	120-800	200-800	0-20	Yes	Single
Microwave	800-2500	80-500	20-50	No	Single
Stove (big element)	1800-3000	0-30	0-5	No	Double
Stove (small	1000-2000	0-30	0-5	No	Double

element)					
Oven	2200-3600	0-30	0-5	No	Double
Kettle	1300-3000	0-30	0-5	No	Single
Clothe dryer	3000-6000	60-250	0-5	Yes	Double
Washer (Front-load)	80-300	<100	65-95	Yes	Single
Washer (Top-load)	300-1000	300-1200	0-20	Yes	Single

Based on the above discussion, Table 5-2 presents an ON-Event filtration condition table for 10 major appliances. The listed P/Q/THD values can be understood as the generic ranges of these appliances in the geographic area of our research location (Edmonton, Canada). The signatures are based on the measurements of different brand/models of appliances from several residential houses. At least 4 appliances of the same kind were measured. It should be noted that measuring all models/brands is impossible due to our limited resources. However, to compensate, the signature ranges are all expanded by at least 20% from our measurement values. For example, during the measurement, the power range of microwave was 1200-2000W, and in Table 5-2, it is modified to 800-2500W in order to be more inclusive. Since the electric attributes of appliances are essentially determined by their functions, they will generally fall into the above ranges. However, in different countries or regions, different electricity voltage levels, climates and even cultures may affect the above signature ranges. Thus, a more local ON-Event filtration table can be updated according to the local measurements.

Appliances may have different working modes. For example, a stove usually has 4 heating elements (two big ones and two small ones) on its panel. The small elements of a stove consume only approximately half the power of the big elements. Since their signatures are quite different, they should be treated as two different types of appliances.

It should also be noted that some appliances may have different working principles which result very different signatures. For example, most modern front-load washers are controlled by a variable-speed drive which outputs significant harmonic contents. In contrast, an old top-load washer behaves more like a regular big motor which outputs higher P/ Q and little THD. These two washers should also be treated as two different types of appliances.

In Table 5-2, P and Q are calculated based on the fundamental components of the current and voltage so that heavy harmonic contents will not affect their values [99]. For harmonic contents, the total harmonic distortion (THD) of current is used. Only the odd orders of the current harmonic contents(i_k) are considered due to their significance and usually a larger order than 9 does not need to be considered [99]-[100].

$$THD = \frac{\sqrt{\sum i_k^2}}{i_1}, k = 3, 5, 7... \tag{5.1}$$

According to the conditions listed in Table I, the suspect ON events of an appliance can then be identified by inspecting the events one by one in its joint data piece. To improve processing efficiency, only the ON events are verified and then considered as potential suspect events after passing inspection, because many transient features such as spikes accompany only ON events.

In order to guarantee that the events of an appliance can be included, the filtration conditions should not be set too strictly. In other words, the suspect events located by using the conditions in Table I may not be able to exclude the non-relevant events caused by other appliances; however, the conditions must allow enough of the events caused by the appliance to be included, because for a specific appliance, without enough event samples, the following procedures will fail or lead to inaccurate results.

5.3 Event Clustering

5.3.1 Definition of event clustering

As discussed in Section II, the suspect events of appliance X are a group of ON-events that are possibly but not necessarily caused by X. Since the conditions are defined as loose ranges and the data for multiple days are used, other events which do not belong to X can be easily included in the suspect events occasionally. The aim of event clustering is to determine which suspect events are the real ones belonging to X. In this chapter, the real events are named as the "*authentic events*" of X.

One basic assumption is that inside a suspect event group of appliance X, the number of events belonging to appliance X should be much larger than the number of events belonging to other appliances. The following examples below explain this assumption: suppose 50 suspect events of a fridge are located from a data piece for 2:00-5:00 AM of a whole week. It is possible that the user wakes up between 2:00-5:00 AM on Tuesday and for an unknown reason, uses an appliance such as a ceiling fan that accidentally meets the same filtration conditions as those of the fridge. Thus this fan's event will be mis-recognized as the fridge's suspect event and included in the fridge's 50-suspect-event group. This scenario might occur occasionally on certain days; however, this scenario is highly unlikely to occur as frequently as the fridge since such operations would be abnormal---he would have to turn on/off the fan not only every day but also as frequently as the fridge's kickins during sleeping time. Another example is a microwave. The user might turn on/off an unknown appliance that draws similar P/Q/Harmonic signals as a microwave does before mealtime, but it is unlikely to occur as often as a microwave. For any of these abnormal scenarios, the proposed unsupervised approach is not intended for and not able to deal with.

Furthermore, in some cases, corrupted events may occur that are usually due to simultaneous occurrences of more than one appliance event. This rarely happens but is possible. For example, the accidental overlapping of two single events belonging to a fridge and furnace will result in a power jump that is roughly the summation of the two events. Normally, the frequency of these corrupted events is much smaller than the authentic events.

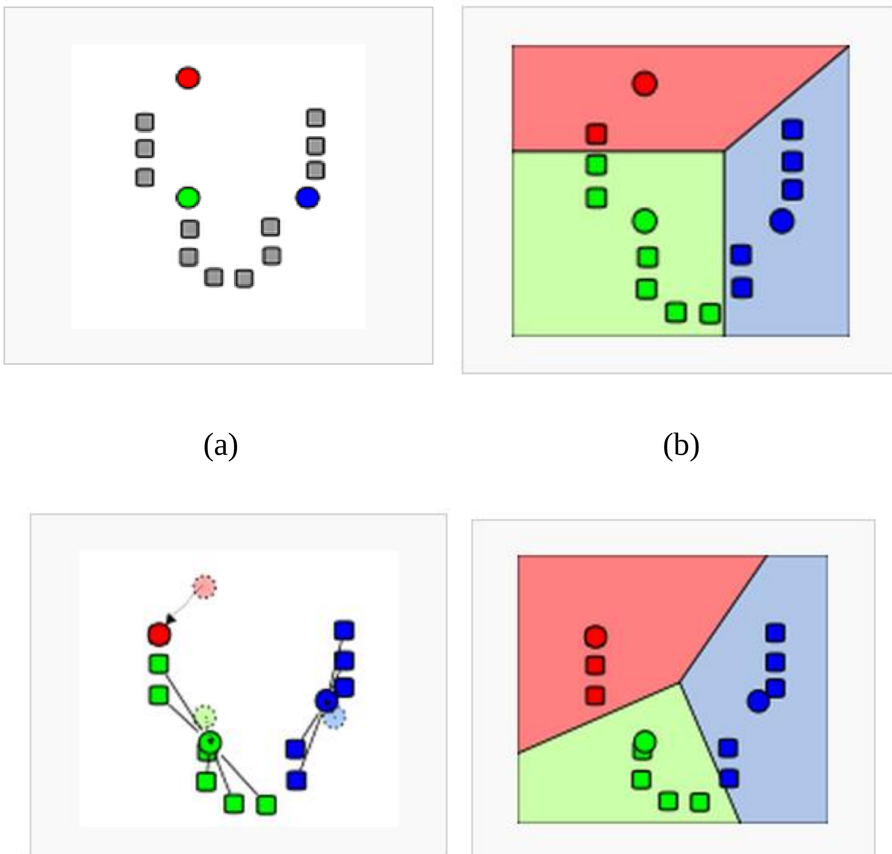
Based on the above assumption, if the number of groups of events and the number of events each group possesses are known, the authentic events can be determined as those in the event group with the maximum number of members. Clustering is an effective tool to obtain the information. A clustering algorithm determines which events are roughly the same and can be grouped together as one cluster. After event filtration, since the clustering space is greatly reduced, many uncertainties and noises can be ruled out and clustering can be done much faster and more accurately compared to applying clustering to all the events of a day in the beginning. For instance, as Table 5-3 shows, suppose 88 suspect events of a fridge are located and after applying clustering algorithm, only four clusters are found. Then from the number of events each cluster owns, the dominant cluster which has 75 event members can be identified as having been caused by the fridge and these 75 event members can be labeled as the fridge's authentic events. On the other hand, smaller clusters are discarded because they are likely to be from other appliances or noises and cannot represent the objective appliance. The clusters and their mean values are listed in Table 5-3. Two types of clustering methods are applicable to our study.

Table 5-3 Example of composition of suspect events

Cluster Index	Mean P (W)	Mean Q (Var)	Mean THD(%)	Number of Events	Physical cause
1	100.3	76.2	10.6%	75	Fridge
2	87.7	67.9	10.3%	10	Fan
3	73.6	58.8	2.2%	2	Motor X
4	189.6	138.5	9.9%	1	Corrupted

5.3.2 Selection of clustering method

Our proposed approach selects the mean-shift algorithm for clustering. Before introducing mean-shift clustering, it is necessary to explain another related clustering method named K-means clustering. Comparison of the two will also be given afterwards.



(a) (b)

(c) (d)

Figure 5.7: Example of K-means algorithm

K-means clustering is a classic method of clustering analysis that can partition n samples into k clusters in which each sample belongs to the cluster with the nearest mean. It is firstly proposed in [106]. An illustration of its standard algorithm is shown as follows:

As shown in Figure 5.7, before the procedure begins, the number of clusters *K* must be initialized. In this example, *K* is set as three. Accordingly, in step (a), three "means" are randomly chosen from the given dataset. In (b), three clusters are created by grouping every sample with its nearest mean. In (c), based on the three clusters that are determined from (b), centroids of the three clusters are calculated and the three old means are now "shifted" towards the three new centroids. In the end, three new clusters are formed in (d). After this, (b) and (c) are repeated until means are not shifted (convergence has been reached).

The above K-means procedure is straightforward. However, it is difficult to be applied to our application---load event clustering. The reason is K-means algorithm normally requires the number of clusters *K* as a priori that is actually the number of load event types in the objective house. However, as a non-intrusive approach, the number of loads in the house is unknown. To cope with this challenge, another clustering approach---mean-shift clustering is adopted. Mean-shift clustering is a density based clustering algorithm. It has been widely applied to the areas of computer vision and pattern recognition, such as intelligence fusion, target following, and image segmentation (Figure 5.8) [107]-[108]. In [109], the most widely used mean-shift clustering was proposed.

The comparison of K-means and mean-shift clustering in general can be found in [110]. It has been mentioned that one advantage of mean-shift over K-means is that there is no need to know and choose the number of clusters before clustering starts. Instead, mean-shift requires the bandwidth parameter that restricts the variance of samples belonging to a certain cluster with respect to the cluster mean. For residential appliance loads, the variation of electric signatures is determined by the system voltage fluctuation level and the variation range can be pre-estimated. In the earliest publication of NILM [55], this phenomenon has already been explained in detail. In other words, the bandwidth required by mean-shift can be easily obtained while the number of clusters required by k-means is difficult. [110] also mentioned mean-shift might be slower than K-means but this is not a strong concern because non-intrusive signature extraction is an off-line process that does not require fast computation.

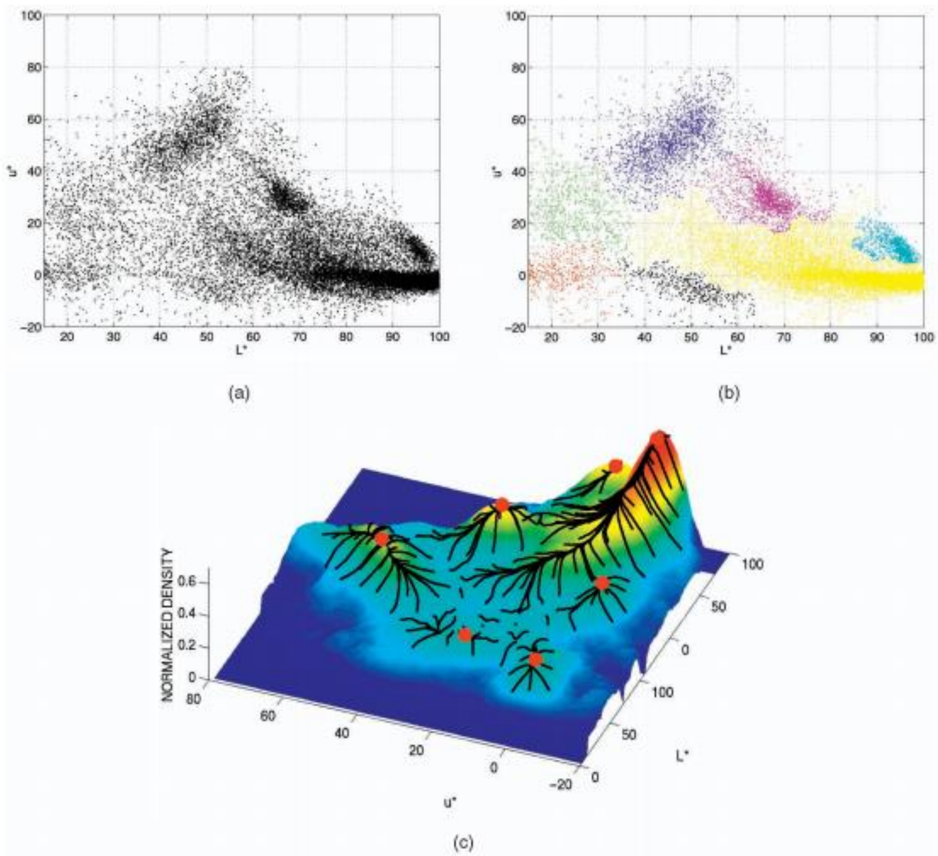


Figure 5.8: Example of Mean-shift clustering applied to image segmentation

Furthermore, [111] explained the above difference and specifically discussed the feasibility of applying mean-shift to residential appliance identification. It demonstrated a simple attempt of mean-shift algorithm on clustering a few appliance operations (the "triangle"-shaped operation and "rectangle"-shaped operations) in a residential house. The trail in this paper is not a complete approach but has proven the feasibility of mean-shift clustering applying to NILM related applications.

For mean-shift clustering, assuming the sample data is a finite set *S* in the n-dimensional Euclidean space, *X*, its kernel function is shown as below:

$$K(x) = \begin{cases} 1, \|x\| \le \lambda \\ 0, \|x\| < \lambda \end{cases} \tag{5.2}$$

The mean of the cluster is

$$m(x) = \frac{\sum_{s \in S} K(s-x)s}{\sum_{s \in S} K(s-x)} \tag{5.3}$$

The difference $m(x) - x$ is defined as the mean shift. The steps of mean-shift clustering algorithm are listed as below:

Step 1: Initialize random samples as cluster centers;

Step 2: Within the ranges defined by a specific radius or bandwidth, calculate the means of each cluster ;

Step 3: For all the clusters, shift the cluster centers towards the new means calculated in step 2;

Step 4: Repeat step 2 and 3 until the centers of all the clusters are not moving anymore. In other words, the means of all the clusters are not updated anymore;

Step 5: Merge similar clusters that have close means together.

In our study, P/Q/THD are selected as three attributes used for clustering purpose. The "mean" linkage type is selected to determine the distance between two clusters. All features are firstly normalized to the range of 1 to 100 through min-max scaling [113]. For mean-shift clustering, its bandwidth parameter is set from 5-20 due to the estimated variation of the electric signatures caused by the local system voltage fluctuations. Generally, for the above two clustering methods, when the power variance inside the suspect event group is large, the bandwidth of mean-shift clustering can be relatively lowered in order to strengthen the cluster differentiation; when the power variance inside the suspect event group is small, the bandwidth of mean-shift clustering can be relatively increased to strengthen the cluster fusion.

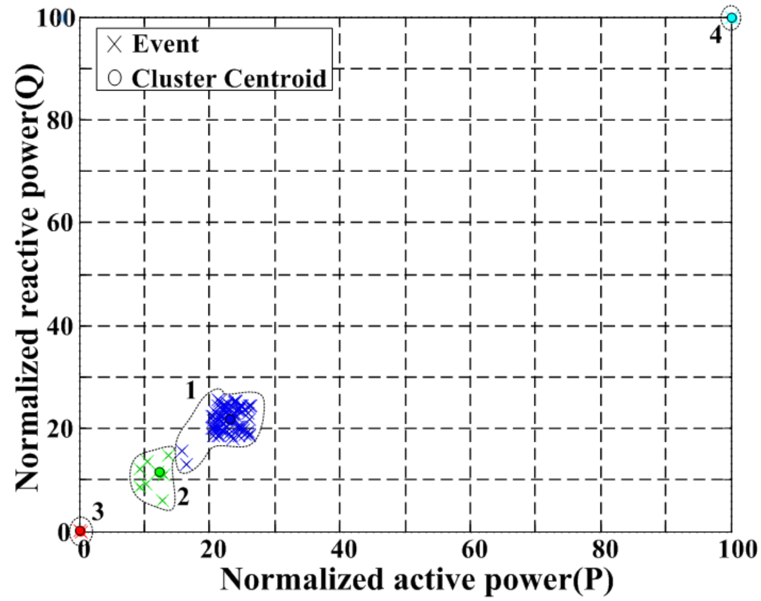
5.3.3 Feature selection for mean-shift clustering

The above standard mean-shift clustering has one drawback due to the Euclidean space it uses where different attributes will be treated as equally important. However, as explained in Table 5-1, depending on which category the appliance type belongs to, different electric features should not be treated equally because some features may be almost noise features. For example, for a fridge, as a linear/reactive appliance, it produces considerable active/reactive powers. But it barely produces any harmonic content and thus the THD feature is actually a noise feature that might be caused by misreading from the meter-side; for a stove, both of its reactive power and harmonic contents should be disabled because it is basically a resistor.

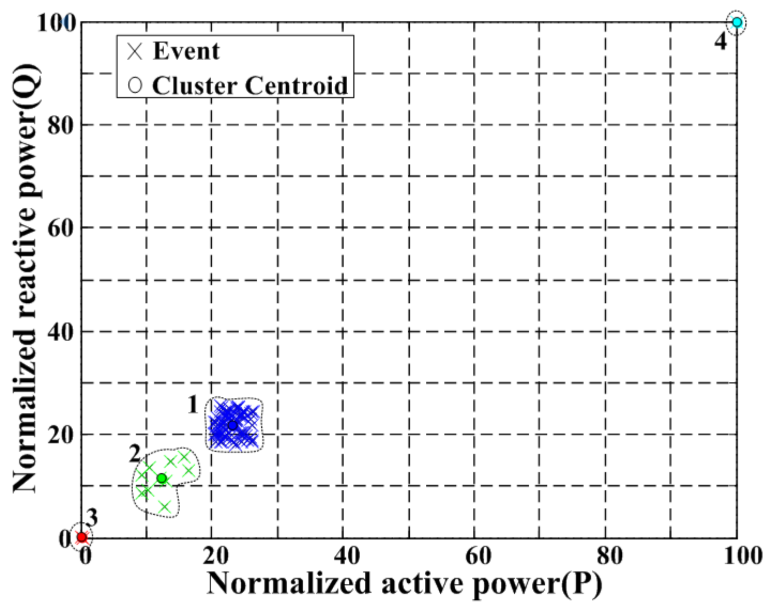
To cope with the above noise feature problem, P/Q/THD features can be easily selected according to Table 5-1. This is because for the proposed non-intrusive signature extraction algorithm, the type of appliance is already known in advance and the selection of features can be therefore determined based on the category the appliance type falls into. The effect of feature selection is similar to using a Minkowski metric weighted space explained in [112] where the "relative distance" on low-weight features contributes little to the total distance.

To make a comparison, both of the mean-shift clustering methods without feature selection and with feature selection were applied to the 87 suspects discussed in Table 5-3. They consist of 75 events from a fridge (1), 10 from a ceiling fan (2), 2 from a motor X (3) and 1 corrupted event (4). The clustering results are shown as P-Q 2D plots in Figure 5.9. It shows that the mean-shift clustering method without feature selection mistakenly includes 2 fans' events in its authentic group (the fridge's event group) because both the fridge and fan have similar THD values. However, because of the fridge's linear/reactive load type, its harmonic content should not contribute because it is usually small and instable, and can easily become mixed up with other linear appliances; in contrast, after feature selection, the THD feature is ruled out and the result is able to exclude the

2 events caused by the fan. Generally, mean-shift with feature selection has a better performance than mean-shift clustering in the proposed application.



(a) Mean-shift clustering without feature selection



(b) Mean-shift clustering with feature selection

Figure 5.9: Effect of feature selection

5.4 Event Association

After the event clustering, the knowledge of the appliance is still incomplete since the other events except for the ON event in an appliance's operation cycle are still unknown. For accurate energy-tracking purposes, all the events and even the event pattern signatures [104] of an appliance have to be obtained. For example, for a fridge, its OFF event needs to be known; for multi-stage appliance such as a furnace, it is even more important since the ON event is only a small part of the furnace's full operation cycle. Its middle events and OFF event also need to be found. Knowing how these middle stage events occur in a particular pattern is also valuable. The overall purpose of event association is to determine the other events which also belong to the objective appliance and hence reconstruct its full *representative operation cycles*. .

The proposed method counts the number of event occurrences and calculates the frequency to judge whether an event is associated with a specific appliance and determines the event's association type. An operation study of actual appliances [104] defines three association types:

1. Single event: It has a strong association with an appliance. Once the appliance is working, this event must appear. However, it appears only once in single operation cycle. As Figure 5.10 shows, a fridge has a fixed OFF event and it accompanies every operation cycle of the fridge but appears only once.

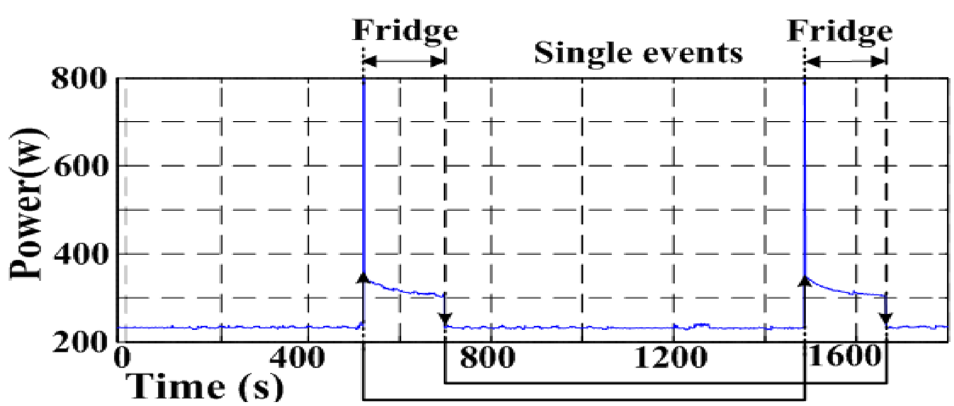


Figure 5.10: Example of single events.

2. Repetitive event: It has a strong association with an appliance. Once the appliance is working, this event must appear. It can appear more than once in a single operation cycle. As shows in Figure 5.11, a furnace has repetitive events caused by its heating elements, which can be triggered multiple times according to the environment temperature.

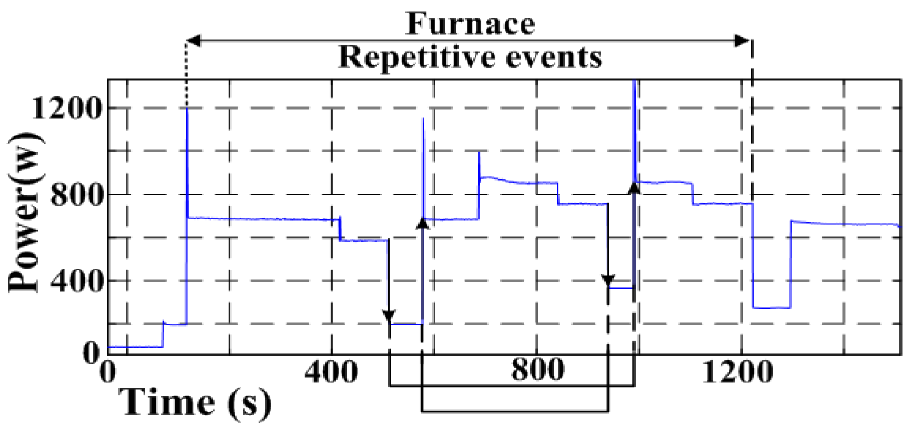


Figure 5.11: Example of repetitive events

3. Occasional event: It has a weak association with an appliance. Once the appliance is working, this event might not occur. As shown in Figure 5.12, when the door of a fridge is open, the light inside the chamber will be automatically switched on; when the door is closed, light will be switched off. Then users can see a small pair of power jumps between ON and OFF. An occasional event is usually caused by an ancillary element of appliance, such as the light of a fridge and the hood of a stove.

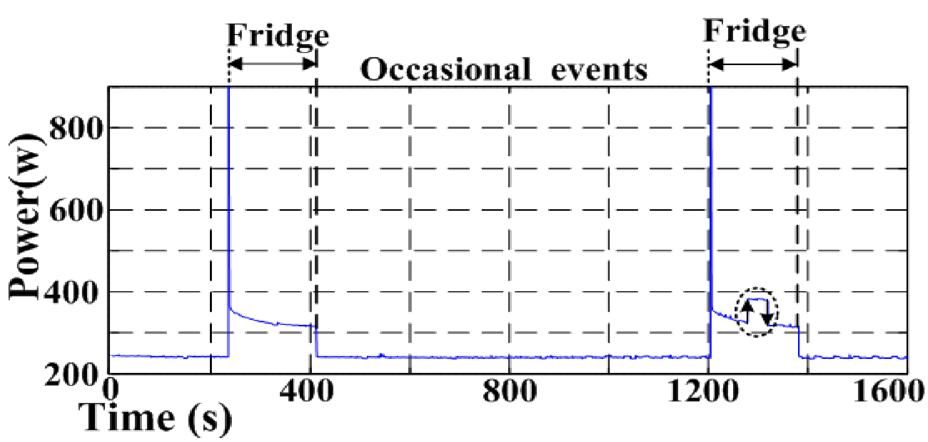


Figure 5.12: Example of occasional events

4. Unrelated event: It has zero association with an appliance because it is caused by another appliance. As shows in Figure 5.13, during a fridge's operation, a big incandescent lamp is also turned on. The ON event of the lamp happens before the OFF event of the fridge and the OFF event of the lamp happens after the OFF event of fridge.

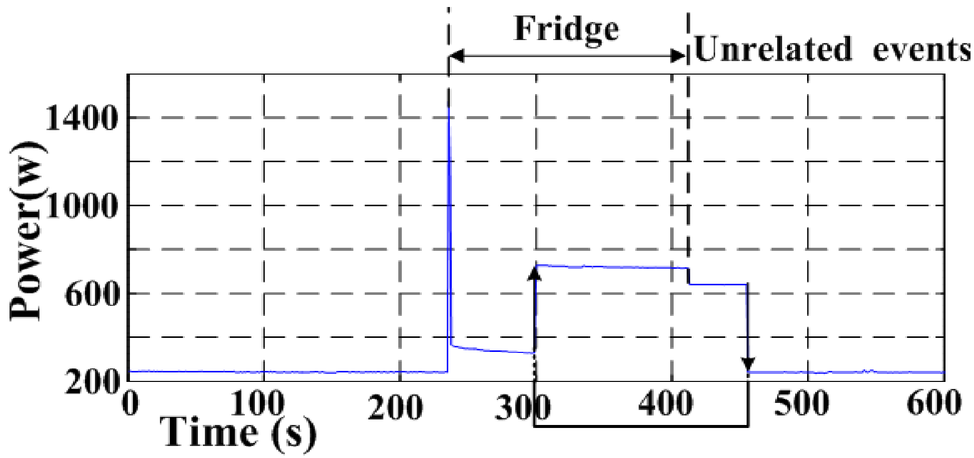


Figure 5.13: Example of unrelated events

Table 5-4 Average duration and data segment length for typical appliances

Load name	Average duration	Data segment length
Fridge(cycle)	15 mins	22.5 mins
Furnace(cycle)	20 mins	30 mins
Microwave	4 mins	6 mins
Stove	25 mins	37.5 mins
Kettle	4 mins	6 mins
Oven	10 mins	15 mins
Washer	45 mins	67.5 mins
Clothes dryer	50 mins	75 mins

The way to locate associated events and determine their association types can be determined by examining the data segments starting from the authentic ON events (determined by the event clustering in the last step). Each data segment is defined as a piece of data that has an authentic ON event in its beginning. The length of the piece is properly defined so it will not be too short (may fail to include all the associated events of the appliance) while it will also not be too long (may include too many unrelated events as interferences). The reasonable lengths can be defined as 1.5 times the average durations of the appliance operation. Table 5-4 shows the average durations of typical appliance operation and their segment lengths.

Suppose there are M authentic events, there will also be M corresponding data segments which may include not only associated events of appliance but other irrelevant events. Applying event clustering methods discussed in Section III to all events in all segments together, clusters of events can be determined. According to the definitions of three association types, theoretically, the criteria in Table 5-5 can be used to examine association of those clusters individually.

Table 5-5 Theoretical criteria for association determination

Association type	No. of events the cluster include (N)
Single event	$N=M$
Repetitive event	$N>M$
Occasional event	$bM<=N<M$
Irrelevant event	$N<bM$ ($0<b<1$)

For single event, they should appear in each data segment and thus event number N should be equal to data segment number M; for repetitive event, since in some data segments, multiple events may appear, event number N should be larger than M; for occasional event, its occurrence should be more frequent than a threshold ruled by b. b is a factor to differentiate if an event is related to objective appliance or not. Considering possible mistaken clustering or missing events due to improper data segmentation, the above theoretical criteria should be corrected using confidence c and additional conditions shown in Table 5-6. Normally, values of b,c can be set as 0.3 and 0.8.

Table 5-6 Criteria for association determination

Association type	No. of events the cluster include (N)
Single event	cM<=N<=M
Repetitive event	N>=cM & n>=2 in at least one certain segment
Occasional event	bM<=N<cM
Unrelated event	N<bM (0<b<1)

An example of how event association works is shown in Figure 5.14 and Table 5-7. For the given appliance, ON-event “1” is the initial authentic event determined by the event clustering step previously. Accordingly, 4 data segments of this appliance are located and each segment has the same length determined referring to Table 5-4. At the beginning of event association, clustering is applied again to all the events included in these segments. As a result, each event can be labeled by the index of the cluster it individually belongs to. After the occurrences of each cluster are counted, the association type can be determined by using criteria in Table 5-6. The results are presented in Table 5-7 .

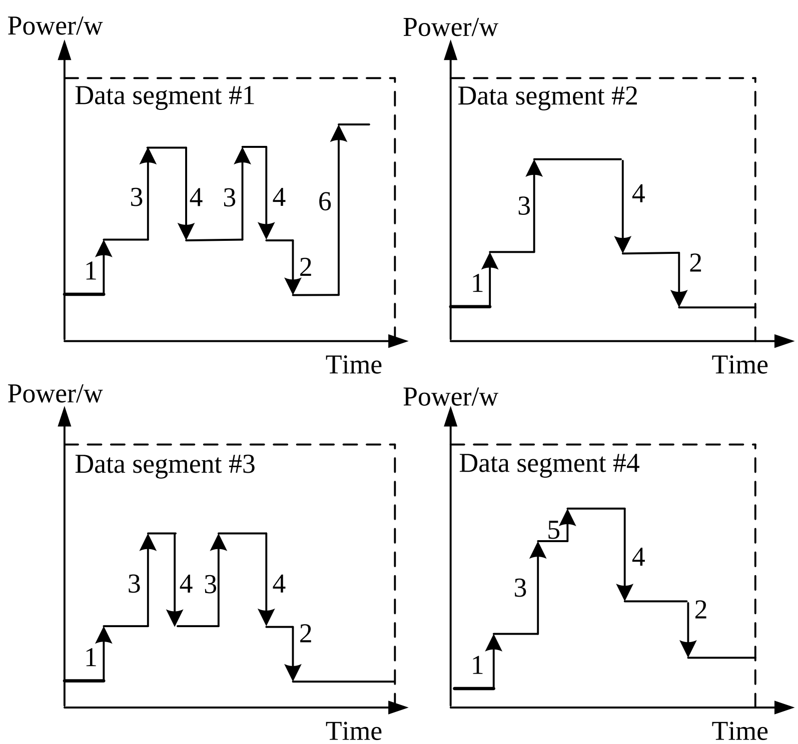


Figure 5.14: An example of event association from 4 data segments

In this example, the authentic operation cycle of the given appliance is determined as 1→3→4→2 (~~~: repetitive). An effective way to improve the above association accuracy is to have more data segments, which may require more data pieces from more days, say a week. The more data segments being used, the less likely an unrelated event is to be incorrectly judged as a related one due to the error of the confidence factors b and c; also, the impact of a possible clustering error from the previous step or the missing events due to improper data segmentation can be reduced to a minimum.

Table 5-7 Example of event association judgment

Item	Number of appearances	Criteria used (b=0.3;c=0.8)	Association type
Data segment	M=4	---	---
Event 1	N=4	cM<=N<=M	Single event
Event 2	N=4	cM<=N<=M	Single event
Event 3	N=6	N>=cM & n>=2	Repetitive event
Event 4	N=6	N>=cM & n>=2	Repetitive event
Event 5	N=1	N<bM	Unrelated event
Event 6	N=1	N<bM	Unrelated event

5.5 Verifications and Discussions

5.5.1 Verification and discussions based on real house #1's data

The above algorithms were tested by using data acquired from a real residential house in Edmonton, Canada for a week with no special attention from the house owner. A laptop-based data acquisition system was hooked the house's electricity panel and continuously collected all the voltages and currents from the two hot phases (A and B) inside. It behaves just like a smart meter. The data were sampled every second. In each second, 6 consecutive cycles are acquired and each cycle has 256 points. Data acquired was then processed using the proposed algorithm and the operation cycles of specific appliances were automatically learned and reconstructed. In order to compare, operation times of appliance activities were manually recorded. By doing this, true operation cycles of appliances could be directly labeled as references. If the labeled events were more than one, their average values were used for comparison. In this house, there were 3027 events (>50W) in Phase A, 2055 in Phase B and 2012 in Phase A-B. The total power of both phases on a typical day is shown in Figure 5.15. Table I is adopted for event filtration. In total, 10 appliances were found and two different fridges were connected to the two different phases.

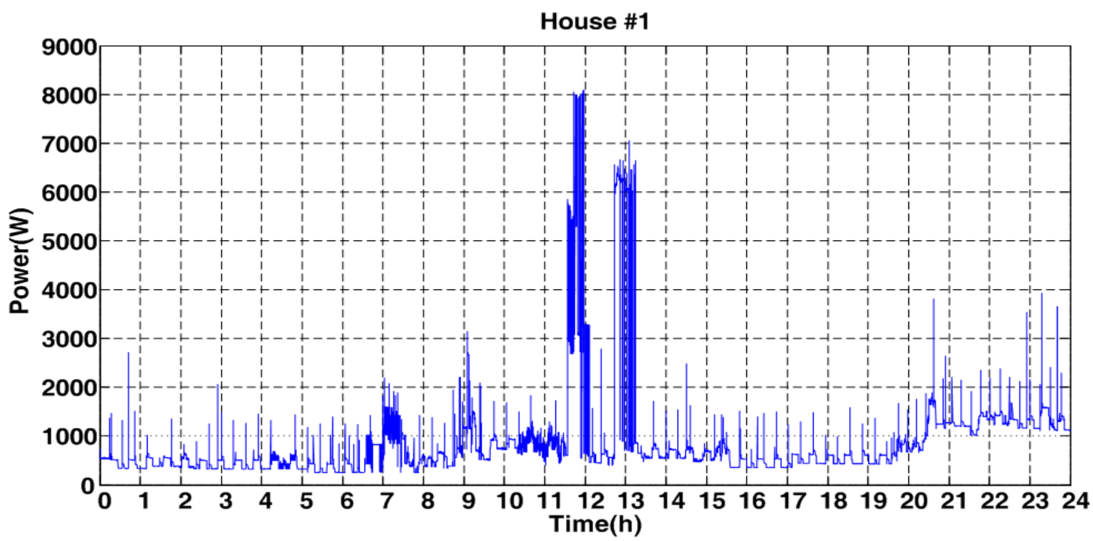


Figure 5.15.The total power on a typical day from house #1

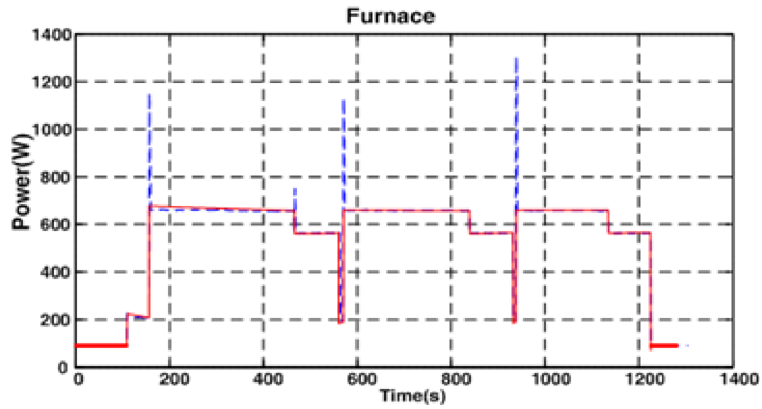
Table 5-8 shows the search window and authentic events determined in the step of event clustering. As can be seen, for each appliance, the number of authentic events (the largest no. of events in a single cluster) is larger than the average number of events per cluster. This result supports the assumption in Section III that authentic events will dominate inside suspect events.

Table 5-8 Search window and results of event clustering

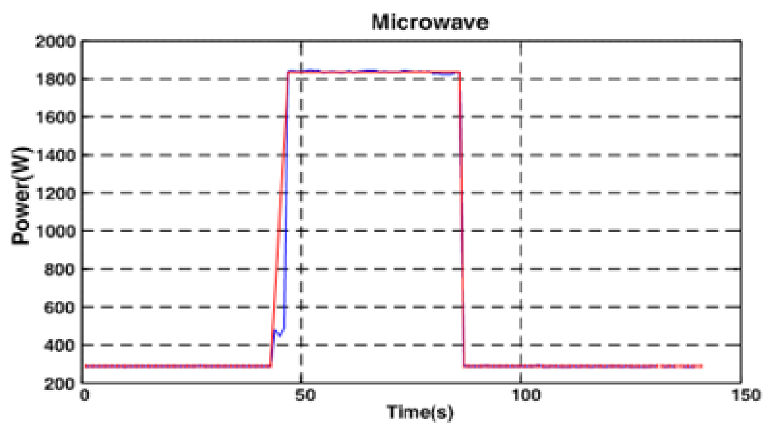
Appliance	Search window	# of suspect events	# of clusters	Largest no. of events in single cluster
Fridge	2:00AM-	69	10	21 (30.4%)

(PhaseA)	5:00 AM			
Fridge (Phase B)	2:00AM-5:00 AM	31	2	7(22.6%)
Furnace	2:00AM-7:00AM 8:00PM-0:00AM	4	1	4(100%)
Micro-Wave	6:00AM-9:00AM; 11:00AM-2:00PM; 4:00PM-8:00PM	54	7	27(50%)
Stove(big element)	4:00PM-8:00PM	517 (Repetitive)	9	505(97.7%)
Stove (small element)	4:00PM-8:00PM	77 (Repetitive)	5	60(77.9%)
Oven	8:00AM-8:00PM	7	5	3(42.9%)
Kettle	6:00AM-9:00AM; 11:00AM-2:00PM; 4:00PM-8:00PM	11	4	8(72.7%)
Clothes dryer	8:00AM-11:00PM (Weekend)	92 (Repetitive)	5	61 (66.3%)
Washer	3 hrs before dryer	380 (Repetitive)	10	189(49.7%)

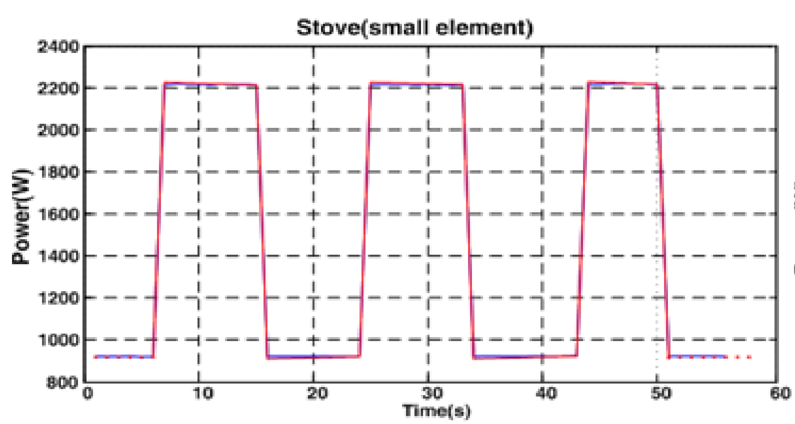
As for the reconstructed operation cycles, the following two verification methods are conducted:



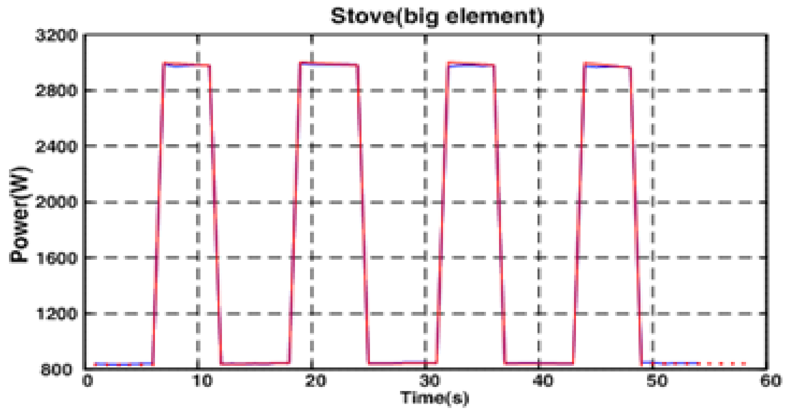
(a) Furnace



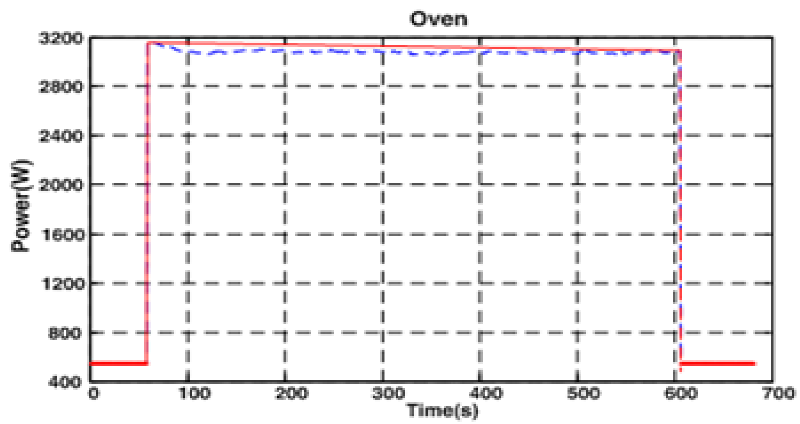
(b) Microwave



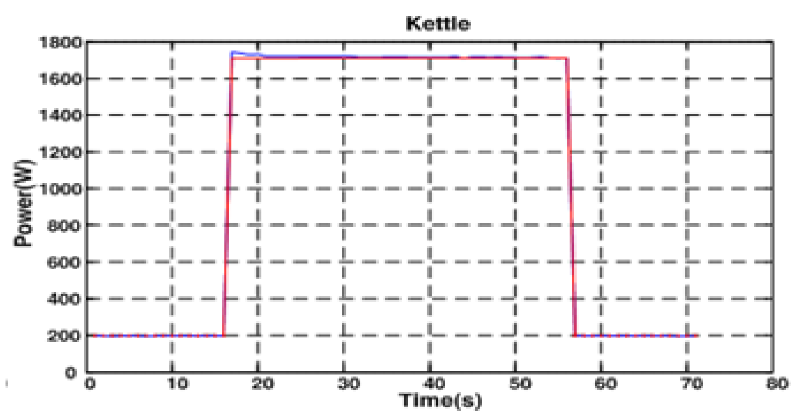
(c) Stove(small element)



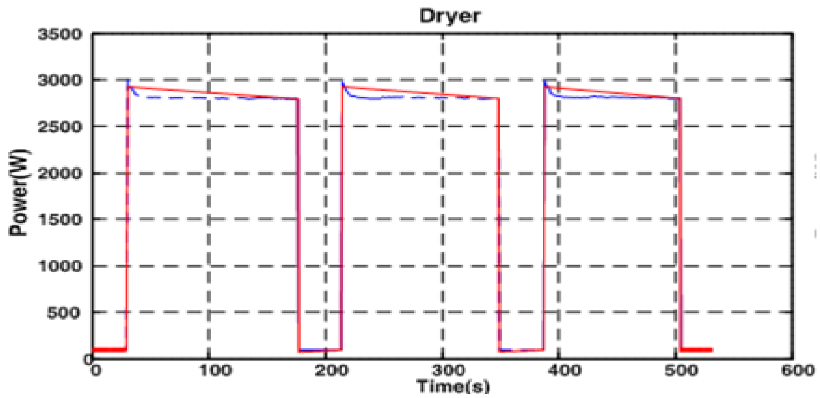
(d) Stove (big element)



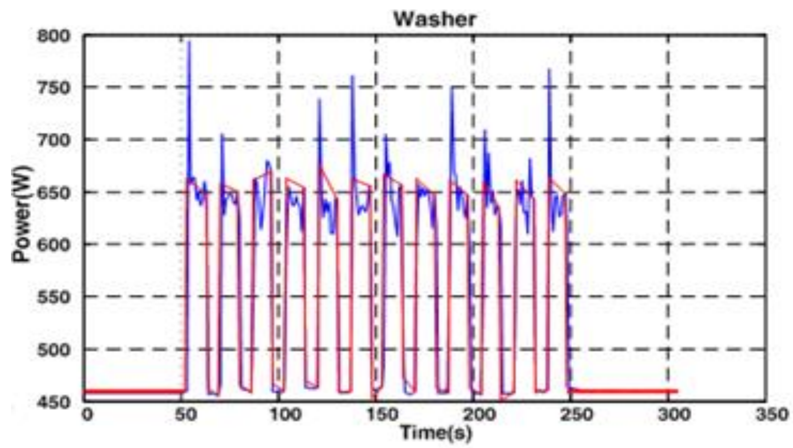
(e) Oven



(g) Kettle



(h) Dryer



(i) Washer

Figure 5.16.Reconstructed cycles (red) vs. Labeled real cycles (blue) in house #1

1. To visualize the effectiveness of the approach, the reconstructed operation cycles are compared with the labeled reference cycles in Figure 5.16. For the sake of convenience, the events determined by using event clustering and association are aligned with the original events in the labeled cycles. As Figure 5.16 shows, they are quite similar. The only difference is that some power variations between neighboring events are not captured, because the proposed approach focuses only on events. However, this missing information is not important since most NILM algorithms do not use it. Particularly, if any transient information accompanied

with an event is needed, the corresponding event sample can be simply loaded for study.

2. An appliance's reconstructed operation cycle contains the electric signatures of all its associated events such as P, Q, THD (but not limited to these, depending on the needs of NILM).

Table 5-9 presents a detailed comparison between the signatures of individual events in the reconstructed cycles and labeled cycles. To compare, the mean values of event clusters are adopted. The errors are calculated with respect to the labeled reference events as true values.

Table 5-9 shows acceptable accuracy between the reconstructed cycles and labeled cycles. Please note the electric signatures of appliances can change within +/- 5% due to system voltage fluctuation. As

Table 5-9 shows, most of the real power errors are lower than 5%. Some errors of reactive power or harmonic THD are a little higher, because the corresponding appliances produce little reactive power or harmonic contents. The true values of these attributes are comparable to signal noises and can greatly fluctuate between measurements, especially when they are small. For example, even if several measurements are directly performed for the kettle's harmonic THD, differences at the listed level may still be observed. Another example is the washer. Since it is controlled by a variable speed driver, which continuously generates many repetitive event pairs with heavy noises (as shown in Fig.7), the inherent signature consistency of these events is lower than that of other appliances. Thus, the above errors are quite normal. Furthermore, they will not affect NILM's identification significantly. Very small or zero weights will be given to these unstable attributes when making identification [104]. Similarly, classifiers such as neural networks [69] will "ignore" these signatures automatically during their training stages because these signatures in their training sets vary a lot too. The mean values being used can also balance out the variations of abnormal events inside clusters. Generally speaking, the electrical signature accuracy is very satisfactory.

Table 5-9 Electric signature error between reconstructed cycles and reference cycles for house #1

Appliance Event		Power Error %	Reactive Power Error %	Harmonic THD Error %
Fridge(PhaseA)	ON	0.54	0.45	-4.41
Fridge(PhaseA)	OFF	1.80	-0.39	-0.58
Fridge(PhaseB)	ON	3.00	7.31	-4.75
Fridge(PhaseB)	OFF	-0.01	-0.04	-5.48
Furnace	ON	0.73	4.11	1.94
Furnace	Middle 1	1.13	-4.91	2.16
Furnace	Middle 2	-1.41	1.40	-3.16
Furnace	Middle 3	1.11	1.44	6.57
Furnace	OFF	-0.13	1.03	-2.72
Microwave	ON	-0.32	0.79	-0.86
Microwave	OFF	-0.85	-2.72	-2.68
Stove(small)	ON	0.52	0.17	2.24
Stove(small)	OFF	-0.41	0.12	-7.46
Stove(big)	ON	-1.15	1.93	-0.91
Stove(big)	OFF	0.14	-4.70	-0.68
Oven	ON	1.21	-0.21	2.98
Oven	OFF	-0.72	-4.49	-3.08
Kettle	ON	-2.19	4.46	-5.84
Kettle	OFF	0.24	-2.81	12.08
Dryer	ON	0.96	0.97	-0.20
Dryer	OFF	0.63	-1.41	-4.89
Washer	ON	-4.21	-0.54	-1.41
Washer	OFF	-5.31	-3.32	-7.29

Also, the speed of the proposed algorithm was found to be fast. Based on a regular desktop PC (Intel 2-Quad CPU 2.33GHz, 4GB memory), the 3-stage time required to process the appliances are recorded. The most time-consuming appliances “Stove(big element)” and “Washer” in house #1 is listed in Table 5-10.

Table 5-10: 3-stage time required for the most time-consuming appliances

Appliance	Event filtration(s)	Event clustering(s)	Event association(s)	Total time(s)
Stove(big)	0.012243	0.441213	0.030468	0.4839
Washer	0.009009	0.359888	0.027966	0.3969

5.5.2 Verification and discussions based on real house #2's data

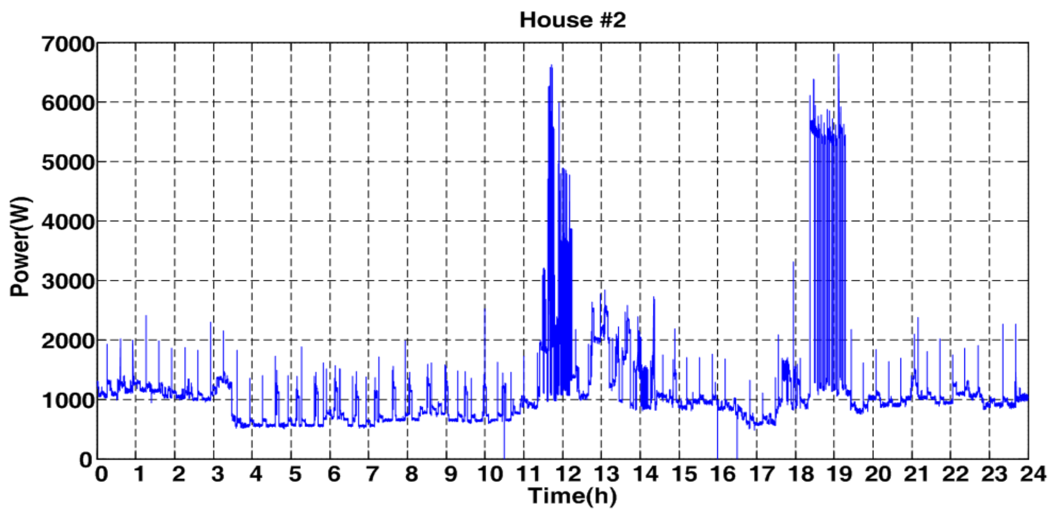


Figure 5.17: The total power data on a typical day from house #2

Similarly to house 1, another real residential house with a smaller family size in Edmonton was also measured for a week. The house contained 1451 events (>50W) in Phase A, 1698 ones in Phase B and 564 ones in Phase A-B. The house's total power on a typical day is shown in Figure 5.17. It should be noted in this house an old top-load washer is used. Its reconstructed cycle referring to a real labeled cycle is shown in Figure 5.18, which shows that between ON and OFF, its power gradually decreases and becomes stable after a few minutes.

By using the proposed approach, 8 major appliances were found (the house has only one fridge). As Table 5-11 shows, the electric signature is quite accurate with respect to the signatures contained in the labeled reference cycles.

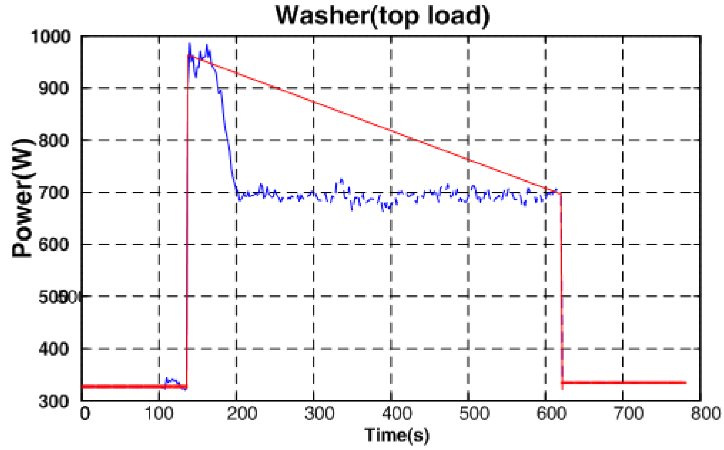


Figure 5.18: Reconstructed cycles (red) vs. l with labeled real cycles (blue dash) for the top -load washer in house #2

Table 5-11 Electric signature error between reconstructed cycles and reference cycles for house #2

Appliance Event		Power Error %	Reactive Power Error %	Harmonic THD Error %
Fridge	ON	1.84	3.04	-5.87
Fridge	ON	-1.93	-1.01	3.33
Furnace	ON	0.82	-2.03	-5.57
Furnace	Middle 1	-1.71	-6.99	1.20
Furnace	Middle 2	-4.37	-5.54	-5.39
Furnace	Middle 3	1.42	-0.40	3.93
Furnace	Middle 4	-0.80	-0.86	-2.55
Furnace	OFF	-2.73	-2.62	-1.58
Microwave	ON	-1.24	-8.78	-1.77
Microwave	OFF	1.91	-6.46	3.25
Stove(small)	ON	4.65	-5.41	2.02
Stove(small)	OFF	4.14	-4.78	0.37
Oven	ON	-1.42	2.88	-5.58
Oven	Middle	-1.47	-4.60	2.85
Oven	OFF	1.04	0.44	1.69
Kettle	ON	1.04	0.44	1.69
Kettle	OFF	0.41	2.75	-3.71
Dryer	ON	1.91	-0.59	-2.99

Dryer	OFF	0.97	0.98	3.33
Washer	ON	5.32	1.85	3.15
Washer	OFF	0.85	0.07	3.57

5.5.3 Verification and discussions based using MIT public dataset

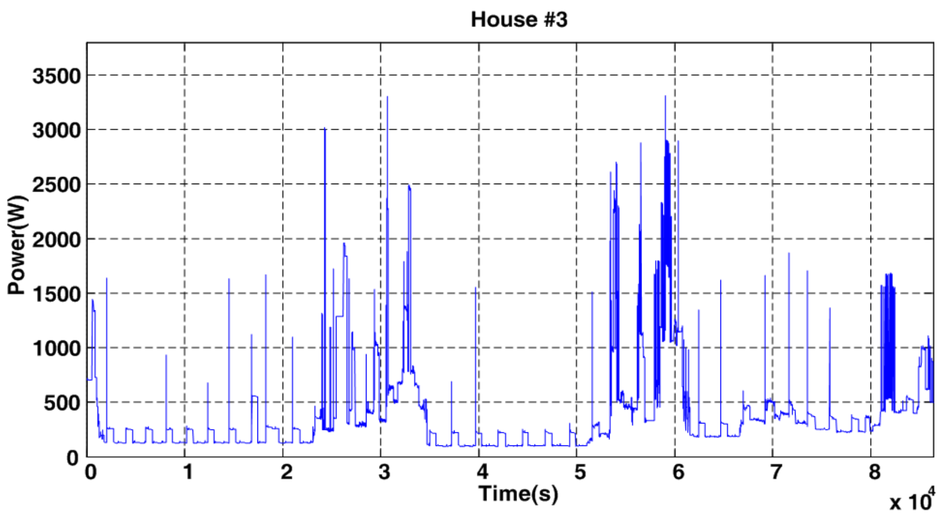


Figure 5.19. The total power data of the first 86400 points from house #3

To further verify the proposed approach, the data from house 3 in the REDD dataset [95] was used. REDD, a public dataset available for NILM research was released by MIT in 2011. The data was acquired from the greater Boston area in US. Please note: 1. In this dataset, many middle hours in a day are missing. For example, activities of the stove, oven and kettle were not recorded. 2. Although REDD labels specific appliances but unfortunately, some appliances are not clearly labeled such as the washer. In general, the dataset contains 1427284 seconds or roughly 16 days. The sampling rate was 275 points/cycle. The house contains 1160 events (>50W) in Phase A, 2497 in Phase B and 523 in Phase A-B. As an example, the total power of the first 86400 seconds (the length of a day) is plotted in Figure 5.19.

In spite of the above issues in the dataset, 5 major appliances can still be extracted using proposed approach. Surprisingly, even when the washer activities were not labeled in the REDD dataset, they were still able to be located and "mined" by using the proposed method. With the help of the proposed method, the washer's cycles were re-labeled. Its reconstructed cycles verses real cycles are plotted in Figure 5.20. Like the washer used in house 1, this is a front-load washer with a variable-speed drive inside.

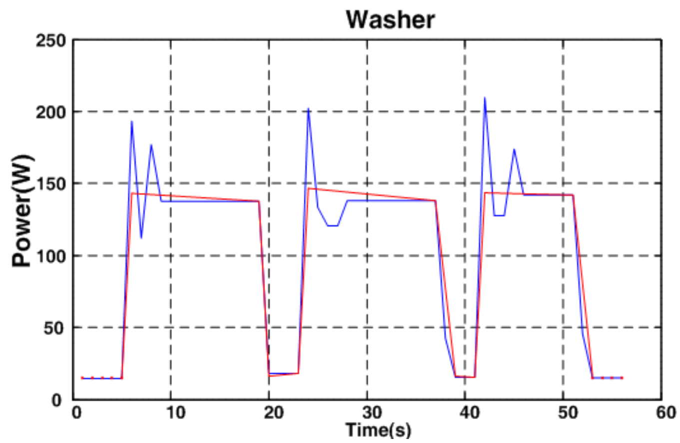


Figure 5.20: Reconstructed cycles (red) vs. real cycles (blue dash) for washer in house #3

As Table 5-12 reveals, the electric accuracy of reconstructed cycles is very good. It should be noted the filtration conditions in Table I used in this case are based on our local measurements in Edmonton, Canada. However, it is found still viable to deal with the house in the Boston area. This finding implies that appliances may share common signature ranges in similar geographic regions (such as northern part of North America).

Table 5-12 Electric signature error between reconstructed cycles and reference cycles for house #3

Appliance Event		Power Error %	Reactive Power Error %	Harmonic THD Error %
Fridge	ON	-1.04	-1.23	-4.16
Fridge	ON	-0.20	-0.55	0.25
Furnace	ON	-2.36	-2.49	-1.92
Furnace	Middle 1	-1.47	-5.46	5.39
Furnace	Middle 2	7.21	-1.81	-0.96
Furnace	Middle 3	2.46	3.88	-7.12
Furnace	OFF	0.61	0.81	-2.68

Microwave	ON	-0.83	1.95	-1.92
Microwave	OFF	-0.56	-3.20	-3.36
Dryer	ON	-1.96	-4.32	3.25
Dryer	OFF	-0.71	2.06	-1.17
Washer	ON	1.08	4.39	1.94
Washer	OFF	1.83	-1.69	-8.03

5.5.4 Verification of event association based on laboratory data

To further test the effectiveness of the proposed event association approach specifically, a space heater was tested in the laboratory. To bring in the interference from other appliances' events, a lamp and a microwave were also connected to the same power supply bar which the space heater was also connected to. Then the aggregated signal of the power supply bar was measured under 4 types of scenarios, each for different times:

Scenario 1: The heater ran with only its heating function on; this scenario was conducted 5 times.

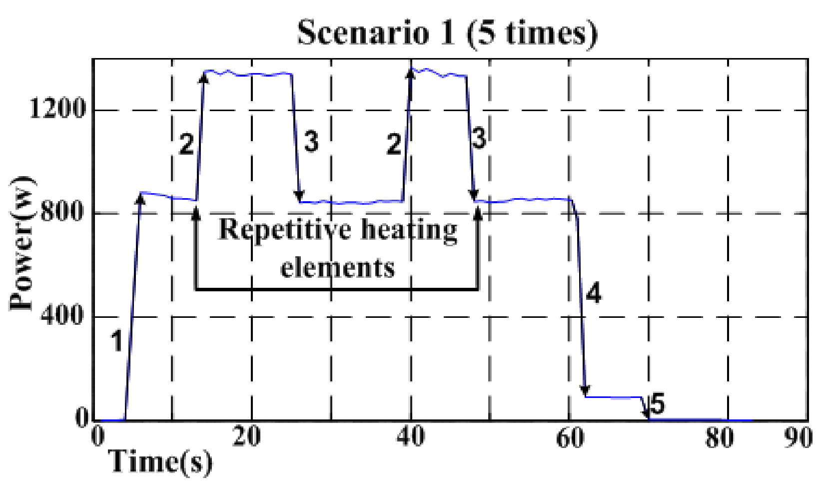


Figure 5.21: Laboratory switching experiment---Scenario 1

Scenario 2: The heater ran with both its heating function on and swaying function on (constantly changing the wind direction); this scenario was conducted 4 times.

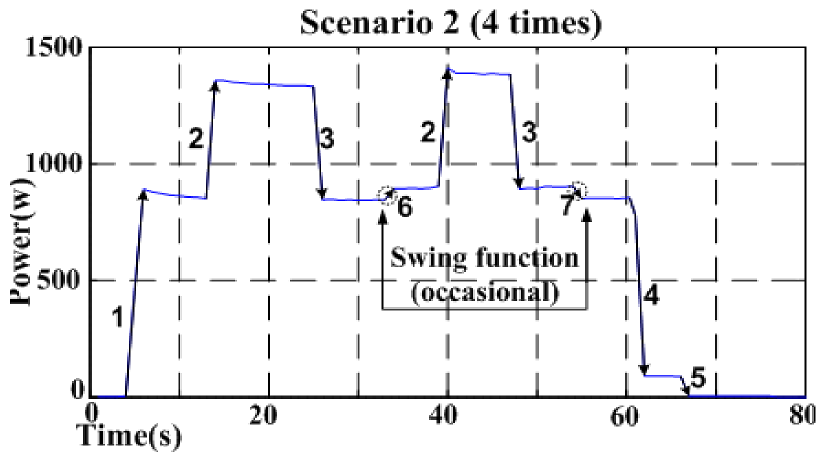


Figure 5.22: Laboratory switching experiment---Scenario 2

Scenario 3: The heater ran with only its heating function on; also, the microwave was switched on in the middle of the heater's operation; this scenario was conducted once.

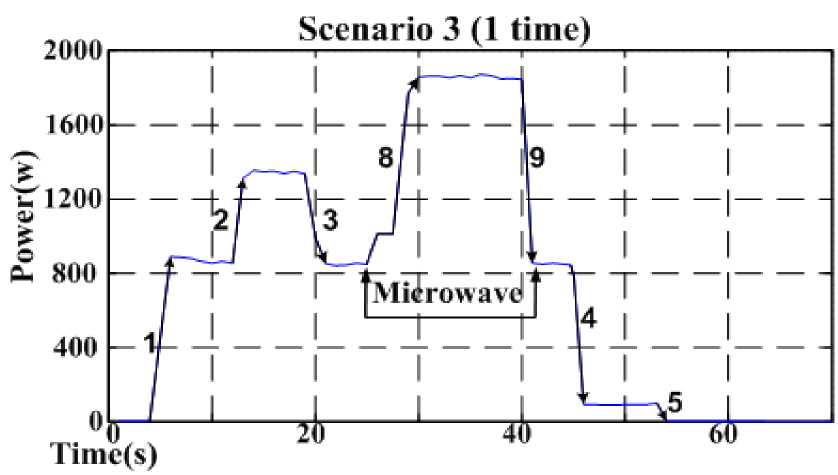


Figure 5.23: Laboratory switching experiment---Scenario 3

Scenario 4: The heater ran with only its heating function on; also, lamp and microwave were both switched on in the middle of the heater's operation; this scenario was conducted twice.

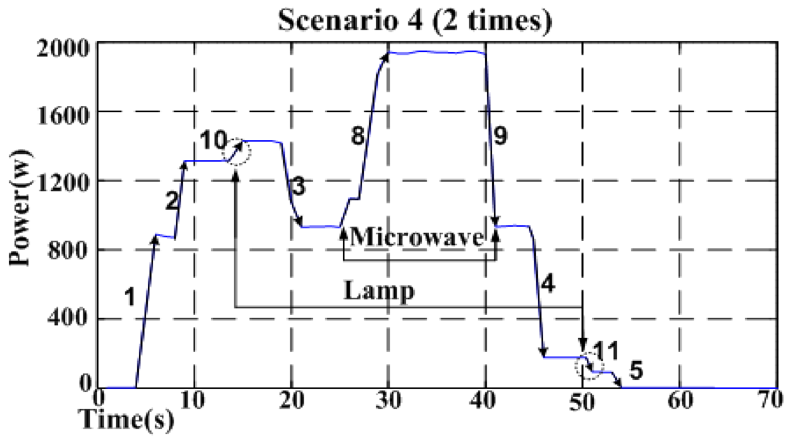
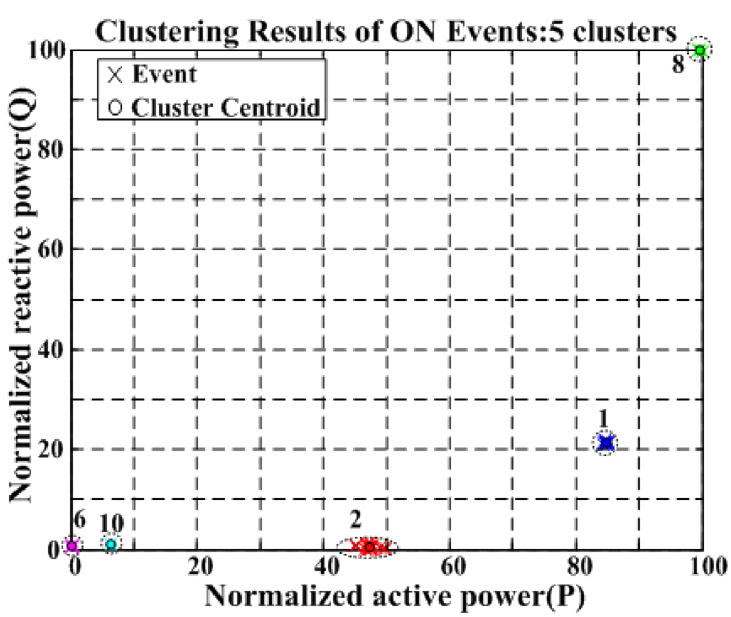
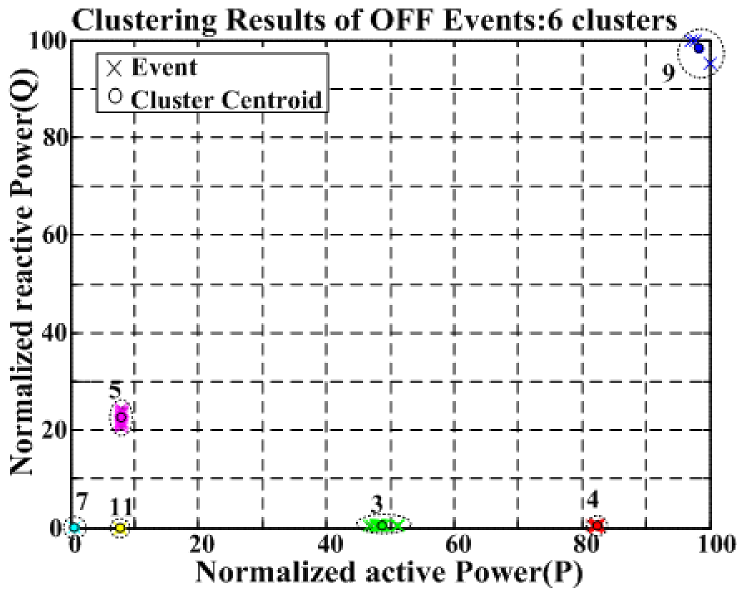


Figure 5.24: Laboratory switching experiment---Scenario 4

The above 4 scenarios simulate not only occasional events (such as those from the optional sway function) but also unrelated events caused by other appliances (in this case, microwave and lamp). In addition, the heating elements of the heater itself can be triggered repetitively. The purpose of this experiment was to test the capability of the algorithm to distinguish all types of event association for a given appliance. Examples of the measured power signals of the above 4 scenarios are plotted from Figure 5.21 to Figure 5.24. All events are also marked by different numbers according to their physical causes.



(a) Clustering results of ON events



(b) Clustering results of OFF events

Figure 5.25: Clustering results of all evens in space heater's 12 segments

Event clustering was applied to the above four scenarios, which included 12 data segments in total. The clustering plots for both the ON and OFF events are shown in Figure 5.25. The numbers marked around clusters are consistent with the numbers marked around events in Fig.12. Also, according to the number of events each cluster owns, event association is judged as shown in Table 5-13. The table shows that the determined event pattern 1→2→3→4→5(6→7) is completely consistent with the physical observations of the heater shown from Figure 5.21 to Figure 5.24 (~~~: repetitive; bracket: occasional event).

Overall, the above verification procedures and discussions show that the proposed signature extraction approach is capable of providing accurate electric and event pattern signatures for specific appliances automatically as long as the amount of feeding data is sufficient(more than a week). With this knowledge, most non-intrusive load identification algorithms can be adopted without intrusive measurements for the training or registration stages. In other words, by combining proposed NISE with a NILM method such as [104], a truly non-intrusive load monitoring solution can be provided.

Table 5-13 Event association judgment For heater(b=0.3,c=0.8)

Item	Number of apperance	Criteria used from Table IV	Event association type
Data segment	M=12	---	---
Event 1	N=12	cM<=N<=M	Single event
Event 2	N=21	N>=cM & n>=2	Repetitive event
Event 3	N=21	N>=cM & n>=2	Repetitive event
Event 4	N=12	cM<=N<=M	Single event
Event 5	N=12	cM<=N<=M	Single event
Event 6	N=4	bM<=N<cM	Occasional event
Event 7	N=4	bM<=N<cM	Occasional event
Event 8	N=3	N<bM	Unrelated event
Event 9	N=3	N<bM	Unrelated event
Event 10	N=2	N<bM	Unrelated event
Event 11	N=3	N<bM	Unrelated event

5.6 Summary

This chapter addressed a novel problem related to the NILM research---the non-intrusive extraction of load signatures. Although most previous NILM studies addressed the identification or classification of load activities, the non-intrusive signature extraction of loads haven't obtained enough research efforts yet. In reality, an intrusive signature extraction process of loads can significantly impact the application of NILM to ordinary households. Thus, more research attentions should be drawn to solve this problem. The proposed approach is an unsupervised non-intrusive approach which can automatically extract load signatures by using the meter-side data and requires almost zero effort from users. The intention of this research was to eliminate or at least reduce the intrusive work load required by most existing NILM methods. The proposed approach uses event filtration, clustering and association to locate suspect events, determine authentic events and associated different events together to reconstruct operation cycle for an objective appliance, respectively. These reconstructed cycles contain most of the electric signatures and event pattern signatures of appliances and are enough for existing NILM approaches for identification and energy tracking purposes. The proposed approach was verified off-line by using the data acquired from 3 actual residential houses and a laboratory experiment. Both electric and event pattern signatures were tested, analyzed and discussed. Generally, the accuracy of the extracted signatures was satisfactory.

This chapter focused on major appliances. Future research could study the smaller or unique appliances. Meanwhile, event filtration conditions can be updated with vast measurements in different geographic regions.

Chapter 6

Energy Estimation of Residential House

This chapter addresses an important application of NILM--- Energy estimation, specifically for residential loads. According to different types of energy components in a house, energy estimation methods for ordinary appliances, specific load group (such as incandescent lights) and background power are respectively discussed. Based on these methods, energy characteristics of several houses are analyzed and presented. The meaning and implications of the results are discussed in the end.

6.1 Overview

The above chapters have systematically discussed the identification of load activities using NILM. With the identified activities, many applications can be achieved. Energy estimation is an important area that is worth studying. It can help the end-users to gain an insight into the usage pattern of residential loads. A good understanding of the usage pattern may further support the decision making on energy management. For the utility side, with this information, the effectiveness of demand response program and Time-Of-Use (TOU) price can also be analyzed.

The overall residential energy consumption in a residential house can be divided into 3 components: ordinary appliances, specific load groups and vampire power:

- Ordinary appliance activities are usually the ones that can be tracked based on NILM's identification results. Thus their energy can be estimated accordingly. This chapter firstly discusses energy estimation methods for this type of energy component.

- Specific load group is the group of appliances that is better to be dealt with as a whole together. Examples are the incandescent light group and fluorescent light group. There are so many lights in a house and it is thus difficult to label, register and identify each individual light. However, as a group, since they share some particular characteristics in common, it is easier to use a specific estimation method to deal with. In this chapter, an independent method to estimate the incandescent light is proposed and discussed.

- Background energy refers to the electric energy consumed by appliances while they are switched off or kept in a standby/always-on mode. For example, some appliances have remote control and digital clock features. Also, some small appliances such as wireless router always stay on in a house and they should also be considered as background energy. This chapter also proposes the estimation method for this type of energy.

Finally, based on the above components, energy characteristics of several residential houses are synthetically analyzed and the findings are also presented. Their implications of the residential house energy characteristics are also discussed with respect to the Time-of Use price and load-shift actions.

6.2 Energy estimation methods for ordinary appliances

After the activities of ordinary appliances are identified either manually or by using NILM algorithms, their energy estimation can be achieved by using three different ways. They will be respectively discussed.

A. Energy estimation using average energy consumption

This method takes the operations times of appliances and its average energy consumption into consideration. The calculation is pretty simple but less accurate. It can be presented in the following formula:

$$E = nE_{ave} \tag{6.1}$$

Where E_{ave} is the average energy consumption of one time operation and n is the operation times that are detected. E_{ave} can be averaged based on several operation samples.

It is a preferred method when:

- The NILM algorithm can only provide the schedule information of appliances. For example, in [61]-[62], since the NILM identification can only identify the ON-event transients of appliances, energy cannot be accurately calculated without knowing the OFF event. This method, however, can provide some estimation results instead.
- The operation process of appliance is relatively fixed or stable. For example, some loads such as HVAC fan and refrigerator are programmed to operate similarly each time. The actual energy consumed within a period can be very close to the result obtained from equation (6.1).

B. Energy estimation using average power

For a certain appliance, this method takes the ON/OFF time and the average power of each operation into consideration. It can be mathematically presented as below:

$$E = \sum_{n=1}^{N} (t_n^{off} - t_n^{on}) P_{ave} \tag{6.2}$$

Where t_n^{on} and t_n^{off} are the ON/OFF time of a certain time operation; N is the total operation times for energy estimation; P_{ave} is the average power and it can be calculated based on several operation samples.

For single-state appliance, P_{ave} can be chosen from the real power signature of the steady-state or the ON/OFF event; for a continuous varying ON/OFF appliance such as a fridge, P_{ave} can be approximated as:

$$P_{ave} = \frac{(P_{on} + P_{off})}{2} \tag{6.3}$$

where P_{on} and P_{off} are the real power signatures of ON and OFF events.

As for the multi-state appliances, the following universal equation can be used no matter what type of appliance it is:

$$P_{ave} = \frac{\sum_{n=1}^{N} E_n}{\sum_{n=1}^{N} (t_n^{off} - t_n^{on})} \tag{6.4}$$

where N is the number of operation samples that are taken into consideration in order to obtain P_{ave}; t_n^{on} and t_n^{off} are the ON/OFF times of each operation sample.

Generally, this method has a good performance for all types of appliances as long as enough samples N are considered in (6.4). However, this method may be less accurate than the third method as follows.

C. Energy estimation using all window events

Compared to method B, this method uses all the events belonging to a certain appliance, which includes not only the ON/OFF events but also the middle events. In addition, this method uses integration of power points to accurately track the power variations between events. This method requires the time information of all events inside an event-window and is only applicable to specific NILM methods such as the proposed event-window based method.

As shown in Figure 6.1, the objective appliance has two related events #2 and # 4 while #1 and #3 are events coming from other appliances. The energy of this appliance is therefore the area summation of A1 and A2. In this case, even if the power curve between event #2 and event #3 is not constant, it can still be calculated by adding all the acquired power points between event #2 and event #3 together with the 270W base power before event #2 subtracted.

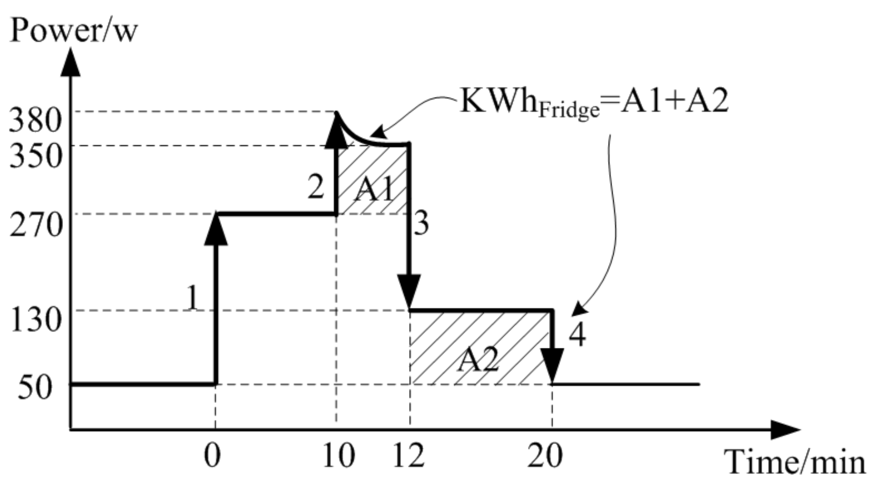


Figure 6.1: Example of energy estimation using all window events

Mathematically, it can be expressed using the following equation:

$$E = \sum_{n=1}^{N} \sum_{m=1}^{M-1} [\sum_{t=t_n^m}^{t_n^{m+1}} P(t) - P_{ref}(t_n^m)(t_n^{m+1} - t_n^m)] \tag{6.5}$$

Where:

N is the total number of operations that are taken into energy calculation;

M is the number of events that are included in an event-window. In the example shown in Figure 6.1, it is three;

P(t) is the acquired total power at a given instant t;

t_n^m is the time instant where event m occurs;

t_n^{m+1} is the time instant where event $m+1$ occurs;

P_{ref}(t) is the reference power level which may change with time. In this example, at the beginning, it is the initial power before the first event (270W). Once an unrelated event such as event #3 occurs, P_{ref}(t) will be updated as the value of initial power plus this event's power. For example, $P_{ref}(10)$ to $P_{ref}(12)$ is 270W; after 12th min, $P_{ref}(12)$ to $P_{ref}(20)$ is 270+(-220)=50W.

To summarize, method A is the simplest and fastest but also the least accurate method. It is especially suitable when the target load has a fixed operation process or when only its schedule information is known; method B is applicable to all types of appliances and it makes a balance between accuracy and simplicity; method C is the most accurate method but it requires the time instant information of all window events. Also, it is the most complicated method.

6.3 Energy estimation method for incandescent lights

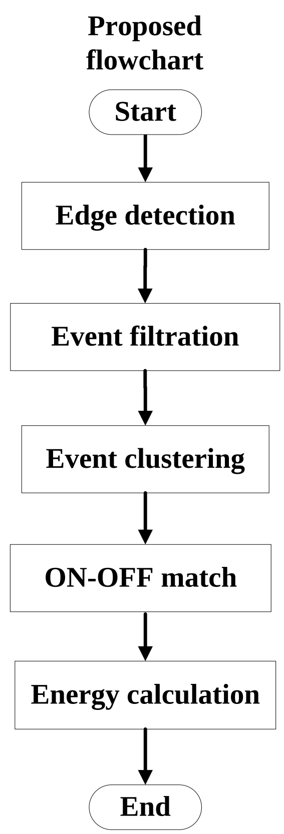


Figure 6.2: Flowchart of Energy estimation methods for incandescent lights

The intention of the proposed method is to obtain the energy consumption of all the incandescent lights (IL) for a given period, for example, a week in a typical residential house. It is better to treat all the IL as a special load group instead of treating each light individually. The main technique the method bases on is the mean-shift clustering. The proposed method consists of 4 core steps: event filtration, event clustering, ON-OFF match and energy calculation Some ideas used in this method can be referred back to Chapter 4. The flow chart is shown in Figure 6.2. Energy estimation for other load groups may also be referred to the incandescent light that is explained as follows.

6.3.1 Event filtration

All the events in a house are firstly captured. Then according to the unique features that most incandescent lights have in common, suspect events of all ILs are picked out. This step is implemented through a filtering step---only the specific events, which can satisfy a certain set of filtration conditions, are further determined as the suspect events of IL. An example of filtration conditions used for IL is listed as below:

- Active power: 30-300W
- Reactive power: <10 var (IL is a resistive load)
- Harmonic level: THD< 0.2 (IL is linear)
- Time of operations: 5:00PM---8:00AM of the next day
- Single-phase connection (all 120V based)

The above conditions can be adjusted according to the actual system environment. For example, depending on different seasons and regions, the length of daylight might be different. Daylight is closely related to the operation of lighting devices. Also, if the power level of incandescent light bulbs in a house is known, the active power range can be refined. For the other conditions such as reactive power, harmonic and single-phase connection, they stay roughly the same for all types of incandescent lights.

6.3.2 Event clustering

Event clustering is then applied to the suspect events of IL obtained by the previous step. The purpose of event clustering is to make clear how many groups/types of ILs there are in this house. Since clustering algorithm is based on common electric characteristics of lights, the events from the same type of IL can be grouped as a “cluster”.

The model used for IL is a single-state load model which has the same absolute electric characteristics for both its ON event and OFF event. The only difference is that their signs are opposite to each other. It implies in an event cluster that is caused by an IL, the number of its OFF members should be approximately equal to the number of its ON members.

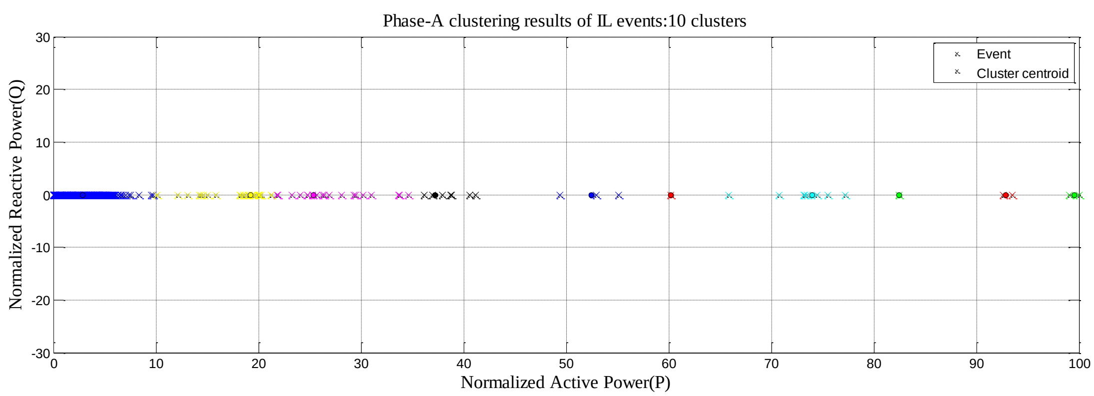


Figure 6.3: Example of clustering results of IL events.

Mean-shift clustering is applied to all the ON events and OFF events together and only their absolute event signatures are considered. In other words, both ON and OFF events of a certain appliance should be found inside a single cluster. During clustering, only active power feature is considered. This is because its reactive power and harmonic contents are too small and can be noises. To rule out the noisy impact from Q, the nominal values of Q of all events are set to zero. A typical clustering map for one phase is shown in Figure 6.3.

It can be seen from the clusters in different colors, there are likely 10 (kinds of) incandescent lights at this phase. Among these clusters, only the ones which satisfy:

- Member number is larger than 4;
- Owns both ON and OFF events.

are selected and taken into the following steps. This is because as a lighting device, it is expected the frequency of light activities is sufficient; also, as a single-state load, it is expected to see both ON and OFF events.

6.3.3 ON-OFF match

This step digs into a certain cluster obtained from the above step and examines its inside ON-OFF events. The purpose is to check if they have roughly the same number of appearances. The philosophy behind this is that ***theoretically, most of incandescent lights (simply resistors) are single state appliances of which an ON event should always match an OFF event***. In reality, exceptions may occur due to the following reasons:

1. Not all the ON and OFF events are successfully captured. Some may be lost;

2. Events coming from other appliances are mistakenly included in. For example, for cluster A, we find 10 ON events and 12 OFF events. It can still be an incandescent light, but two (12-10=2) OFF events could result from a third appliance, for example the middle events of a furnace.

When one of the above is true, mismatch on numbers of ON and OFF will happen.

Hypothetically, if all true ON and OFF events can be identified and selected, accurate energy calculation can be done. However, it is impossible to do so. Alternatively, as a statistic estimation method, all possible combinations of ON and OFF events can be calculated and an energy distribution can be obtained. An example is shown as below:

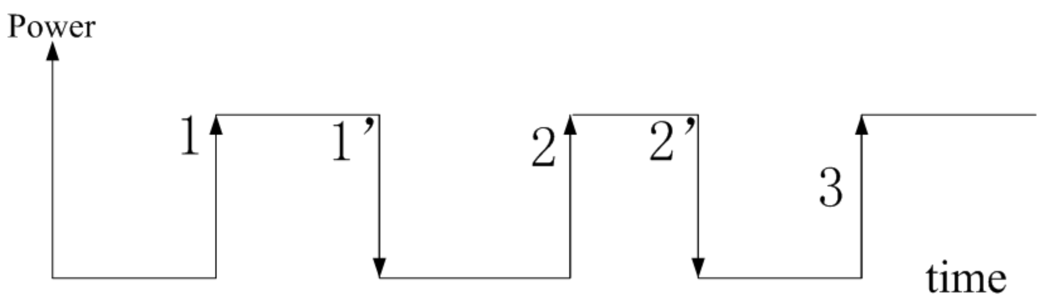


Figure 6.4: ON/OFF pattern of light A

Light A has the above pattern in a given studied period. It has 3 possible ON events (1,2,3) but only 2 possible OFF events (1' and 2'). For the 3 ON events, any two of them can be selected as a combination. Thus, there are 3 possible combinations (1,2), (1,3) and (2,3). Each of these combinations can be matched with its OFF event--- it only has one combination (1',2'). The 3 ON combination and the OFF combination can form the following 3 ON-OFF matches:

Table 6-1 3 possible ON-OFF matches of light A

Match index	ON events	OFF events	Energy
1	(1,2)	(1',2')	0.5 kWh
2	(1,3)	(1',2')	0.6 kWh
3	(2,3)	(1',2')	-0.3 kWh

The above 3 matches will result different energy results. Thus, an energy distribution can be formed.

Generally speaking, for a cluster which has *m* OFF events and *n* ON events ($m>n$), the total number of possible combinations can be calculated using factorial:

$$N_{cmb} = \frac{m!}{n!(m-n)!} \tag{6.6}$$

It should be noted, when *(m-n)* is large, N_{cmb} can be a very large number and obtaining all the N_{cmb} combinations can result in a heavy computation burden. Alternatively, in actual implementation, 10000 different ON-OFF matches are randomly generated and they can roughly represent all the possible combinations.

In addition, some ON-OFF match can result in a negative energy such as the #3 match shown in Table 6-1. It is because the selected OFF events are before the selected ON events. This kind of matches should be ruled out from the possible ON/OFF matches because the energy is not allowed to be negative.

6.3.4 Energy calculation and the distribution plot

It should be noted, for a given ON-OFF match, paring of ON and OFF event will not make a change on the energy calculation result. Still taking the above light A as example, for match#1---(1,2)→(1',2'), we have:

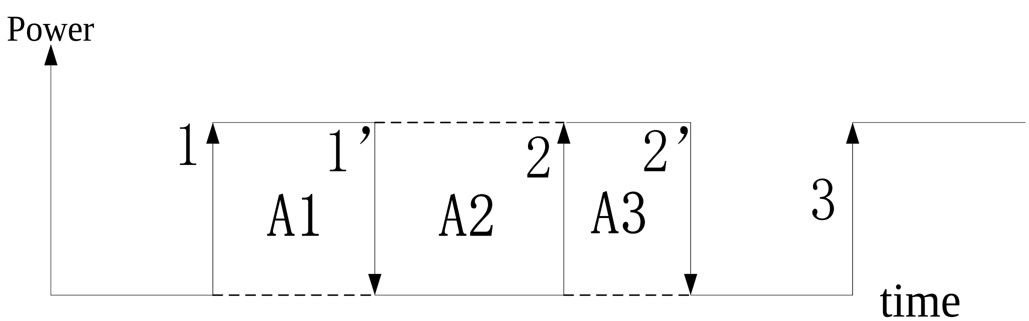


Figure 6.5: Energy blocks of light A

Scenario 1: if 1 is paired with 1' and 2 is paired with 2', its energy is

$$E_1 = A_1 + A_3 \tag{6.7}$$

Scenario 2: if 1 is paired with 2' and 2 is paired with 1', its energy is

$$E = (A_1 + A_2) - (A_3 + A_2) = A_1 + A_3 = E_1 \tag{6.8}$$

As can be seen, it can be concluded that no matter which ON is paired with which OFF, as long as the ON-OFF match is determined, the energy is always the same. In this example, match #1's energy is always E_1 no matter which ON pairs with which OFF. But it can be different from match #2 and match #3 since events from different matches can happen at different times.

Hence, the energy calculation is very simple:

1. For each qualified cluster, calculate the energy consumption for all possible matches;

2. Add the energy of all the qualified clusters together to get overall IL energy values in this house. In other words, an estimation of energy distribution can be obtained. This is shown in the following section.

6.3.5 Results

A. House #1

In house #1, 6 days in September are processed. The estimate of energy distribution, which is due to different possible ON-OFF matches, is shown in Figure 6.6. In total, 10000 random generated ON-OFF matches are taken into calculation. In this histogram, the width of per unit interval is 0.1kWh.

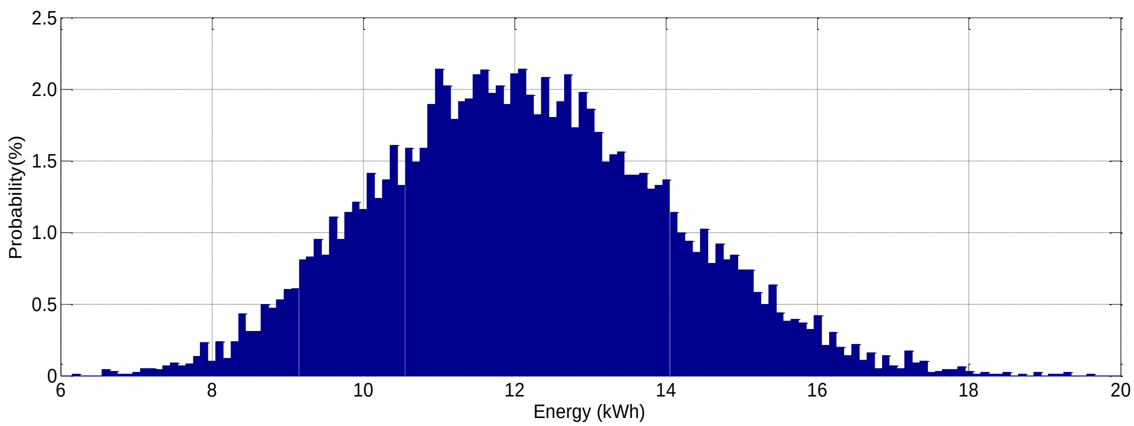


Figure 6.6: Energy distribution of IL in house #1

The minimum, mean and maximum values from the above distribution are shown as below:

$$\begin{cases} E_{\min} = E_{\min}^{phA} + E_{\min}^{phB} = 1.24 + 3.45 = 4.69kWh \\ E_{mean} = E_{mean}^{phA} + E_{mean}^{phB} = 6.52 + 5.60 = 12.12kWh \\ E_{\max} = E_{\max}^{phA} + E_{\max}^{phB} = 11.66 + 10.95 = 22.61kWh \end{cases} \quad (6.9)$$

The mean energy for IL is 12.12 kWh. This value is 10.54% of the total energy of this house in the 6 days, which is about 115 kWh.

B. House #2

In house #2, 7 days in June are processed. The energy distribution that is due to different possible ON-OFF matches is shown in Figure 6.6. In total, 10000 random generated ON-OFF matches are taken into calculation. In this histogram, the width of per unit interval is 0.1kWh.

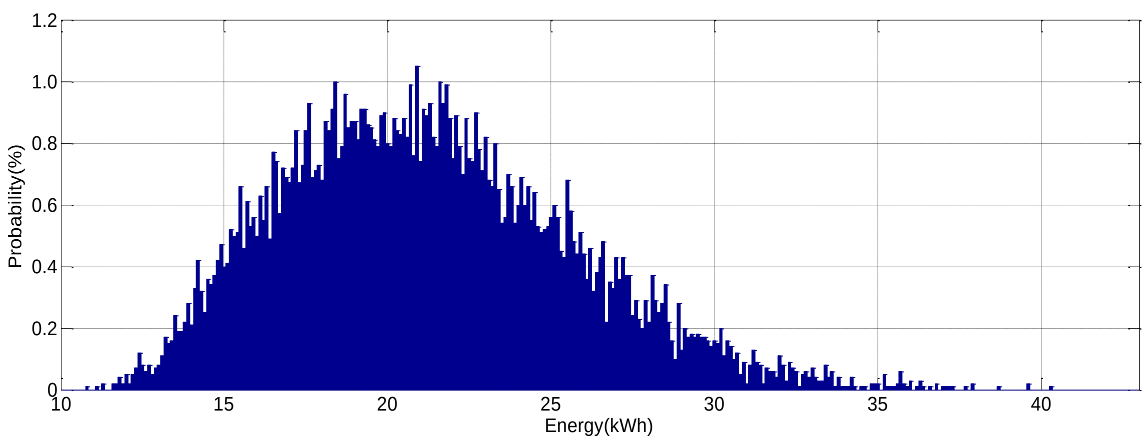


Figure 6.7: Energy distribution of IL in house #2

The minimum, mean and maximum values from the above distribution are shown as below:

$$\begin{cases} E_{\min} = E_{\min}^{phA} + E_{\min}^{phB} = 5.40 + 5.02 = 10.42kWh \\ E_{mean} = E_{mean}^{phA} + E_{mean}^{phB} = 10.61 + 12.68 = 23.29kWh \\ E_{\max} = E_{\max}^{phA} + E_{\max}^{phB} = 20.14 + 23.42 = 43.56kWh \end{cases} \quad (6.10)$$

The mean energy for IL is 23.29 kWh. This value is 16.88% of the total energy of this house in the 7 days, which is about 138 kWh.

Generally, the above results comply with the statistic results obtained in [114],[115]. Unfortunately, there is no feasible measurement method to accurately validate the above results.

6.4 Energy estimation method for background energy

6.4.1 Minimal power based method

A simple way to estimate the background energy is proposed. Since background power is the summation of the stand-by power and always-on load power, the minimal power point when the other loads are all turned off is considered as the background power.

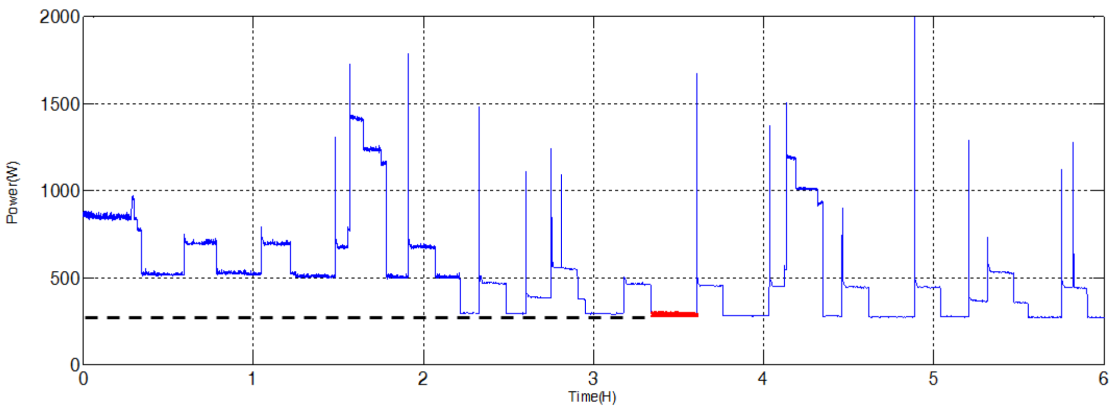
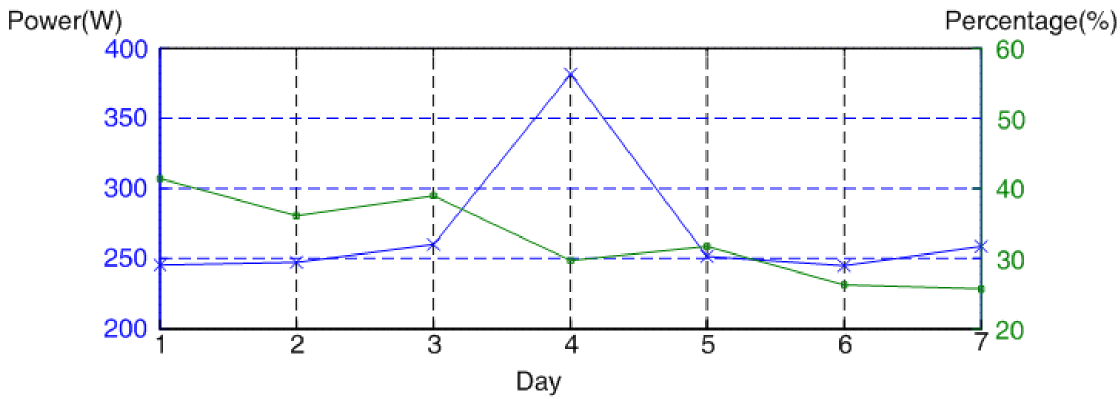


Figure 6.8: Example of background power

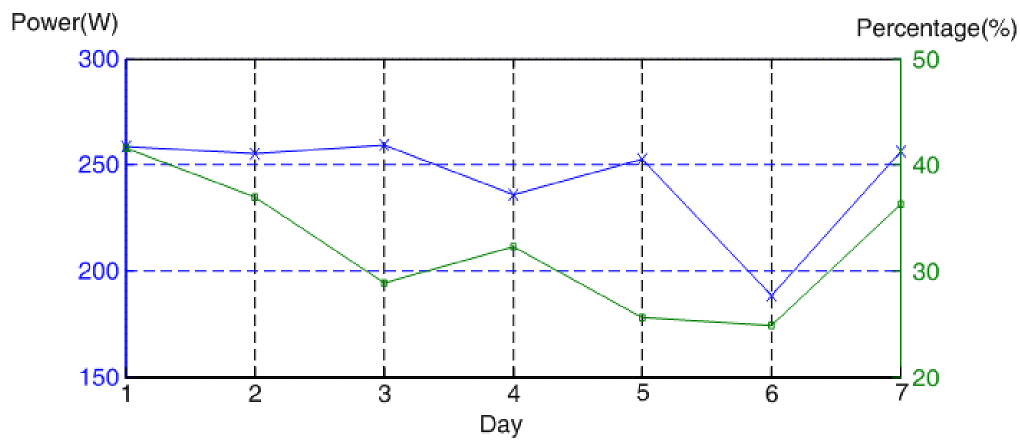
As shown in Figure 6.8, during sleeping time, most of the appliances are turned off, the minimum power found is defined as the background power. Generally, to locate the background power, one can search the lowest power from 0:00AM to 6:00 AM. Background energy is calculated by multiplying the located lowest power by 24 hours per day.

6.4.2 Results

The above method is applied to two houses and each for 7 days. The extracted background power is shown as below:



(a) House #1



(b) House #2

Figure 6.9: Background power extracted from 2 houses

As can be seen, the background power is an important energy component in a house. Its power can be larger than 200W and can take up 20% to 40% of the average total power.

6.5 Energy estimation of residential houses

6.5.1 House #1

After all the energy components are estimated using the above introduced methods, the total energy of a house can be broken down into different parts. A pie-chart of energy consumption for a week in a residential house is plotted and shown in Figure 6.10.This residential house is located in Edmonton, Canada and the processed week is in June. Please note, other lights such as fluorescent lights and compact fluorescent lights are roughly estimated as ¼ of incandescent light according to their energy efficiency [116]-[118].

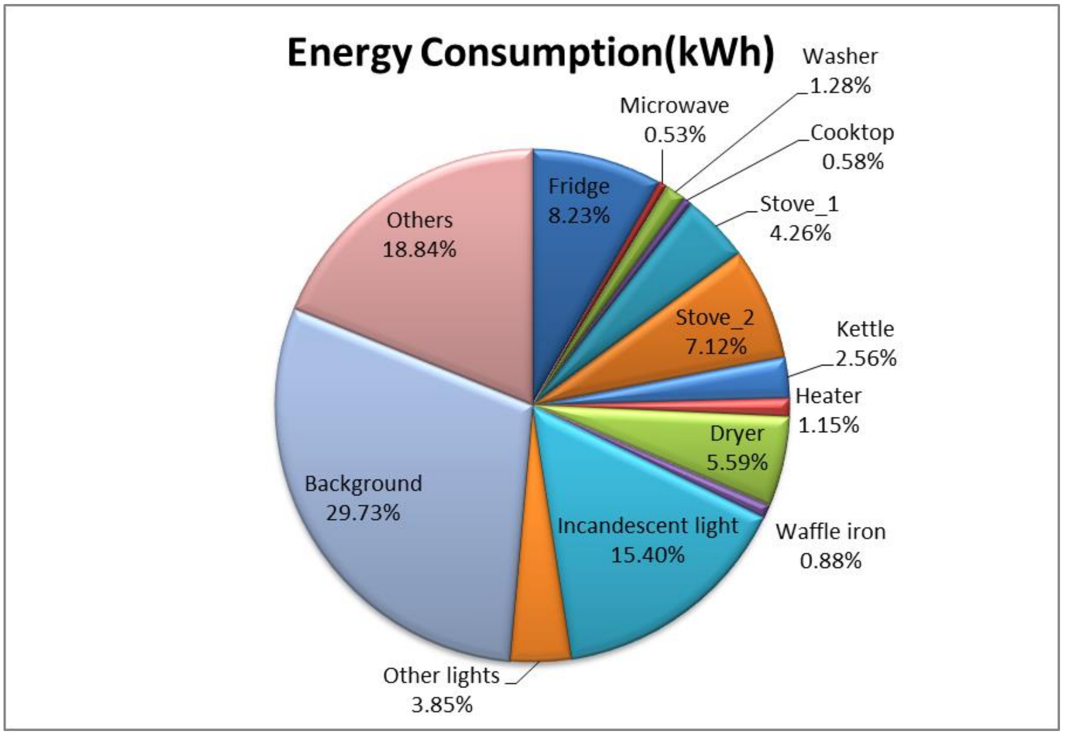


Figure 6.10: Energy consumption pie-chart for house #1

Assuming the electricity price per kWh is 9.9 Canadian cents (a typical mid-peak price according to [16]), the electricity expenses in terms of dollars spent by major individual appliances, appliance group and background power can also be obtained. The results are shown in Table 6-2.

Table 6-2 Energy consumption for house #1

Appliance	Energy Consumption(kWh)	Percent (%)	Dollar
Fridge	11.34	8.23	1.47
Microwave	0.73	0.53	0.10
Washer	1.76	1.28	0.23
Cooktop	0.80	0.58	0.10
Stove(low power)	5.87	4.26	0.76
Stove(high power)	9.80	7.12	1.27
Kettle	3.53	2.56	0.46
Heater	1.59	1.15	0.21
Dryer	7.70	5.59	1.00
Waffle iron	1.21	0.88	0.16
Incandescent light	21.21	15.40	2.76
Other lights	5.30	3.85	0.69
Background	40.95	29.73	5.32
Others	25.95	18.84	3.37
Total	137.74	100.00	17.91

As can be seen, a single fridge takes a large portion of energy which is 8.23%. Although the fridge’s power in this house is only about 140W, it is running throughout 24 hours a day and the accumulated energy is significant.

Some kitchen appliances like microwave, kettle and waffle iron (in total 3.97%) do not consume a lot although they have a large power (>1000W). This is because for each time operation, it only lasts for minutes or sometimes even shorter.

Other kitchen appliance such as stove (15.67%) contributes a lot to the total energy consumption because it is connected to double live phases and its cooking time lasts much longer, for example, 30 minutes.

For laundry activities, washer does not consume a lot while dryer consumes an important portion of energy (5.59%).

Lighting is another significant energy component. Incandescent lights along with other lights consume 19.25% of the total energy.

"Background energy" and "others" take almost a half of the house's total consumption including stand-by power, always-on appliances and all the other appliances such as TVs, computers and dishwasher which have not been registered or processed by NILM.

6.5.2 House #2

House #2 is another residential house located in Edmonton, Canada. A week in May was processed. The results are shown in Figure 6.11 and Table 6-3.

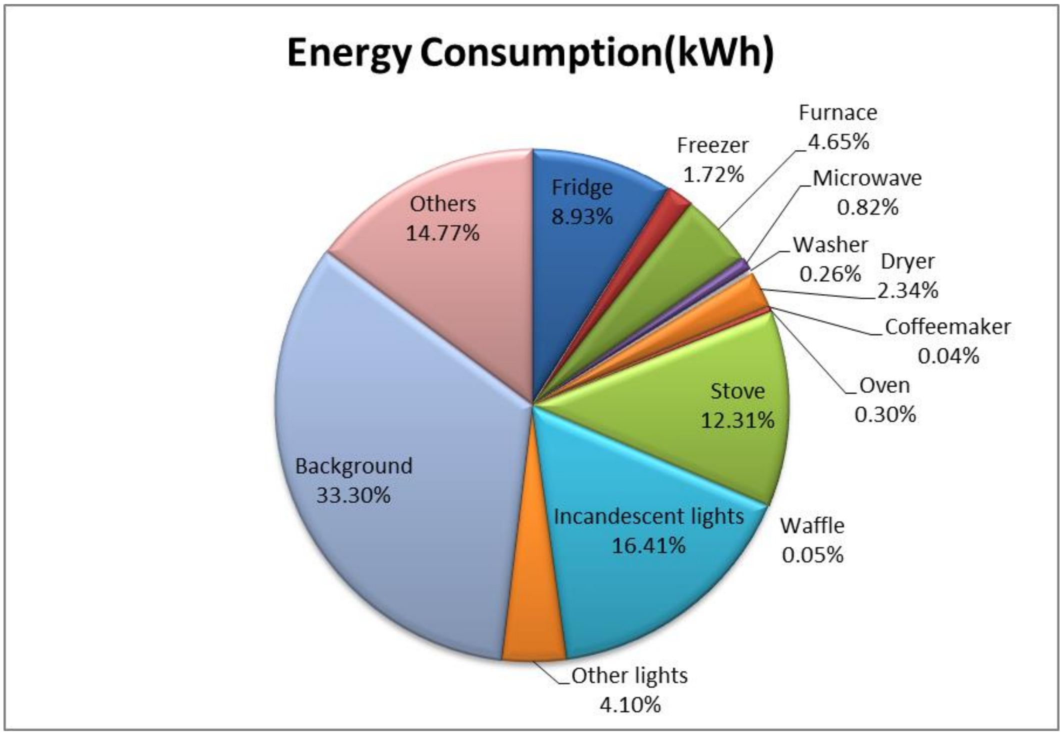


Figure 6.11: Energy consumption pie-chart for house #2 in spring

Table 6-3 Energy consumption for house #2 in spring

Appliance	Energy Consumption(kWh)	Percent (%)	Dollar
Fridge	10.98	8.81	1.09
Freezer	2.12	1.70	0.21

Furnace	5.72	4.59	0.57
Microwave	1.01	0.81	0.10
Washer	0.32	0.26	0.03
Dryer	2.87	2.31	0.28
Coffeemaker	0.05	0.04	0.00
Oven	0.37	0.30	0.04
Stove	15.14	12.15	1.50
Waffle iron	0.06	0.05	0.01
Incandescent lights	20.18	16.19	2.00
Other lights	5.04	4.05	0.50
Background	40.95	32.85	4.05
Others	18.16	14.57	1.80
Total	124.66	100.00	12.34

As can be seen, in May, furnace is still used in this house due to the cold temperature in Edmonton. It consumes 4.59% of the total energy. Compared to fridge and freezer (10.51%), it consumes less because its frequency of operation is lower than the frequency of fridge and freezer.

Similar results for kitchen appliances and lighting can be found compared to house #1. The consumption by dryer is smaller than house #1 and it is possibly due to smaller washing load in this house. Also, some clothes may be dried outside in the sun.

The percentage of “background energy” and “others” (47.42%) is also similar to house #1.

6.5.3 Seasonal changes of house #2

A week in late October from House #2 was also processed in comparison with spring data to reveal the seasonal changes. The results are shown as below:

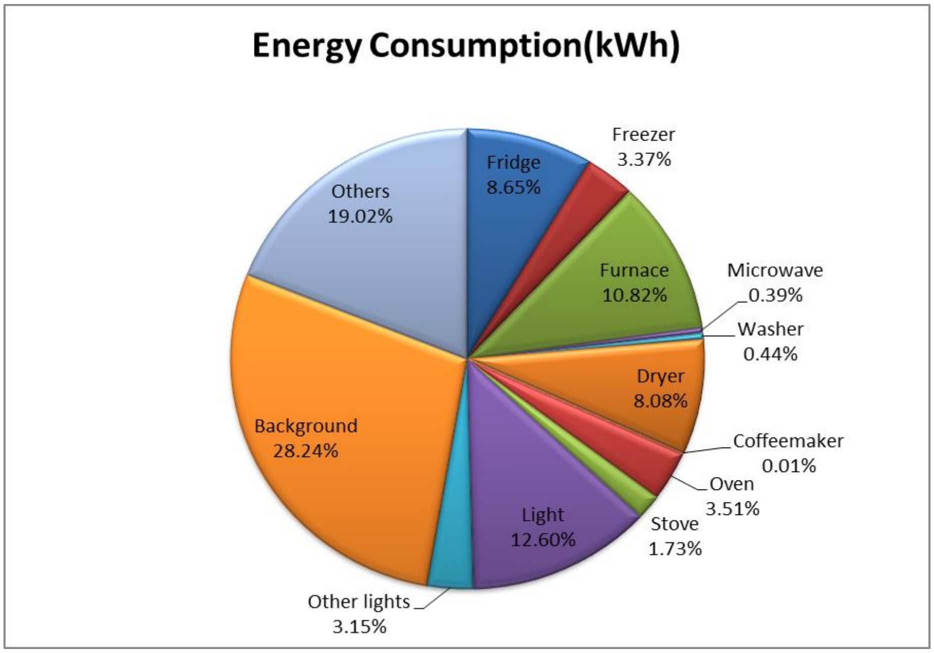


Figure 6.12: Energy consumption pie-chart for house #2 in fall

Table 6-4 Energy consumption for house #2 in fall

Appliance	Energy Consumption(kWh)	Percent (%)	Dollar
Fridge	11.50	8.65	1.14
Freezer	4.48	3.37	0.44
Furnace	14.38	10.82	1.42
Microwave	0.51	0.39	0.05
Washer	0.58	0.44	0.06
Dryer	10.74	8.08	1.06
Coffeemaker	0.01	0.01	0.00
Oven	4.66	3.51	0.46
Stove	2.30	1.73	0.23
Incandescent Light	16.75	12.60	1.66
Other lights	4.19	3.15	0.41
Background	37.54	28.24	3.72
Others	25.28	19.02	2.50
Total	132.93	100.00	13.16

Generally, the total energy consumed in fall is higher than the energy consumer in spring. This is partly due to the significant increase of the furnace operation. The energy consumed by furnace increases from 5.72 kWh to 14.38 kWh. This is easy to be understood since the house temperature in October is much lower than spring. In the meanwhile, the energy of fridge and freezer increases from 13.1 kWh to 15.98 kWh. Also the dryer increases significantly from 2.87 kWh to 10.74 kWh. This is probably because the washing load in fall are generally higher than spring (more clothes). Also in late October, clothes are not likely to be dried outside. The lighting and stove/oven energy consumed by this particular week is smaller than spring and this is believed to be caused by some unknown change of house owner's behaviors such as less time in the house and more dining in restaurants during the studied week.

6.6 Residential energy characteristics and its implications to TOU price

Based on the above estimated results, some common characteristics of residential electric energy consumption are summarized as below:

1. High power loads do not necessarily consume large amount of energy. Examples are microwave and kettle. This is because their operation duration and frequency can be very small.

2. Small power loads can still consume a lot of energy due to energy accumulative effect. Examples are fridge and freezer.

3. "Automatic" appliances such as fridge, freezer and furnace usually consume a lot of energy because they run throughout 24 hours per day.

4. Double-phase appliances such as stove, oven and dryer usually consume a lot of energy. There are two reasons: their power can be extremely high, for example, over 3000 watts; their operation durations are quite long, for example, 30 minutes to one hour.

5. Washer does not consume a lot of energy as originally expected. This is because modern front-load washer's power (100-200W) is not large and it is a VFD type device which does not generate a lot of heat like dryer.

6. The residential energy consumption changes in different seasons. One example is the furnace. It is also expected air conditioner used in some warmer areas has a similar change.

7. Lighting does consume a significant amount of energy. Our results are consistent with the statistic results in[115]---lighting consumes a significant part of total electrical energy worldwide. 20 to 50 percent consumed is due to lighting in homes and offices.

8. Background energy also takes a large portion of energy. According to [120], the large number of such devices and their being continuously plugged in resulted in energy usage of 8 to 22 percent of all appliance consumption in different countries, which is 32 to 87W. On top of the stand-by power, there are also some always-on devices such as the wireless router and our data logging devices used at the meter-side (about 100W)

Demand Periods	RPP Time-Of-Use Price	Winter Season November 1, 2012 To April 30, 2013
On-Peak	11.8 cents/kWh	**Weekdays:** 7 a.m. to 11 a.m. 5 p.m. to 7 p.m.
Mid-Peak	9.9 cents/kWh	**Weekdays:** 11 a.m. to 5 p.m.
Off-Peak	6.3 cents/kWh	**Weekdays:** 7 p.m. to 7 a.m. **Weekends & Holidays:** All Day

Figure 6.13: A summary of TOU prices in spring, 2013 by Hydro One

In terms of the electricity bill cut and money saving, it seems the existing Time-of-Use (TOU) price may not be able to create an effective and sufficient incentive for ordinary house owners to change their electricity usage behavior. However, before, TOU and TOU based smart meters are considered as effective tools to encourage load shifting and energy saving for residential customers. According to [16], the current TOU price is shown in Figure 6.13.

Now considering house #2 as an example, in fact there are many appliances that cannot be shifted from On/Mid-Peak hours to Off-Peak hours. For example, fridge, freezer and furnace are automatically controlled according to the actual environment/chamber temperature and house owners cannot easily shift them. In addition, lighting cannot be shifted because its usage is determined by the daylight conditions.

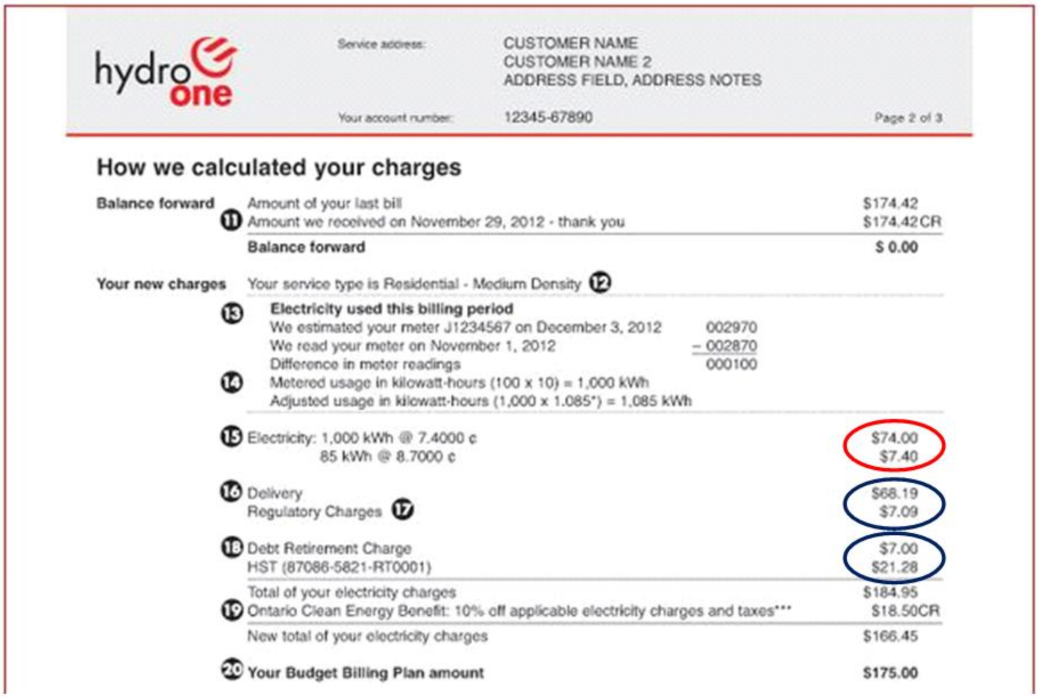
In Table 6-3, the only available appliances for load-shifting are microwave, washer, dryer, coffeemaker, oven, stove and waffle iron. Hypothetically, changing all of them from On-peak price to Off-peak price can save only $19.82 \times (11.8 - 6.3) \times 4(weeks) = 4.36$ dollars per month.

However, even for this little amount of money, the required load-shift can greatly affect customers' comfortableness and sometimes is even impossible. For example, cooking is normally always before mealtime and cannot be shifted.

On the other hand, based on Table 6-3, replacing all the incandescent lights with fluorescent/compact fluorescent lights may be able to reduce the energy by about 12%. But this requires additional money investment since fluorescent lights are much more expensive than incandescent lights [118].

In addition, a typical electricity bill from Hydro One [119] is shown in Figure 6.14. It can be seen that electricity charge ($81.40) is only 49% of the electricity bill. The delivery charge, regulatory charge and retirement charge are relatively fixed charges that do not change with the actual amount of electricity spent by the residential customers.

Overall speaking, the effectiveness of TOU prices for residential houses in order to encourage energy usage behavior change is suspected based on our results.



hydro one

Service address: CUSTOMER NAME
CUSTOMER NAME 2
ADDRESS FIELD, ADDRESS NOTES

Your account number: 12345-67890

Page 2 of 3

How we calculated your charges

Balance forward	Amount of your last bill		$174.42
	11 Amount we received on November 29, 2012 - thank you		$174.42CR
	Balance forward		**$ 0.00**
Your new charges	Your service type is Residential - Medium Density 12		
	13 **Electricity used this billing period**		
	We estimated your meter J1234567 on December 3, 2012	002970	
	We read your meter on November 1, 2012	– 002870	
	Difference in meter readings	000100	
	14 Metered usage in kilowatt-hours (100 x 10) = 1,000 kWh		
	Adjusted usage in kilowatt-hours (1,000 x 1.085*) = 1,085 kWh		
	15 Electricity: 1,000 kWh @ 7.4000 ¢		$74.00
	85 kWh @ 8.7000 ¢		$7.40
	16 Delivery		$68.19
	Regulatory Charges 17		$7.09
	18 Debt Retirement Charge		$7.00
	HST (87086-5821-RT0001)		$21.28
	Total of your electricity charges		$184.95
	19 Ontario Clean Energy Benefit: 10% off applicable electricity charges and taxes***		$18.50CR
	New total of your electricity charges		$166.45
	20 **Your Budget Billing Plan amount**		**$175.00**

Figure 6.14: Electricity billing example by Hydro One

6.7 Summary

In this chapter, different energy estimation methods were proposed for three major energy components in a residential house --- ordinary appliances, appliance group (incandescent lights) and background energy.

Using these methods, meter-side data for several weeks which were acquired from two local houses were processed. The energy estimation results were presented. Also, common characteristics were revealed and summarized.

Based on the revealed characteristics, hypothetical energy savings by load shift according to the existing TOU price was calculated and analyzed. It was found that the existing TOU may not be able to effectively encourage residential load shift behaviors after taking the customer's comfortableness and the actual saving amount into account.

Chapter 7

Conclusions and Future work

7.1 Thesis Conclusions and Contributions

This thesis has approached topics related to event detection, load signature studies, non-intrusive load monitoring algorithm, non-intrusive signature extraction algorithm, energy estimate methods and the characteristics of residential house. The thesis has presented a complete solution for power system load decomposition, especially for residential end-users. Two major closely related issues—non-intrusive load monitoring and non-intrusive signature extraction were discussed in detail. But not limited to this, some other questions related to event detection, load modeling, load group energy estimation, and residential energy characteristics were also answered in the thesis.

The main conclusions and contributions of this thesis are summarized as follow:

- Two data-segmentation based event detection methods were proposed and studied. Instead of looking for state transitions directly, the proposed methods look for stable data segments in which a certain level of signal noise is also acceptable. The proposed slope method can effectively solve the existing event detection challenges such as slope-type event, slow event and signal noise caused misdetection.

- Common special issues of event detection have received attentions in this thesis. A simple method to detect double-phase event was proposed. Also, the causes of adjacent events and event overlap issues were studied and possible solutions were discussed. It was found that increasing the data-acquisition sampling rate could effectively reduce the occurrence of the above problems.

- Based on a review of the drawbacks of existing steady-state and transient load models, an event-window load model was proposed. Studies for event-window signatures were carried out. It was found that the event-window model along with its signatures could accurately describe the whole operation process of a complex load.

- A novel event-window based NILM algorithm was proposed and discussed. The new algorithm:

    - Can effectively identify the activities of complex loads including continuous-varying and multi-state loads;

    - And does not require a complicated local training process. In addition, no re-training process is required after the load inventory is changed;

  Extensive verifications based on three houses and a public dataset were conducted. It was found the average identification rate is promising even when the aggregated signal contains a certain level of noises.

- A novel NILM related problem--non-intrusive signature extraction for major residential loads--was proposed and studied. The previous measurement based signature extraction methods were actually intrusive. This critical obstacle prevents NILM from being practically applied. Unfortunately, in the past, researchers did not pay enough attentions to this problem. In this thesis, this problem was systematically discussed, and a clustering based solution was presented.

- Energy estimation of residential houses was studied and investigated from the following perspectives:

    - Three energy estimation methods for ordinary appliances were proposed and compared.

- A clustering based energy estimation method for load groups such as incandescent lights was presented.
- An energy estimation method for background energy was presented.
- Residential energy characteristics and their implication for the Time-of-Use price were summarized and discussed. It was found that the existing TOU price might not encourage residential load shift behaviors after taking both the customer's comfortableness sacrifice and the actual amount saved into account.

7.2 Suggestions for future work

As with any study, something can always be done to extend and improve the research. Several extensions and modifications of this thesis can be explored as follows:

- Future studies can be extended into the commercial and industry compound load area. This thesis focuses mainly on residential compound loads since many commercial and industry loads have their own sub-load monitoring equipment already. However, a cost-effective non-intrusive solution may still be of interest. In the appendix, some preliminary studies for a typical commercial building were presented. More studies including extensive field data based verification can be completed in the future.
- The energy estimation method for load groups can be improved:
    - This thesis focused on the incandescent light load group. Other load groups such as the florescent load group for a residential house, the computer load group for an office building, and the HAVC load group for an industrial building can be further studied.

  - In this thesis, for the energy estimation method proposed for incandescent light group, no field data based verification was conducted because of the limited available data acquisition devices that can be used to effectively measure all the incandescent lights. This problem may be solved by using advanced in-direct sensing devices, which may also be applied to other types of load groups.

- Event-misdetection by event overlap is still a problem that could affect the accuracy of the proposed event-window based NILM algorithm. More sophisticated hardware and algorithm solution can be studied for improvement.

- Comparative studies between NILM algorithm and the statistical data based approach [34]-[35] can be conducted to clarify the advantages and disadvantages on the accuracy and easiness aspects of each approach.

- More research should focus on non-intrusive load signature extraction. More types of loads, more geographic regions and different sensitivity studies should be provided because clearly this very important area has been neglected by the previous NILM researchers.

- The energy characteristics of commercial and industry loads can be further studied, and the potential application of load decomposition to demand response and load shedding can be further discussed.

Chapter 8

References

[1] David Schatsky and Clint Wheelock, “Home Energy Management”, available online at: http:// www.pikeresearch.com/research/

[2] David Kathan and et al. , “Assessment of Demand Response and Advanced Metering”, Federal Energy Regulatory Commission, Dec. 2008

[3] S. Massoud Amin and Bruce F. Wollenberg, “Toward a Smart Grid”, IEEE Power & Energy Magazine, Sept/Oct 2005, pp. 34-41.

[4] H. Farhangi, “The Path of the Smart Grid”, IEEE Power & Energy Magazine, Jan/Feb 2010, pp. 18-28.

[5] United States Department of Energy, “The Smart Grid: An Introduction”, available online at: http://www.oe.energy.gov/SmartGridIntroduction.htm

[6] Litos Strategic Communication, “The Smart Grid: An Introduction”, United States Department of Energy, 2004.

[7] S. Rahman, “Smart Grid Expectations”, IEEE Power & Energy Magazine, Sept/Oct 2009, pp. 83-85.

[8] A. Mahmood, M. Aamir and M.I. Anis, “Design and Implementation of AMR Smart Grid System”, IEEE Electrical Power and Energy Conference, 2008, pp. 1-6.

[9] ACEEE, “Study: Smart Meters Alone Not Enough to Significantly Reduce Power Use”, available online at:

http://tdworld.com/customer_service/aceee-smart-meter-study-0610/

[10] Andy Berrisford, “New Technologies for End-use Load Research at BC Hydro”, CIGRE Conference Vancouver 2010.

[11] Gridwise, “A Smart Grid:Cost Reduction” ,Gridwise Alliance, available on line at: http://www.gridwise.org/smartgrid_costreduction.asp

[12] ComEd, “Residential Real-Time Pricing (RRTP) program overview”, available online at: https://rrtp.comed.com/about/

[13] Ameren Illinois, “Power Smart Pricing” available online at: http://www.powersmartpricing.org/

[14] IESO, “Time-of-Use prices”, available online at: http://www.ieso.ca/imoweb/siteshared/tou_rates.asp

[15] Hydro One, “Introducing Time-of-Use Prices” , available online at:

http://www.hydroone.com/TOU/Documents/Introducing_TOU_Prices.pdf

[16] Hydro One , “Time-of-Use Prices” , available online at: http://www.hydroone.com/TOU/Pages/Default.aspx

[17] Burlington Hydro, “Time of Use: A New Way to Think About Electricity”, available online at: http://www.burlingtonhydro.com/your-business/time-of-use.html

[18] Guelph Hydro , “Time-of-Use Rates for Electricity”, available online at: http://www.guelphhydro.com/en/residential/timeofuserates.asp

[19] Ontario Ministry of Energy, “Smart Meters and Time-of-Use Prices”, available online at: http://www.energy.gov.on.ca/en/smart-meters-and-tou-prices/

[20] Milind Naphade and et al., “Smarter Electricity Portal Pilot Study Report”, IBM Research, available online at: http://www.cityofdubuque.org/DocumentCenter/View/15984

[21] ISTAT, Time Use Survey, Notiziarion. 10, 1990 (in Italian).

[22] ISTAT, Household consumption: year 1987, vol. 14, 1989 (in Italian).

[23] Ipsos-RSL and Office for National Statistics, United Kingdom Time Use Survey, 3rd ed., UK Data Archive, Colchester, Essex, September 2003

[24] Ian Richardson and Murray Thomson, "CREST Domestic Electricity Demand Model", available online at: https://dspace.lboro.ac.uk/dspace-jspui/handle/2134/5786

[25] Bell, M. Swinton, M.C. Entchev, E. Gusdorf, J. Kalbfleisch, W. Marchand, R.G. Szadkowski, F., "Development of Micro Combined Heat and Power Technology Assessment Capability at the Canadian Centre for Housing Technology", pp. 48. 2003-12-08.

[26] M. Armstrong, M. Swinton, H. Ribberink, I. Beausoleil-Morrison and J. Millette, "Synthetically derived profiles for representing occupant driven electric loads in Canadian housing", Building Performance Simulation, vol.2, pp. 15–30, 2009.

[27] Pratt, R., et al. Description of Electric Energy Use in Single-Family Residences in the Pacific Northwest – End-Use Load and Consumer Assessment Program (ELCAP). DOE/BP-13795-21, Richland, WA: Pacific Northwest National Laboratory, 1989.

[28] CSA, "Energy Consumption Test Methods for Household Dishwashers", CSA Standard CAN/CSA-C373-92, 1992.

[29] CSA, "Test Method for Measuring Energy Consumption and Drum Volume of Electrically Heated Household Tumble-Type Clothes Dryers", CSA Standard CAN/CSA-C361-92, 1992.

[30] Statistics Canada, "Selected dwelling characteristics and household equipment", available online at: http://www40.statcan.gc.ca/l01/cst01/famil09b-eng.htm

[31] I. Richardson, Murray Thomson, David Infield and Conor Clifford, "Domestic electricity use: A high-resolution energy demand model," Energy and Buildings, vol. 42, pp. 1878-1887, 2010.

[32] C. Walker and J. Pokoski, "Residential load shape modeling based on customer behavior," Power Apparatus and Systems, IEEE Transactions on, pp. 1703-1711, 1985.

[33] Capasso, A.; Grattieri, W.; Lamedica, R.; Prudenzi, A., "A bottom-up approach to residential load modeling," Power Systems, IEEE Transactions on , vol.9, no.2, pp.957,964, May 1994

[34] “DOE-2”, available online at: http://gundog.lbl.gov/dirsoft/d2whatis.html

[35] Microsoft Corp., “Microsoft-Hohm”, available online at: http://www.microsoft-hohm.com/Default.aspx?wa=wsignin1.0

[36] John N. Sharood, Graham Bailey, D. Mitchell Carr, James Turner and David Peachey. “Appliance retrofit monitoring device with a memory storing an electronic signature”, U.S. Patent 6 934 862, Aug 23, 2005.

[37] Digi Smart Plug, “XBee Smart Plug” , available online at: http://www.digi.com/wiki/developer/index.php/XBee_Smart_Plug

[38] Blue line innovations, “BlueLine Electricity Power Cost Monitor”, available online at:

http://www.blueline innovations.com/Products/PowerCost-Monitor/

[39] AlertMe Energy, “How it works”, available online at: http://www.alert me.com/products/energy/how-it-works

[40] Energy Inc., “TED The Energy Detective”, available online at: http://www.theenergydetective.com/

[41] Echelon, “Smart Metering”, available online at: http://www.echelon.com/solutions/smartmetering/han_solutions.htm

[42] Engadget, “LG Thinq linqs your smart appliances with WiFi and smartphone apps“, available online at: http://www.engadget.com/2011/01/04/lg-thinq-linqs-your-smart-appliances-with-wifi-and-smartphone-ap/

[43] TriplePundit, "Study Shows Bright Green Future for Smart Appliances", available online at: http://www.triplepundit.com/2010/03/smart-appliances-could-be-the-next-big-thing/

[44] A. Zaballos, A. Vallejo, M. Majoral, and J. M. Selga,“Survey and Performance Comparison of AMR Over PLC Standards,” IEEE Transactions on Power Delivery, 2009, vol. 24, pp.604-613.

[45] Q. Y. Liu, B. Z. Zhao, Y. R. Wang, and J. Hu, “Experience of AMR Systems Based on BPL in China,” IEEE International Symposium on Power Line Communications and Its Applications, 2009, pp.280-284.

[46] IEEE Std. 802.15.4-2003, Wireless Medium Access Control(MAC) and Physcial Layer(PHY) Specifications for Low Rate Wireless Personal Area Networks[S], 2003.

[47] W. Qiu, E.Skafidas, and P. Hao, “Enhanced tree routing for wireless sensor networks,” Elsevier on Ad Hoc Networks, 2009, vol. 7, pp.638-650.

[48] K. Taehong, K. Daeyoung, P. Noseong, Y. Seong-eun and T.S. Lopez, “Shortcut Tree Routing in ZigBee Networks,” Proceedings of the Second International Symposium on Wireless Pervasive Computing, February 2007.

[49] Y. Kim, T. Schmid, Z. M. Charbiwala, and M. B. Srivastava, "ViridiScope: Design and implementation of a fine grained power monitoring system for homes", in Proc. 11th Int. Conf. Ubiquitous Comp., pp. 245–254, 2009

[50] A. Rowe, M. Berges, and R. Rajkumar, "Contactless Sensing of Appliance State Transitions Through Variations in Electromagnetic Fields", BuildSys 2010, pp. 19-24

[51] D. Jung, and A. Savvides, "Estimating Building Consumption Breakdowns using ON/OFF State Sensing and Incremental Sub-Meter Deployment". Proceedings of the 8th ACM Conf. Embedded Networked Sensor Systems, pp. 225-238, 2010

[52] Tobias Ryberg, “Smart Metering in Western Europe”, M2M Research Series, available online at: http://www.triplepundit.com/2010/03/smart-appliances-could-be-the-next-big-thing/ .

[53] Jeffery F. Fye and Mark A. Johnson, “Solid-state Electricity Meter”, United States Patent, US 6,734,663 B2, May 11 2004.

[54] Aclara TWACS Technology Specification Sheets, available online at: http://www. aclara.com/AclaraPLS/Pages/specsheets.aspx.

[55] Hart, G.W., "Non-intrusive Appliance Load Monitoring", Proceedings of the IEEE, vol. 80, No 12, pp. 1870 - 1891, 1992

[56] C. J. Andrews, “Nonintrusive monitoring device for electric loads in commercial buildings”, MIT Laboratory for Electrical and Electromagnetic Systems Technical Report, 1988.

[57] F. Sultanem, "Using appliance signatures for monitoring residential loads at meter panel level", IEEE Trans. Power Delivery, vol. 6, no. 4, pp. 1380-1385, 1991

[58] Hannu Pihala, "Non-intrusive appliance load monitoring system based on a modern kWh-meter", VTT Energy licentiate thesis, Technical research center of Finland, 1998

[59] Jian Liang, Simon K. K. Ng, Gail Kendall, and John W. M. Cheng, "Load Signature Study—Part I: Basic Concept, Structure, and Methodology", IEEE Trans. Power Delivery, vol. 25, no. 2, pp.551-560, 2010

[60] Jian Liang, Simon K. K. Ng, Gail Kendall, and John W. M. Cheng, "Load Signature Study—Part I: Basic Concept, Structure, and Methodology", IEEE Trans. Power Delivery, vol. 25, no. 2, pp.551-560, 2010

[61] R. W. Cox, S. B. Leeb, S. R. Shaw, and L. K. Norford, "Transient event detection for nonintrusive load monitoring and demand side management using voltage distortion," in Proc. 21st IEEE Applied Power Electronics Conference and Exposition (APEC), 2006

[62] S.B. Leeb, S.R. Shaw and J.L. Kirtley, "Transient event detection in spectral envelope estimates for nonintrusive load monitoring," Power Delivery, IEEE Transactions on , vol.10, no.3, pp.1200,1210, Jul 1995

[63] S.B. Leeb, S.R. Shaw and S.R.Shaw, "Multiprocessing Transient Event Detector For Use In A Nonintrusive Electrical Load Monitoring System", United States Patent, US 5,717,325, Feb. 10,1998.

[64] S. Leeb , J. L. Kirtley Jr., M. S. LeVan and J. P. Sweeney "Development and validation of a transient event detector", AMP Journal Technol., vol. 3, pp.69 -74 1993

[65] S.N. Patel, T. Robertson, J.A. Kientz, M.S. Reynolds, and G.D. Abowd, "At the Flick of a Switch: Detecting and Classifying Unique Electrical Events on the Residential Power Line", In: UbiComp 2007, pp. 271-288

[66] S. Gupta, M. S. Reynolds, S. N. Patel, "ElectriSense: Single-Point Sensing Using EMI for Electrical Event Detection and Classification in the Home," Conference on Ubiquitous Computing, pp. 139-148, 2010

[67] H.Y. Lam, G. S. K. Fung, W. K. Lee, “A Novel Method to Construct Taxonomy of Electrical Appliances Based on Load Signatures,” IEEE Transactions on Consumer Electron., vol. 53, no. 2, pp. 653-660, 2007

[68] Yi-Ching Su; Kuo-Lung Lian; Hsueh-Hsien Chang, "Feature Selection of Non-intrusive Load Monitoring System Using STFT and Wavelet Transform," IEEE 8th International Conference on e-Business Engineering (ICEBE), pp.293-298, 2011

[69] D.Srinivasan, W.S.Ng and A.C.Liew, "Neural-network-based signature recognition for harmonic source identification", IEEE Transactions on Power Delivery, Volume 21, Issue 1, pp. 398 – 405, 2006

[70] Duan, J.; Czarkowski, D.; Zabar, Z., "Neural network approach for estimation of load composition," Circuits and Systems, 2004. ISCAS '04. Proceedings of the 2004 International Symposium on , vol.5, no., pp.V-988,V-991 Vol.5, 23-26 May 2004

[71] Yu-Hsiu Lin, Men-Shen Tsai, “Application of Neuro-Fuzzy Pattern Recognition for Non-intrusive Appliance Load Monitoring in Electricity Energy Conservation”, WCCI 2012 IEEE World Congress on Computational Intelligence, pp.686-692, 2012.

[72] Suzuki, K.; Inagaki, S.; Suzuki, T.; Nakamura, H.; Ito, K., "Nonintrusive appliance load monitoring based on integer programming", SICE Annual Conference, 2008, pp. 2742 – 2747, 2008

[73] Baranski, M.; Voss, J., "Genetic algorithm for pattern detection in NIALM systems," Systems, Man and Cybernetics, 2004 IEEE International Conference on , vol.4, no., pp.3462,3468 vol.4, 10-13 Oct. 2004

[74] M.L. Marceau, and R. Zmeureanu, "Nonintrusive load disaggregation computer program to estimate the energy consumption of major end uses in residential buildings", Energy Conversion & Management, vol. 41, pp. 1389-1403, 2000

[75] Prudenzi, "A Neuron Nets Based Procedure for Identifying Domestic Appliances Pattern-of-use from Energy Recordings at Meter Panel", IEEE Power Engineering Society Winter Meeting, pp. 941-946, 2002

[76] HH Chang; Kun-Long Chen; Yuan-Pin Tsai; Wei-Jen Lee; , "A New Measurement Method for Power Signatures of Nonintrusive Demand Monitoring and Load Identification," Industry Applications, IEEE Transactions on , vol.48, no.2, pp.764-771, March-April 2012

[77] Yu-Hsiu Lin, Men-Shen Tsai, “Application of Neuro-Fuzzy Pattern Recognition for Non-intrusive Appliance Load Monitoring in Electricity Energy Conservation”, WCCI 2012 IEEE World Congress on Computational Intelligence, pp.686-692, 2012.

[78] A.G. Ruzzelli, G.M.P. O’Hare, A. Schoofs, C. Nicolas, "Real-Time Recognition and Profiling of Appliances through a Single Electricity Sensor", In 7th Annual IEEE Communications Society Conference on Sensor, Mesh, and Ad Hoc Communications and Networks (SECON’10), 2010

[79] H. Murata , T. Onoda, "Applying Kernel Based Subspace Classification to a Non-intrusive Monitoring for Household Electric Appliances", Artificial Neural Networks - ICANN 2001

[80] Marchiori, D. Hakkarinen, H. Qi, and L. Earle, “Circuit level load monitoring for household energy management,” IEEE Pervasive Computing, vol. 10, no. 1, pp. 40-48, 2011

[81] M. Berges, E. Goldman, H. S. Matthews and L. Soibelman, “Training Load Monitoring Algorithms on Highly Sub-Metered Home Electricity Consumption Data,” 12th International Conference on Computing in Civil and Building Engineering, Beijging, 2008.

[82] M. Berges, E. Goldman, H. S. Matthews, and L. Soibelman, "Learning Systems for Electric Consumption of Buildings," in ASCE International Workshop on Computing in Civil Engineering, Austin, TX, pp. 1-10, 2009

[83] M. Weiss, T. Staake, F. Mattern, "Leveraging smart meter data to recognize home appliances" IEEE Pervasive Computing and Communication (PerCom), Switzerland, 2012

[84] K. D. Lee, S.B. Leeb, L.K. Norford, P.R. Armstrong, J. Holloway, S.R. Shaw, "Estimation of Variable-Speed-Drive Power Consumption From Harmonic Content," , IEEE Trans. Energy Conversion, vol.20, no.3, pp. 566- 574, 2005

[85] W. Wichakool, A.T. Avestruz, R.W. Cox, S.B. Leeb, "Resolving Power Consumption of Variable Power Electronic Loads Using Nonintrusive Monitoring," Power Electronics Specialists Conference, IEEE PESC 2007, pp.2765-2771, 2007

[86] W. Wichakool, A.T. Avestruz, R.W. Cox, S.B. Leeb, "Modeling and Estimating Current Harmonics of Variable Electronic Loads", , IEEE Trans. Power Electronics, vol.24, no.12, pp.2803-2811, 2009

[87] S.R. Shaw, S.B. Leeb, L.K. Norford, R.W. Cox, "Nonintrusive Load Monitoring and Diagnostics in Power Systems", IEEE Trans. Instrumentation and Measurement, vol.57, no.7, pp.1445-1454, 2008

[88] U.A. Orji, Z. Remscrim, C. Laughman, S.B. Leeb, W. Wichakool, C. Schantz, R. Cox, J. Paris, J.L. Kirtley, L.K. Norford, "Fault detection and diagnostics for non-intrusive monitoring using motor harmonics", Applied Power Electronics Conference and Exposition (APEC), pp.1547-1554, 2010

[89] J. Paris, Z. Remscrim, K. Douglas, S. B. Leeb, R. W. Cox, S. T. Gavin, S. G. Coe, J. R. Haag, and A. Goshorn, "Scalability of non-intrusive load monitoring for shipboard applications," in American Society of Naval Engineers Day 2009, National Harbor, 2009

[90] R. W. Cox, P. Bennett, D. McKay, J. Paris, and S. B. Leeb, “Using the non-intrusive load monitor for shipboard supervisory control,” in IEEE Electric Ship Technologies Symposium, Arlington, VA, May 2007

[91] D.C. Bergman, Jin Dong , J.P. Juen, N. Tanaka, C.A. Gunter, A.K. Wright, "Nonintrusive Load-Shed Verification," Pervasive Computing, IEEE , vol.10, no.1, pp.49-57, 2011

[92] P. K. Parikh, R. W. Cox, and J. M. Anderson, "Using Voltage Distortion Measurements for Demand-Side Management", 12th Workshop on Control and Modeling for Power Electronics (COMPEL), pp-1-5, 2010

[93] G. Kalogridis, C. Efthymiou, S.Z. Denic, T.A. Lewis, R. Cepeda, "Privacy for Smart Meters: Towards Undetectable Appliance Load Signatures", First IEEE Inter©national Conf. on Smart Grid Communications (SmartGridComm), 2010 , pp.232-237, 2010

[94] M. Zeifman, and K. Roth, “Nonintrusive Appliance Load Monitoring: Review and Outlook,” IEEE Trans. Consumer Electronics, vol.57, no.1, pp. 76-84, 2011

[95] J. Zico Kolter and Matthew J. Johnson. REDD: A public data set for energy disaggregation research. In proceedings of the SustKDD workshop on Data Mining Applications in Sustainability, 2011

[96] N. Barry, and E.McQuade, "Temperature control using integer-cycle binary rate modulation of the AC mains," Industry Applications, IEEE Transactions on , vol.31, no.5, pp.965-969, Sep/Oct 1995

[97] S.W. Heunis, and R. Herman, “A probabilistic model for residential consumer loads,” Power Systems, IEEE Transactions on, vol. 17, no. 3, pp. 621-625, 2002.

[98] Chenzhi Fan, Dun Sun and Mei Tong, “Circuit Theory” , China Machine Press, Edition 2, 2008

[99] "IEEE Recommended Practices and Requirements for Harmonic Control in Electrical Power Systems," *IEEE Std 519-1992*

[100] IEEE Working Group on Nonsinusoidal Situations: Effects on Meter Performance and Definitions of Power, “Practical Definitions for Powers in Systems with Nonsinusoidal Waveforms and Unbalanced Loads: a Discussion”, IEEE Transactions on Power Delivery, vol. 11, no. 1, January 1996, pp. 79-101.

[101] Hooman Erfanian Mazin, Data-based Harmonic Source Identification, PhD Thesis. University of Alberta, Fall 2011.

[102] Chen Jiang, A Probabilistic Bottom-up Technique for Modeling and Simulation of Residential Distributed Harmonic Sources, Master of Science Thesis. University of Alberta, Spring 2012.

[103] M.Dong, H.E.Mazin and W.Xu, "Data segmentation algorithms for a time-domain harmonic source modeling method," Electrical Power & Energy Conference (EPEC), 2009 IEEE , vol., no., pp.1-6, 22-23 Oct. 2009

[104] M.Dong, Paulo C.M.Meria, W.Xu and Walmir Freitas, “An Event Window Based Load Monitoring Technique for Smart Meters”, IEEE Transactions on Smart Grid, vol.3,No.2, pp 787-796, June 2012.

[105] M.Dong, Paulo C.M.Meria and W.Xu , “Non-intrusive Signature Extraction for Major Residential Loads”，IEEE Transactions on Smart Grid, accepted and in press , Feb. 2013

[106] Hartigan, J. A.; Wong, M. A., "Algorithm AS 136: A K-Means Clustering Algorithm". Journal of the Royal Statistical Society, Series C (Applied Statistics) 28 (1): 100–108, 1979

[107] Comaniciu, D., Meer, P., "Mean shift: a robust approach toward feature space analysis," Pattern Analysis and Machine Intelligence, IEEE Transactions on , vol.24, no.5, pp.603-619, May 2002

[108] Hung-Kuo Chu, Tong-Yee Lee, "Multiresolution Mean Shift Clustering Algorithm for Shape Interpolation," Visualization and Computer Graphics, IEEE Transactions on, vol.15, no.5, pp.853-866, Sept.-Oct. 2009

[109] Y. Z. Cheng, “Mean shift, mode seeking, and clustering,” IEEE Trans.Pattern Anal. Mach. Intell., vol. 17, no. 8, pp. 790–799, 1995.

[110] M.A. Little and N.S.Jones, "Generalized Methods and Solvers for Piecewise Constant Signals: Part I". Proceedings of the Royal Society A.,2011

[111] Wang Zhenyu and Zheng Guilin, "The application of mean-shift cluster in residential appliance identification," Control Conference (CCC), 2011 30th Chinese , vol., no., pp.3111,3114, 22-24 July 2011

[112] Amorim, R. C. and Mirkin, B, "Minkowski metric, feature weighting and anomalous cluster initializing in K-Means clustering". Pattern Recognition 45 (3): 1061–1075, 2012

[113] A. P. Engelbrecht, Computational Intelligence. Chichester, UK: John Wiley & Sons, Ltd, 2007

[114] "National Lighting Inventory and Energy Consumption Estimate", prepared for U.S. Department of Energy, National Technical Information Service (NTIS),U.S. Department of Commerce, Sep.2002

[115] Hawken, Paul; Lovins, Amory; Lovins, L. Hunter (2000), “Natural Capitalism”, Back Bay Press, Time Warner Book Group. ISBN 0-316-35300-0

[116] Design Recycle Inc., "LED Lights vs. Incandescent Light Bulbs vs. CFLs", available online at: http://www.designrecycleinc.com/led%20comp%20chart.html

[117] Earthfriends, “CFL Vs. Incandescent”, available online at: http://www.energyclassroom.com/pdfs/EC_CO_AC_LightBulbCost.pdf

[118] Energyclassroom ,“Light Bulb Cost Comparison”, available online at: www.energyclassroom.com/pdfs/EC_CO_AC_LightBulbCost.pdf

[119] “My Home Bill Explained”, available online at: www.hydroone.com/MyHome/MyAccount/MyBill/Pages/default.aspx

[120] Energy review 2006, Department of Trade and Industry, UK

[121] R.Fernandes, I.Silva, M.Oleskovicz , "Load Profile Identification Interface for Consumer Online Monitoring Purposes in Smart Grids," Industrial Informatics, IEEE Transactions on , vol.PP, no.99, pp.1,1, 0

Appendix

Load Disaggregation Method for Commercial Buildings

The appendix proposes a load decomposition method specified for commercial buildings. Limited simulations are also given for basic verification purpose. The scope of the entire study is preliminary and the main contribution is to present a novel thinking on non-intrusive load group monitoring. Suggestions for the future work are also given in the end.

A.1 The proposed method

The proposed technique aims to break down the total commercial building power usage into the level of individual load groups such as all lighting devices and all computing devices in the building.

Two basic assumptions are:

- A commercial building has only a few appliance groups and categories although the total number of appliances can be fairly large.
- From the statistical perspective, a certain load group which has many individuals will follow a generally stable signature pattern and individual appliance difference amongst it is negligible when treated collectively.

For example, an office building has 500 computers, there are two brands *A* and *B*---300 computers with brand *A* and 200 with brand *B*. Brand *A* computer has a total harmonic distortion (THD) of 30% while brand *B* has a THD of 40%. However, since the mixed ratio of brand *A* and *B* generally stays the same, as a group, all computer devices approximately always produce a stable aggregated THD, say 37%. The prerequisite of this assumption is that brand *A* and brand *B* computers have almost the same behavior characteristics (People do not use them differently). Statistically, for any random given time, the number of brand *A* computers being used vs. the number of brand *B* computers being used are generally still 3:2. Thus, the aggregated THD signature can stay approximately the same for this load group “computer devices”. However, if for example, brand A is only used in night-time and brand B is only used in day-time, they have to be treated as two independent groups. In reality, this rarely happens since for most cases the same function oriented appliances behave consistently.

Based on the above assumptions, some group signatures such as THD and individual harmonic distortion (IHD) as well as power factor can be used to label specific certain load groups.

The figure below shows an example of load division in an office building.

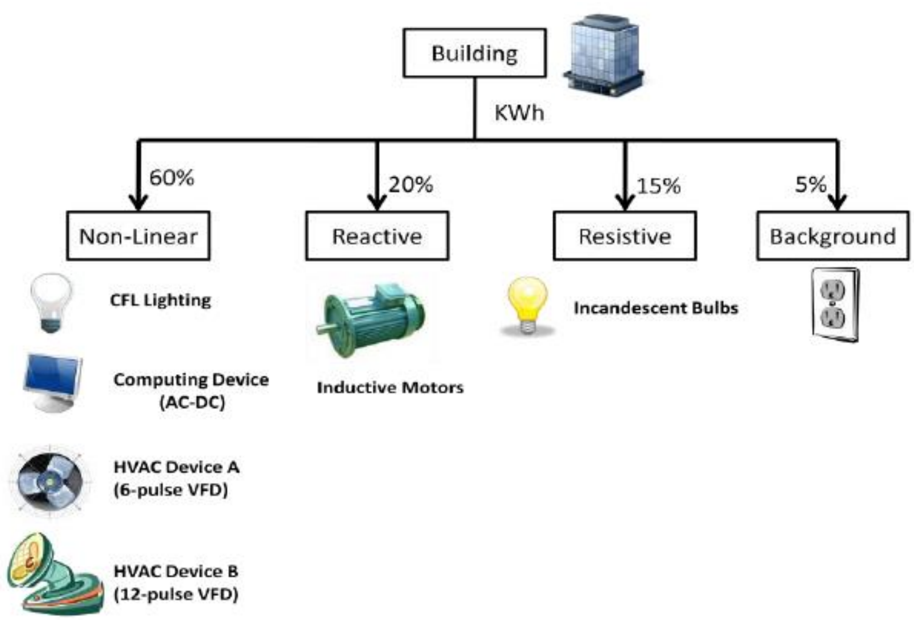


Figure A-1: Example of load division in an office building

In total, its 1500 loads can be classified into the 7 groups under 4 categories---non-linear, reactive (linear), resistive (linear) and background. Each of these categories has their own unique collective group signatures such as IHD ratio for non-linear load category and power factor for reactive load category. Based on the collective group signatures, aggregated profile acquired from the building’s meter-side can be decomposed into individual groups through mathematic methods. Different load categories can be treated specifically as below:

A. Background load

Background load can be understood as the summation of stand-by power and always-on loads. Its value can be approximated as the lowest point at night, say 1:00AM---5:00 AM when no people are acting in the building.

B. Non-linear load

As shown in Figure A-1, non-linear load in an office building might be composed of the following parts:

- CFL lighting
- Computer Device (AC-DC rectifier based)
- HVAC-A (6-pulse variable frequency driver based)
- HVAC-B (12-pulse variable frequency driver based)

From harmonic perspective, the above groups have different harmonic spectrums with each other, however, within each group, the appliances roughly have similar harmonic patterns in terms of normalized magnitude (IHD) and phase angles and can present a relatively stable mixed harmonic pattern. The pattern is often determined by the power electronic circuit it uses. For example, most modern computer devices such as desktops and laptops use single-phase AC-DC rectifiers. HVAC equipment can have either 6-pulse or 12-pulse VFD drives. Thus it is treated as two independent groups A and B separately.

To obtain an accurate representative harmonic pattern for each group, measurement based on a certain ratio can be further performed for each group. For example, suppose there are three types of computers A,B,C being used. In this office building, they have a mixture ratio as 200:300:400 (900 computers in total). Accordingly, 2 type-A computers,3 type-B computers and 4 type-C computers can be randomly chosen and put together. Then their collective IHDs and phase angles can be measured. This measured representative harmonic pattern can approximately represent the aggregated pattern of all computer devices in this building at any given time. The table below is an example of representative spectrums of all four groups (CFL /Computer /HVAC-A / HVAC-B).

Table A-1 Representative Spectrums of load groups

Harmonic order	Representative Spectrums of CFL /Computer /HVAC-A / HVAC-B	
	Normalized Magnitude (IHD)	Phase Angle (degree)
1	1.00 / 1.00 / 1.00 / 1.00	21 / 0 / 0 / -28
3	0.81 / 0.79 / 0.03 / 0.02	54 / 2 / 0 / 0
5	0.57 / 0.49 / 0.61 / 0.02	106 / 4 / -175 /-128
7	0.45 / 0.20 / 0.33 / 0.01	169 / 9 / -172 /51
9	0.44 / 0.12 / 0.01 / 0.01	-135/ 0 / 0 / 0
11	…	…

In the meanwhile, total harmonic signal of the commercial building after the background load deducted can also be analyzed and contents of harmonic orders can be extracted, which should be equal to the summation of the four non-linear groups. The following equations can be listed:

$$\begin{cases} I_3 = I_{CFL} \times I_{CFL3(p.u.)} + I_{CMP} \times I_{CMP3(p.u.)} + I_{HA} \times I_{HA3(p.u.)} + I_{HB} \times I_{HB3(p.u.)} \\ I_5 = I_{CFL} \times I_{CFL5(p.u.)} + I_{CMP} \times I_{CMP5(p.u.)} + I_{HA} \times I_{HA5(p.u.)} + I_{HB} \times I_{HB5(p.u.)} \\ I_7 = I_{CFL} \times I_{CFL7(p.u.)} + I_{CMP} \times I_{CMP7(p.u.)} + I_{HA} \times I_{HA7(p.u.)} + I_{HB} \times I_{HB7(p.u.)} \\ I_9 = I_{CFL} \times I_{CFL9(p.u.)} + I_{CMP} \times I_{CMP9(p.u.)} + I_{HA} \times I_{HA9(p.u.)} + I_{HB} \times I_{HB9(p.u.)} \\ \ldots \end{cases} \tag{A.1}$$

where the left side are the harmonic contents of total aggregated signal. The per-unit value can be read from Table A-1. For example, $I_{CFL3(p.u.)}$ is complex value $0.81\angle 54^\circ$. $I_{CFL}, I_{CMP}, I_{HA}, I_{HB}$ are unknown variables that need to be solved. If there are only four groups, harmonic contents of four orders (I_3, I_5, I_7, I_9) are sufficient to solve the variables. However, more harmonic contents can also be used to convert the problem to an over-determined problem in which case some optimization methods such as least square method can be applied to improve the accuracy. Once $I_{CFL}, I_{CMP}, I_{HA}, I_{HB}$ are solved, their power and energy profiles can be immediately obtained.

C. Reactive (linear) load

Based on the above discussions, the summation of reactive power from the non-linear load category can also be calculated. Apart from non-linear load and background load, the rest aggregated reactive power all comes from the reactive (linear) loads. Similar to the measurement taken to get representative harmonic spectrum, representative power factor can also be measured. For example, a couple of refrigerators and fans can be measured together according to a mixing ratio. Thus power profile of reactive load can be obtained. Also, based on the mixing ratio, refrigerators and fans can be further separated.

D. Resistive (linear) load

In the end, the residual power apart from background load, non-linear load and reactive load belongs to resistive load.

A.2 Simulation results

Due to limited resources, two simulations are conducted based on the bottom-up model and program developed for multiple residential houses in [102]. However, it is still sufficient to verify the basic working principle for commercial buildings and reveal some interesting findings. In this case, a commercial building is assumed to include 6 groups of loads--- 800 CFL lights, 300 computers, 100 furnaces, 100 inductive motors, 2400 incandescent lights and background power. Their powers and mixing ratios are listed in Table A-2.

The program uses Monte Carlo based simulation to generate the aggregated signal of all the loads according to certain statistical time of use probability profiles for different loads [102].

Table A-2 Load group inventory and composition

Load group	Mixture appliances	Power of appliances	Mixing ratio	Total Number
CFL	ID-6	29.47 W	50 %	800
	ID-12	14.66 W	50 %	
Computing device	PC ID-1 +Monitor ID-1	98.76 W+32.43 W	33.3%	300
	PC ID-2 + Monitor ID-1	90.85 W+25.27 W	33.3%	
	PC ID-3 + Monitor ID-12	73.71 W+29.87 W	33.3%	
Furnace	ID-1	519.3W	40 %	40
	ID-2	648.9W	60 %	60
Inductive Motor	ID-1	5050.8W	100%	100
Incandescent Light	ID-2	59.28 W	100%	2400
Background	---	15000 W	---	---

Simulation 1: At any given moment of a day, the mixing ratios are kept fixed and the same with the above table’s values. Also, the background power is removed. This is a very ideal case.

Using the proposed method explained in A.1, the power curve of individual load group can be separated from the aggregated signal. In simulation 1, it is found the separated load group power has exactly the same power has the actual one. An example is shown in Figure A-2.This is because the mixing ratios are constant. When solving the problem using equations such as (A.1), the exact results can be obtained.

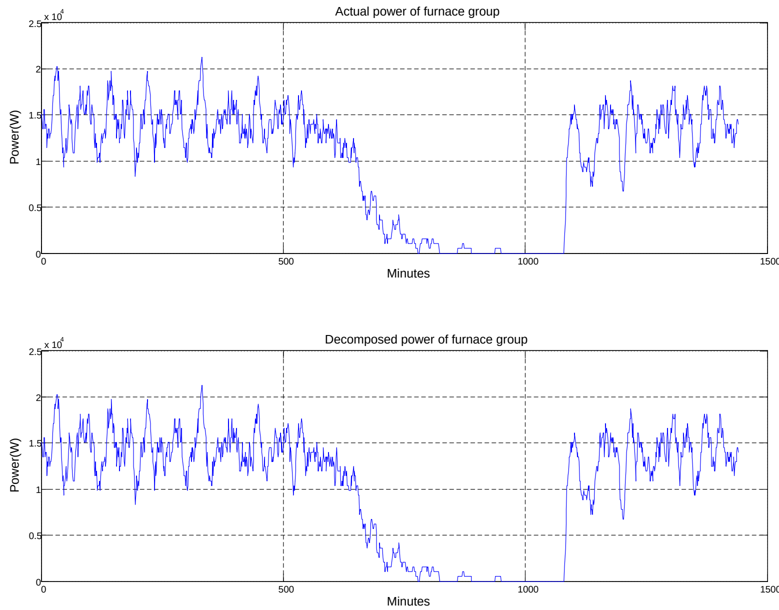
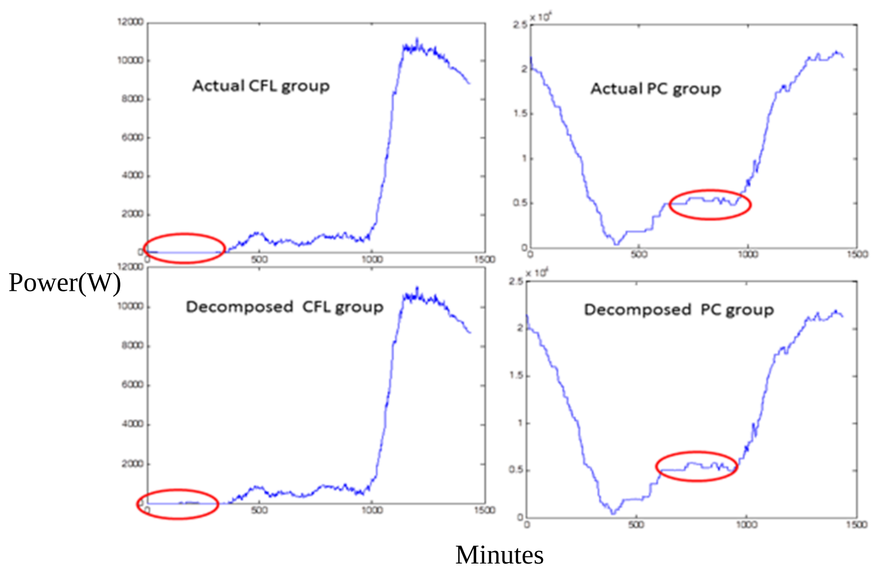
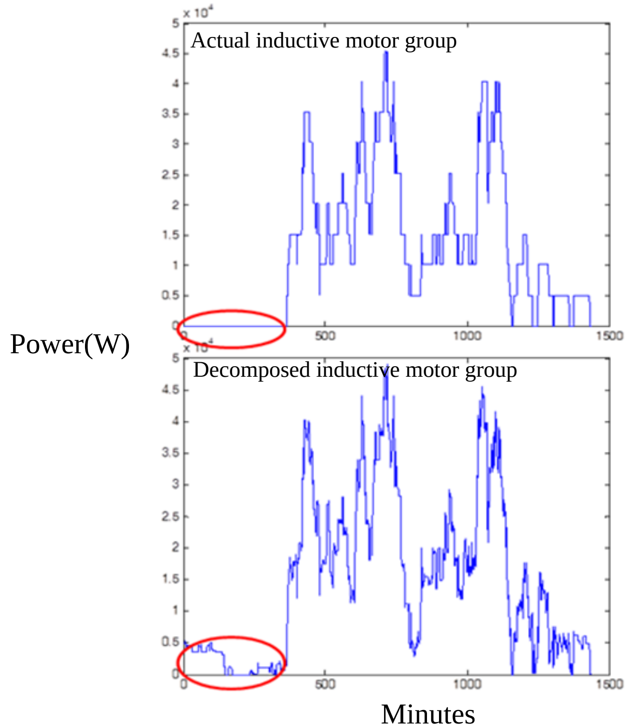


Figure A-2: Example of actual power vs. decomposed power in simulation 1

Simulation 2: At any given moment of a day, a single appliance is switched ON/OFF randomly according to its own time of use probability profile. In other words, the instant mixing ratio is unknown. However, statistically, instant ratio should fluctuate around the average ratio that is the pre-defined mixing ration listed in Table A-2. This scenario is much closer to reality, under which it is only possible to know the static mixing ratio but impossible to know the dynamic mixing ratio at a certain given instant.



(a) Non-linear load groups



(b) Reactive (linear) load groups

Figure A-2: Examples of actual power vs. decomposed power in simulation 2

As can be seen from Figure A-2, the error on power curves becomes larger, especially for reactive load groups. Quantitative comparison can be seen in Table A-3.

Table A-3 Load group inventory and composition

Group type	Trend correlation coefficient	Actual Energy (KWH)	Decomposed Energy (KWH)	Energy Error (%)
CFL group	0.9999	70.6625	68.7657	2.7
Computing device	0.9999	243.8629	245.1459	0.5
HVAC group	0.9981	242.8680	245.3361	1.0
Inductive Motor group	0.9853	292.0204	352.5564	20.7
Incandescent Light	0.9977	571.7378	525.5483	8.1

As can be seen from Table A-3, the above decomposed results are close to the actual results. The error for non-linear load category such as CFL group and computing devices is very low. However, for reactive and resistive loads, errors are larger. This is because they do not have as plentiful information as the harmonic spectrum. In other words, they cannot be accurately solved or optimized by using equations like (A.1). But the results can still be used as reasonable references. Besides, it is found that the trend shapes of all types of load groups are very similar to the actual ones.

A.3 Suggestions for future work

Some suggestions are given for future work on load disaggregation research for commercial buildings:

- For different types of commercial buildings, more realistic load groups and categories can be investigated.
- More realistic simulations that consider the actual time of use probability profiles in commercial buildings [121] can be done.
- Field data based verification should also be performed. This is the most challenging part since the number of loads in a building is massive and difficult to be measured directly. Energy auditing based methods may be considered.
- The application of the proposed method may also be extended to industry compound loads.