Personalized Energy Services

A Data-Driven Methodology towards Sustainable, Smart Energy Systems

Srirangam Narashiman, Akshay

DOI
10.4233/uuid:a17fa324-8783-4578-a838-0f53c8061ddf

Publication date
2017

Document Version
Final published version

Citation (APA)
Srirangam Narashiman, A. (2017). *Personalized Energy Services: A Data-Driven Methodology towards Sustainable, Smart Energy Systems*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:a17fa324-8783-4578-a838-0f53c8061ddf

Personalized Energy Services

A Data-Driven Methodology towards Sustainable, Smart Energy Systems

PERSONALIZED ENERGY SERVICES

A DATA-DRIVEN METHODOLOGY TOWARDS SUSTAINABLE, SMART ENERGY SYSTEMS

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. ir. K. C. A. M. Luyben,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op
maandag 29 mei 2017 om 12:30 uur

door

Akshay Uttama Nambi SRIRANGAM NARASHIMAN

Bachelor of Engineering, Computer Science and Engineering,
Visvesvaraya Technological University, India
geboren te Mysuru, India.

This dissertation has been approved by the

promotor: Prof. dr. K. G. Langendoen
copromotor: Dr. R. Venkatesha Prasad

Composition of the doctoral committee:

Rector Magnificus	chairman
Prof. dr. K. G. Langendoen	Delft University of Technology
Dr. R. Venkatesha Prasad	Delft University of Technology

Independent Members:

Prof. dr. N. Bessis	Edge Hill University
Prof. dr. K. Blok	Delft University of Technology
Prof. dr. J. Kuri	Indian Institute of Science
Prof. dr. P. Palensky	Delft University of Technology
Prof. dr. ir. G. P. J. Verbong	Eindhoven University of Technology

Author email: akshay.uttama@gmail.com

ISBN 978-94-6186-813-8

An electronic version of this dissertation is available at
http://repository.tudelft.nl/.

This work was carried out in the TU Delft graduate school.

This work was carried out in the ASCI graduate school. ASCI dissertation series number 371.

This work was funded by iCore, a project sponsored by the EU FP7 program.

“An idea that is developed and put into action
is more important than an idea
that exists only as an idea."

Edward de Bono

Contents

SUMMARY

The rapid pace of urbanization has an impact on climate change and other environmental issues. Currently, 54% of the global population lives in cities accounting for two-thirds of global energy demand. Sustainable energy generation and consumption is the top humanity's problem for the next 50 years. Faced with rising urban population and the need to achieve energy efficiency, urban planners are focusing on sustainable, smart energy systems. This has led to the development of Smart Grids (SG) that employs intelligent monitoring, control and communication technologies to enhance efficiency, reliability and sustainability of power generation and distribution networks.

While energy utilities are optimizing energy generation and distribution, consumers play a key role in sustainable energy usage. Several energy services are provided to the consumers to know households' hourly energy consumption, estimate monthly electricity cost and recommendations to reduce energy consumption. Furthermore, advanced services such as demand response, can now control and influence energy demand at the consumer-end to reduce the overall peak demand and re-shape demand profiles. The effectiveness and adoption of these services highly depend on the consumers' awareness, their participation and engagement. Current energy services seldomly consider consumer preferences such as their daily behavior, comfort level and energy-consumption pattern. In this thesis, we investigate development of personalized energy services that strive to achieve a balance between efficient-energy consumption and user comfort.

Personalization refers to tailoring energy services based on individual consumers' characteristics, preferences and behavior. To develop effective personalized energy services a set of challenges need to be tackled. First, fine-grained data collection at user and appliance level is required (*data collection challenge*). Mechanisms should be devised to collect fine-grained data at various levels in a non-intrusive way with minimal sensors. Second, personalized energy services require detailed user preferences such as their thermal comfort level, appliance usage behavior and daily habits (*user preference challenge*). Accurate learning models to derive user preferences with minimal training and intrusion are required. Third, energy services developed needs to be easily scalable, from one household to tens and thousands of households (*scalability challenge*). Mechanisms should be developed to tackle the deluge of data and support distributed storage and processing. Fourth, energy services should deliver real-time feedback or recommendations so that users can promptly act upon it (*real time challenge*). This calls for development of distributed and low complexity algorithms.

This thesis moves away from traditional SG services – which hardly consider consumer preferences and comfort – and proposes a novel approach to develop effective personalized energy services. The proposed energy services provide actionable feedback, raise awareness and promote energy-saving behavior among consumers.

In this thesis, we follow a bottom-up data-driven methodology to develop personalized energy services at various scales – (i) nano: individual households, (ii) micro: buildings and spaces, and (iii) macro: neighborhoods and cities. To this end, we present our approach –

physical analytics for sustainable, smart energy systems – that combines IoT data, physical modeling and data analytics to develop intelligent, personalized energy services. Physical analytics fuses data from various Internet of Things (IoT) devices such as smart meters, smart phones and smart watches, along with physical information such as household type, demographics and occupancy to infer energy-usage patterns, user behavior and discover hidden patterns. This approach is used to learn and model user preferences and energy usage, subsequently, employed to develop personalized energy services.

This thesis is organized into three parts. Part I describes how to derive fine-grained information with minimal sensors and intrusion. We present two novel algorithms *viz.*, LocED and PEAT that derive fine-grained information from appliance and user level, respectively. This real-time information is used to raise awareness on energy-usage behavior among occupants. Part II presents personalized energy services targeted at households and buildings. We develop services that shift and/or reduce energy consumption and cost by considering individual consumers' preferences and comfort. These energy services are aimed at providing actionable feedback to occupants towards sustainable energy usage. Part III presents energy services targeted at neighborhood and city level. These energy services aim to identify target consumers in a neighborhood based on their energy-usage pattern and preferences for various DR programs. Finally, we present data-processing architectures that investigate how to cope with the overwhelming data generated from smart meters towards design and development of sustainable, smart energy systems.

This thesis advocates that the design and development of energy services should follow personalized approach with consumer preferences and comfort given paramount importance. Results show that the personalized energy services developed has significant potential to raise awareness, reduce energy consumption and improve user comfort in smart – homes, buildings and neighborhoods.

SAMENVATTING

De snelle urbanisatie wereldwijd heeft een grote impact op het klimaat en leefmilieu. Op dit moment leeft 54% van de wereldbevolking in stedelijk gebied, en is verantwoordelijk voor twee derde van het totale energieverbruik. Duurzame energieopwekking en gebruik is het belangrijkste probleem van de mensheid voor de komende 50 jaar. Gegeven de immer uitdijende steden en de noodzaak om zuinig om te gaan met energie, zoeken stadsontwikkelaars de oplossing steeds meer in duurzame, slimme energiesystemen. Deze trend heeft geleid tot de ontwikkeling van zogeheten slimme energienetwerken (Smart Grids) gebaseerd op slimme meters, geavanceerde beheer- en communicatietechnologieën, duurzame stroombronnen, en betrouwbare transport netwerken.

Terwijl energieleveranciers de opwekking en transport optimaliseren, spelen gebruikers een hoofdrol bij het verduurzamen van het energieverbruik. Diverse diensten spelen hier op in door gebruikers inzicht te geven in hun dagelijkse/maandelijkse stroomverbruik, en advies te geven hoe het energieverbruik verminderd kan worden. Verder worden er geavanceerde diensten aangeboden, zoals Demand Response, die de vraag controleren en sturen om de piekbelasting in het energienetwerk te spreiden en te verlagen. De effectiviteit van deze maatregelen hangt sterk af van de betrokkenheid van de consumenten, hun medewerking en hun gebruikspatroon. In dit proefschrift onderzoeken we de ontwikkeling van gepersonaliseerde diensten die beogen om de juiste balans te vinden tussen energieverbruik en comfort.

Personalisatie verwijst naar het toespitsen van diensten op de individuele gebruikspatronen, voorkeuren, en gedragingen van consumenten. Om tot een effectieve oplossing te komen moeten de volgende uitdagingen overwonnen worden. Ten eerste is het noodzakelijk om gedetailleerde data te verkrijgen over het daadwerkelijke stroomverbruik per apparaat en per gebruiker. Dit meten moet zo onopvallend mogelijk en met minimale inspanning van de gebruiker gerealiseerd worden, en tegen minimale kosten. Ten tweede moet er per persoon inzicht verkregen worden in diens gewenste gebruikersinstellingen, bijv. de ideale kamertemperatuur, diens gebruik van elektrische apparatuur, en diens dagelijkse gedragingen. Zelflerende modellen zijn vereist om het achterhalen van deze informatie zo gemakkelijk mogelijk te maken. Ten derde moeten slimme energiediensten naadloos kunnen opschalen van één huishouden tot duizenden. Dit vereist het ontwikkelen van oplossingen die de grote hoeveelheid gegenereerde data (lokaal) kunnen opslaan en verwerken. Ten vierde moeten energiediensten ontwikkeld worden die in realtime werken ten einde de gebruikers onmiddellijk van advies te kunnen voorzien. Dit vraagt om simpele algoritmen die ter plaatse uitgevoerd kunnen worden.

De aanpak in dit proefschrift is dus een trendbreuk t.o.v. traditionele energiediensten die de gebruikers buiten beschouwing laten. Het gebruik van persoonlijke voorkeuren en energieconsumptie data opent de weg naar een nieuwe klasse van diensten die veel efficiënter zijn dan de huidige generatie. De voorgestelde gepersonaliseerde diensten resulteren in concrete adviezen, verhogen het bewustzijn, en promoten energiezuinige leefstijl onder de gebruikers ervan.

We volgen een data-gedreven aanpak bij de ontwikkeling van gepersonaliseerde diensten op diverse niveaus, te weten (i) nano: individuele huishoudens, (ii) micro: gebouwen en openbare ruimtes, en (iii) macro: hele wijken en steden. Centraal staat de "physical analytics" methode die, voor onze slimme energiediensten, Internet of Things (IoT) data combineert met fysische modellering en data-analyse om zo te komen tot slimme, gepersonaliseerde energiediensten. We combineren de data van verschillende IoT apparaten zoals slimme meters, smartphones en smartwatches met demografische data, en huishoud type en bezetting om gebruikspatronen en voorkeuren van consumenten te achterhalen, die vervolgens gebruikt worden om modellen te ontwikkelen die ingebouwd worden in onze gepersonaliseerde energiediensten.

Dit proefschrift is opgedeeld in drie delen. Deel I beschrijft hoe gedetailleerde gebruiksinformatie gemeten kan worden met een minimum aan sensoren en medewerking van de consumenten. We introduceren LocED en PEAT, twee strategieën om respectievelijk gebruiksdata van apparaten en personen te meten. Dit gebeurt realtime om consumenten zo direct mogelijk bewust te maken van hun energieverbruik. Deel II richt zich vervolgens op het benutten van deze informatie in gepersonaliseerde diensten op de schaal van huishoudens en (kantoor)gebouwen. I.h.b. ontwikkelen we diensten die energieverbruik verschuiven of verminderen, en dus kosten besparen, aan de hand van persoonlijke voorkeursinstellingen voor comfort. Deze diensten resulteren in advies omtrent concrete acties voor bewoners om hun stroomgebruik te verduurzamen. Deel III ten slotte beschrijft diensten die toegepast kunnen worden op de schaal van wijken en steden. I.h.b. richten deze diensten zich op het identificeren van groepen gebruikers die een profiel hebben dat geschikt is voor verschillende demand-response technieken. We sluiten af met de introductie van een reeks data-processing architecturen voor het verwerken van de overweldigende hoeveelheid data die door slimme energiemeters gegenereerd wordt, om het mogelijk te maken duurzame, intelligente energiesystemen te kunnen ontwerpen en implementeren.

De essentiële bijdrage van dit proefschrift is het pleidooi om een gepersonaliseerde aanpak te kiezen bij het ontwikkelen van duurzame energiesystemen en gebruik te maken van individuele voorkeursinstellingen en consumptiepatronen. De resultaten van diverse (simulatie) experimenten laten zien dat deze gepersonaliseerde aanpak grote kansen biedt om het bewustzijn van gebruikers te verhogen, hun verbruik te verminderen, en het comfort te verhogen in toekomstige, slimme huizen, gebouwen, wijken en steden.

1

INTRODUCTION

"The best way to predict the future is to create it."

Peter Drucker

ADVANCEMENTS in science and technology have played a key role in transforming our lives and cities over the past few decades leading to urbanization. Urbanization refers to a population shift from rural to urban areas in the interest of having a better quality of life and economic opportunities. Underpinning this transformation are two key drivers, access to affordable electricity and effective social policy. Currently, urban population accounts for 54% of the total global population, up from 34% in 1960, and continues to grow [15]. This rapid pace of urbanization has an impact on climate change and other global environmental issues. For example, cities account for more than two-thirds of the global energy demand and result in 60-80% of global greenhouse gas (GHG) emissions [17]. The effects of urbanization have increased awareness for sustainable practices.

Energy is the major factor for development and urbanization. Sustainable energy generation and consumption is the top humanity's problem for the next 50 years [110]. In recent years, researchers, industry and government organizations have focused on sustainable energy systems. A sustainable energy system aims to lower carbon emissions on the supply-side (energy utilities) and improve distribution infrastructure along with lowering energy consumption on the demand-side (consumers). With the ever-increasing energy demand, utility companies started investigating peak demand-time periods, and ways to encourage consumers to reduce and/or shift energy consumption giving rise to the concept of modernizing the power grid.

Traditional power grids are primarily used to carry power from a few generators to a large number of consumers. In contrast, a modernized power grid – Smart Grid (SG) – employs intelligent monitoring, control, communication and self-healing technologies to enhance efficiency, reliability and sustainability of power generation and distribution networks [9] (See Figure 1.1). In this direction, several energy services are proposed such as integration of renewable energy sources reducing usage of fossil fuels and GHG emissions [87], improving transmission and distribution infrastructure towards efficient electricity transmission with lower costs [49], and reducing/shifting energy consumption by automated

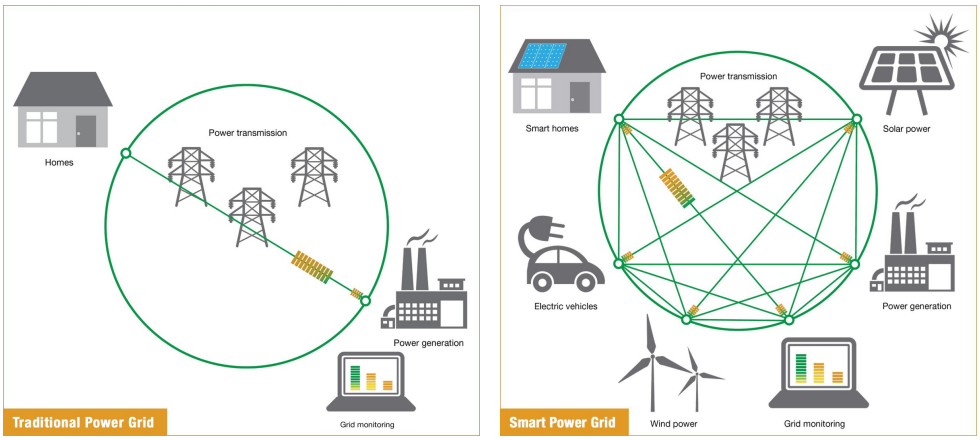


Figure 1.1: Traditional power grid and smart gird [1].

demand management [91]. For example, washing machines do not need to run at a specific time-period and can be turned on automatically while the consumer is asleep, or at work.

A key driver for sustainable energy usage needs to come from the demand-side – households, buildings, industries, neighborhoods – who must change their consumption patterns and adopt energy-saving techniques. Prevalent efforts mainly focus on improving infrastructure to support bi-directional communication between consumers and utilities. Thus enabling real-time feedback to the consumers on power consumption, power quality (e.g., steady supply voltage) and pricing details. A wide literature of works [26, 68, 122] focuses on sensing real-time energy consumption at households, automated appliance controlling based on the electricity price, integrating renewable sources to balance energy demand and thwarting blackouts. While these techniques help in lowering carbon emissions, increasing usage of renewable resources and smoothing peak demands at the demand-side, they seldomly consider consumer preferences such as their daily behavior, comfort level, energy consumption pattern, social belief and ties. The effectiveness and adoption of SG techniques/services highly depends on the consumers' awareness, their participation and engagement [64, 112]. Thus user-centric design and development of SG services need to be considered for building the sustainable, smart energy system.

With the rapid advancements in embedded systems and wireless technologies, the vision of Mark Weiser is becoming a reality. In his monumental work [117], he envisioned "ubiquitous computing" i.e., personal computers that integrate seamlessly into a user's environment, enriching his everyday life by automating many of his routine tasks. User environments such as households and buildings are impregnated with embedded devices to capture the context and adapt the ambience around a user accordingly to improve his experience. This has led to the development of smart homes and smart buildings, in general, smart spaces. Researchers and industry are striving to create services where spaces / environments are active (or even proactive) and take autonomous decisions *without a user in the loop*. Several automated services are being developed such as, automated controlling of appliances, opening and closing of blinds based on lighting conditions, and pre-heating

rooms when the temperature falls below a threshold. These services over time were found to be importunate and no more effective due to lack of consideration of user preferences. Hence, services should provide the necessary information to the users[1] empowering them to make informed decisions (bringing them in the loop). Rather than the space taking autonomous decisions, it should provide suggestions to the users for subsequent actions. For example, rather than preventing the usage of washing machine when the electricity price is high, services should provide necessary information to the user, where users decide to participate or not based on their preferences.

Consider a smart-home scenario, where appliances in the household are monitored individually using a sensor. Energy-consumption information from individual appliances can now be provided to the users in real time. Services such as real-time feedback, remote control of appliances and detection of faulty appliances are being developed to understand energy usage and to improve user comfort. Smart homes can now operate certain appliances depending on the power availability and its associated cost. For example, a washing machine can be controlled in such a way that it will only start operating when there is enough power in the grid and when the cost is under a certain threshold. Although these techniques curb the energy demand they *hardly* consider preferences and comfort of consumers. A consumer might not be willing to shift the usage of an appliance to the proposed time periods. Current techniques give higher importance to cost savings and/or demand reduction rather than consumer comfort. Hence, there is a need to bring the user-in-the-loop (UIL) for making informed decisions and developing personalized energy services. In order to be effective, energy services should strive to achieve a balance between energy efficiency and user comfort.

This thesis is motivated by such a scenario: the need to provide effective, personalized energy services to users while striking a balance between energy consumption and user comfort. The development of energy services should consider individual user preferences and provide actionable feedback. The essence of this thesis is to raise awareness, reduce energy consumption and improve user comfort in smart – homes, buildings and neighborhoods.

1.1. SMART GRID ECOSYSTEM

In order to understand the challenges in developing personalized energy services, we will first introduce some basic knowledge about smart grids (SG) and its services. The traditional power grid comprises of an interconnected network of power systems that carries electricity from power plants to consumers. In contrast, SG takes advantage of ICT to integrate the power infrastructure with an information infrastructure [45]. The information infrastructure supports sensing, computation, control and information exchange capabilities. Smart grids (SG) are energy networks that can monitor energy flows and adjust to changes in energy supply and demand accordingly.

Smart grids are not only about the modernization of traditional power girds but also about enhancing cooperation among various actors. Actors in SG include energy generators, consumers and prosumers (those that do both), where each operates autonomously, but needs to communicate with others to balance energy supply and demand. Smart grid is an ensemble of several services (see Figure 1.2) such as demand response, demand fore-

[1] We use the terms user, consumer and occupant interchangeably.

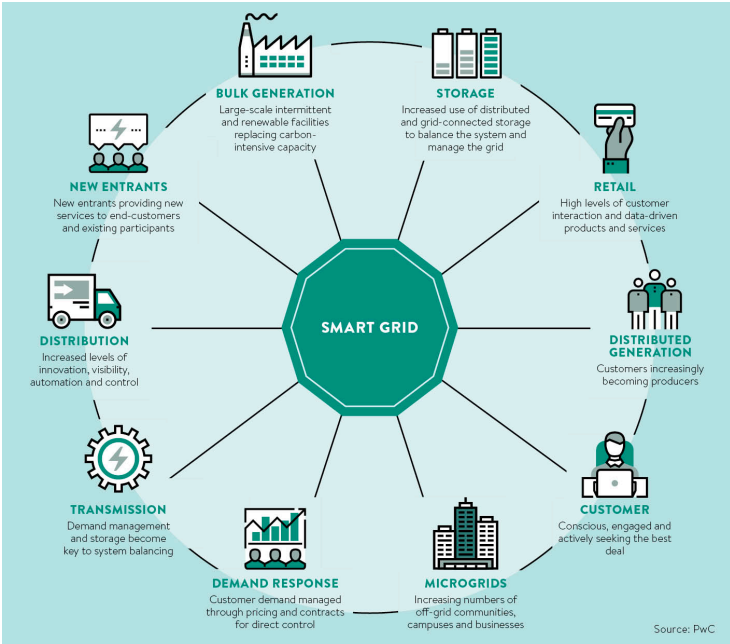


Figure 1.2: Smart Grid Ecosystem with various services [16].

cast, emergency management, anomaly detection, and adaptive pricing. We enlist some key energy services and the issues associated for wide adoption in SG.

Smart Metering. A fundamental building block for all SG services is smart metering or advanced metering infrastructure (AMI). AMI comprises of electricity meters that measure, collect and analyze data about energy usage [45]. The EU aims to replace at least 80% of the traditional meters with smart meters by 2020 [9]. The electricity meters or smart meters can analyze energy consumption and provide real-time feedback to the users. AMI supports bidirectional communication between these devices and utility providers for a wide variety of services. Some of the services include providing immediate feedback on power usage, power quality, and pricing details.

Issues: Several commercial smart meters are available with different communication and sensing mechanisms. Standards are needed to ensure interoperability among various smart meters. Low-cost smart-meter design, reliable data communication between the smart meters to utilities, and enforcing consumer-data protection and privacy are some active research topics.

Demand Regulation. Demand regulation, or response (DR), is a technique that can control and influence energy demand at the consumer-end to reduce the overall peak demand, reshape demand profiles and increase the robustness of the grid. The real-time pricing information communicated to the smart meters can be used to control and adjust the demands at the consumer premises. These reductions can result in lesser strain on the grid and decrease the need for high-cost generation resources. DR techniques can be broadly classified

into direct load control (DLC) and incentive based [115]. In DLC utilities can directly control the consumer appliance to tackle peak-demand reductions, for example, preventing the usage of washing machines during the peak time periods. DLC is a simple and effective way to support the operation of the power grid, however, it penalizes the user comfort. Incentive based DR techniques provide pricing incentives to encourage consumers to shift usage to off-peak hours. The effectiveness of these techniques highly depends on consumer awareness and participation.

Issues: DR techniques require a robust, secure and reliable communication infrastructure for its operation. The effectiveness of DR programs relies on accurate measurement of energy consumption and modeling of consumer preferences. Active research topics include developing distributed or decentralized DR programs, DR programs that target minimization of the total power consumption and/or the maximization of the user comfort and design of simpler, dynamic pricing schemes.

Demand Forecast. Demand forecasting (DF) refers to the prediction of power consumption levels in the next hour, next day, or up to a week ahead for either a household, building, neighborhood or city. The avalanche of data available from individual smart meters can be analyzed effectively for demand forecasting. Demand forecasting can be classified into (i) very short-term forecasting: from minutes to several hours and used in controlling the demand, (ii) short-term forecasting: from hours to weeks and used to adjust energy generation and demand, (iii) medium-term and long-term forecasting: from months to years used for asset management [58].

Issues: Demand forecasting techniques allow proper energy-generation planning and adaptation. Algorithms that take into consideration consumer demographics, preferences, weather information, and historical consumption information need to be developed for accurate prediction of energy consumption.

Microgrids. Microgrids are localized, small-scale grids that enable the integration of distributed energy resources (DER) such as solar and wind. DER transforms a centralized, producer-controlled energy network to a distributed, consumer network, where consumers not only use the energy but also produce – prosumers. The use of local energy generation to support the energy demand helps to reduce energy losses in transmission and distribution. Microgrids can operate autonomously and are typically low-voltage DC/AC grids. Numerous energy management techniques are proposed to determine when to borrow energy from the grid, when to store or sell excess energy generated from DER, and when to trade energy between the households in a microgrid.

Issues: The adoption of microgrids is hindered mainly due to the cost of elements such as solar panels, wind turbines, energy storage, and advanced controlling. Development of distributed energy trading and management techniques, and standardization efforts towards implementation of robust microgrids are some active research problems.

Over the past decade, research in academia and industry has focused on developing the aforementioned SG services. We now highlight some of the challenges that exist across various SG services.

1) Communication. Numerous sensors such as smart meters are being deployed to monitor and control devices at utility and consumer ends. Reliable communication between these

devices forms an essential component for various SG services. Due to a lack of standardization, various technologies can be applied at household, neighborhood, and city level. One has to identify the appropriate communication technology (e.g., Zigbee, WiFi), data collection protocol and Quality of Service (QoS) parameters (e.g., latency, bandwidth) based on the requirements of the SG services.

2) Interoperability. Interoperability is the ability of diverse systems to work together, exchange information and cooperate. Interconnection of a potentially large number of disparate energy-generation sources, distribution networks, consumers, and prosumers is a key feature of SG. There is currently no common understanding for interoperability between these system components. Systems that are independent of the physical medium, manufactures and the type of devices need to be designed.

3) Security. With the large-scale roll out of smart meters across the globe, they become an attractive target for malicious hackers. The lack of standards for secure, reliable communication between these devices has led to several vulnerabilities. Hackers compromising a smart meter can manipulate energy-usage information, associated cost and feed misinformation to the utilities. Energy-usage misinformation would mislead utilities in making incorrect decisions and can harm the electrical infrastructure with excess production or blackouts.

4) Privacy. Smart meters can measure the fine-grained energy consumption of a household, leading to privacy concerns. Energy-usage information can reveal consumer habits and behaviors such as when they are at home and watching TV. Standardization activities along with data anonymization need to be designed to protect privacy-sensitive data.

5) Personalization. The effectiveness of several energy services such as demand regulation, demand management and energy trading depend on consumer preferences and behavior. Personalization refers to tailoring energy services based on individual consumers' characteristics, preferences and behavior. Depending on the energy service, personalization can be fine-tuned to groups or segments of consumers. Several techniques need to be designed to collect consumer data at various levels (activities, preferences, behaviors) and model them efficiently, for wide adoption of energy services.

1.2. PROBLEM STATEMENT

Energy services provided by traditional SG applications target to reduce energy consumption. Widespread adoption and effectiveness of energy services depend on consumer preferences and comfort, for example, a DR scheme fine-tuned based on consumer preference has a higher probability of success [81, 119]. We argue that the design and development of energy services should follow a personalized approach with consumer preferences and comfort given paramount importance. The design of user-centric, personalized energy services will raise awareness, promote energy conservation behavior and has the potential to reduce the total energy consumption.

Apart from the challenges described previously, we highlight the key challenges in developing personalized energy services towards building sustainable, smart energy systems.

1) Data collection. The fine-grained data collection at various levels (e.g., appliance, user) is crucial towards the development of personalized energy services. A large number of sensors

needs to be deployed to collect fine-grained energy-usage information of households, appliances and occupants. This setup is cumbersome to maintain and has a high cost. Thus, obtaining fine-grained data at various levels in a non-intrusive way with minimal sensors is challenging.

2) User preferences. User preferences such as thermal comfort level, appliance usage behavior, and daily habits, need to be obtained for designing personalized energy services. Traditionally, user preferences are collected either by explicitly asking via a survey, or indirectly by observing and interpreting user actions with the system. Accurate learning models need to be developed to derive user preferences in an indirect manner using data collected from various sensors such as smart meters, smartphones and wearables.

3) Scalability. Energy services need to be designed at various scales – households, buildings, neighborhood, cities – to develop a sustainable energy system. As the number of households increases the data produced grow multifold. Thus, it is challenging to analyze and handle the deluge of data. Further, traditional services that are centralized face high latency and require large bandwidth to collect the data and have difficulties to scale.

4) Real time. Energy services such as feedback on consumption and DR events should be analyzed and delivered to consumers in near real time. Hitherto, most of the services had high latency due to centralized systems. Thus, distributed and low complexity algorithms are required to develop real-time energy services.

Given the aforementioned challenges, this thesis addresses the following question.

How to develop effective personalized energy services?

Approach. Our take is that to develop effective personalized energy services, fine-grained data collection at various levels with minimal sensors, coupled with accurate user preference models, and low complexity algorithms are needed. This idea is distilled to define physical analytics for sustainable, smart energy systems. Physical analytics (PA) [7] is an approach that combines IoT data, physical modeling and data analytics to develop intelligent, personalized energy services (see Figure 1.3). The three pillars of physical analytics are listed below.

1) IoT data. This includes the avalanche of data collected from various devices/appliances in the smart-grid ecosystem. At the energy level, smart meters and smart plugs are used to collect energy consumption and generation data. At the user level, smartphones and wearable devices are used to collect user preferences and comfort.

2) Physical modeling. This includes the modeling of user interactions in the physical world either with devices or other users or with physical space, such as household, buildings and neighborhood. Physical modeling utilizes data-driven techniques to study the effects of user interactions and predict their behavior.

3) Data analytics. It combines raw data from various actors (producers, consumers, prosumers) across scales (households, buildings, neighborhood) to discover hidden patterns, preferences and relationships. The knowledge derived is utilized in developing effective personalized energy services.

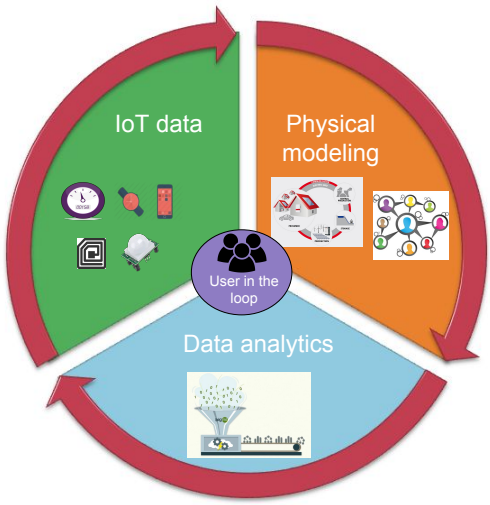


Figure 1.3: Physical Analytics for Sustainable, Smart Energy Systems.

Using the aforementioned concepts, we present and evaluate various personalized energy services for occupants to provide actionable feedback, raise awareness and promote energy-saving behavior. The proposed energy services follow a data-driven, distributed approach with low-complexity algorithms and are scalable, from one household to tens and thousands of households.

Design principle. This thesis presents personalized energy services towards the development of sustainable, smart energy systems. We follow a bottom-up approach wherein, energy services are developed at various scales – (i) nano: individual households, (ii) micro: buildings and spaces, and (iii) macro: neighborhoods and cities. Our belief is that developing persuasive energy services can effectively raise awareness, change users' attitudes and behavior, rather than coercion or automated energy services.

1.3. THESIS CONTRIBUTIONS AND OUTLINE

This thesis tackles the problem of developing effective personalized energy services, which underpins the road to sustainable, smart energy systems. In order to develop sustainable, smart energy systems we follow a data-driven approach, where energy services are designed to adapt based on consumer needs and preferences. Furthermore, we employ real-world data collected from several households and buildings for the development of personalized energy services. This thesis is organized into three parts. Part I forms the basic building block, which describes how to derive fine-grained information with minimal sensors and intrusion at households and buildings (*data collection and real time challenge*). This fine-grained information is then used to develop energy services at various scales. Specifically, Part II describes energy services tailored to individual consumer preferences (*user preferences and real time challenge*), and Part III presents energy services tailored to group or segment of consumers at a neighborhood or city level (*scalability challenge*).

PART I – FINE-GRAINED DATA COLLECTION

In this part, we address the fundamental problem in developing personalized energy services: how to obtain fine-grained information with minimal sensor deployment. Specifically, we focus on obtaining energy-usage information from appliance level and user level.

Energy Disaggregation – Chapter 2 Providing detailed appliance-level energy-consumption information may lead consumers to understand their energy-usage pattern. With this information users can be encouraged to change their behavior to save 5-15% of electricity usage [47, 54]. Smart meters deployed at households can only provide aggregate energy consumption and fail to provide appliance-specific usage. The naive way of obtaining appliance-level information is by deploying a sensor for each appliance. Such a deployment is intrusive, cumbersome to maintain, and has high cost. To this end, we propose the *Location-aware energy disaggregation framework (LocED)* that estimates appliance-level energy consumption from aggregated smart-meter data and user-occupancy information. Traditional energy-disaggregation algorithms are centralized, have high computational complexity, and consider only a subset of the appliances in the household [29, 63]. LocED overcomes these challenges by utilizing room-level user-occupancy information. The key idea is that an appliance usage in a household highly depends on the interaction between the occupants and the appliance. The energy-disaggregation complexity in LocED is reduced by constraining the appliances considered based on the current user location. LocED was evaluated across multiple real-world datasets and state-of-the-art algorithms. LocED achieves around 80% disaggregation accuracy across all appliances and around 20% increase in disaggregation accuracy as compared to the state-of-the-art algorithms.

- S. N. A. U. Nambi, A. R. Lua, and R. V. Prasad, *LocED: Location-aware Energy Disaggregation Framework*, in 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments, BuildSys, 2015.

Energy Apportioning – Chapter 3 Recent studies highlight the advantages of providing energy-consumption information to individual occupants of the household to promote energy savings and has the potential to reduce the energy consumption up to 20% [40, 54]. Current energy-disaggregation algorithms focus mainly on the energy consumption of buildings or households as a whole. However, providing user-level energy-consumption information in real time is a challenging task due to the need for collection of fine-grained information at various levels. To this end, we present the *Personalized Energy Apportioning Toolkit (PEAT)* that combines readily available data from the ubiquitous sensors (smart – meters, phones, watches) present in the household to derive fine-grained user-level energy-consumption information. PEAT combines energy disaggregation with indoor localization and activity monitoring to determine when an appliance is being used, and which occupant is currently using the appliance. PEAT employs simple classification techniques and inference algorithms to process and analyze data from the smart meter, smartphones and smartwatches to derive per-occupant, per-appliance energy usage. PEAT was empirically evaluated in a household and student housing. PEAT achieves 92.6% energy-apportioning accuracy with only location information of the occupants. Furthermore, the energy-apportioning accuracy is around 95% if both location and activity information are available.

- S. N. A. U. Nambi, A. R. Lua, L. A. G. Godinez and R. V. Prasad, *PEAT: Personalized Energy Apportioning Toolkit for Shared Spaces* [under review].

PART II – DEMAND REGULATION IN SMART HOMES AND BUILDINGS

This part presents personalized energy services that are targeted to shift and/or reduce energy consumption and cost by considering individual consumer preferences and comfort.

Demand shifting – Chapter 4 Energy utilities can balance energy supply and demand by nudging consumers to shift their demands to off-peak hours for load balancing and monetary benefits. Existing DR techniques aim to reduce electricity cost by scheduling the demand of the household based on the electricity prices (real time or day ahead) without considering consumer preferences and their appliance-usage patterns. We present a decentralized demand-regulation scheme that can, (i) determine appliance-level information and user preferences for appliance usage, using only aggregated energy consumption from the smart meters and (ii) propose a demand-scheduling algorithm that minimizes the user discomfort and electricity cost based on day-ahead hourly pricing. To tackle the changing consumer preferences, three coefficients (flexibility, sensitivity and dependency) are developed to analyze user preferences and appliance-usage patterns from historic energy-consumption data. The key idea is that the proposed day-ahead schedule should adhere to these three coefficients and resemble the historic energy-consumption pattern of the consumer (as the historic energy-consumption pattern would have been executed by the consumers previously). The proposed algorithm was empirically evaluated across multiple real-world datasets and saves up to 30% electricity cost.

- S. N. A. U. Nambi, A. R. Lua, and R. V. Prasad, *Decentralized Energy Demand Regulation in Smart Homes*, in IEEE Global Communications Conference, GLOBECOM, 2016.

Demand reduction – Chapter 5 HVAC (heating, ventilation, and air conditioning) and artificial lighting systems account for about 25-40% of electricity usage in residential and commercial buildings [19]. Thus efficient usage of the HVAC and lighting is a major step towards reducing energy consumption. Traditional energy-management systems operate within a conservative range or fixed set-point that is amenable to a large number of people providing only an average comfortable environment. To this end, we describe a smart system called *indoor Lighting and Temperature Controller (iLTC)*, that achieves a fine balance between preferred comfort levels of users and energy efficiency. iLTC decides energy-optimal operating set-points based on the knowledge derived from comprehensive temperature and lighting-comfort functions of individuals. The system learns the preferences of each individual based on human perception of comfort through the developed smartphone App. We evaluated iLTC with 21 participants housed in multiple rooms along with qualitative user evaluation. iLTC's set-point selection can reduce energy consumption up to 39% and 60% by the HVAC and lighting systems, respectively, compared to the fixed set-point mechanism.

- C. Sarkar, S. N. A. U. Nambi, and R. V. Prasad, *iLTC: Achieving Individual Comfort in Shared Spaces*, in International Conference on Embedded Wireless Systems and Networks, EWSN, 2016.

PART III – DEMAND REGULATION IN NEIGHBORHOODS

This part explores the design and development of novel energy services that can identify target consumers in a neighborhood for various DR programs. Data-driven techniques are proposed to determine consumer preferences and characteristics using only aggregated energy-consumption data.

Temporal demand regulation – Chapter 6 Energy consumption is highly influenced by consumer behavior and their characteristics. Rather than selecting all the households in a neighborhood for a DR event, an effective DR mechanism should first identify the set of target consumers and then apply the DR technique. In a neighborhood or community with thousands of households, heterogeneity in consumer characteristics hinders identifying consumers for specific DR programs such as reduction in average energy consumption, reduction in demand peaks, etc. Current techniques employ manual survey or behavioral economics to identify target consumers. To this end, we propose a novel mechanism – temporal demand regulation (TDR) – to analyze and classify households based on their historic energy-consumption data. We present a new concept of computing the demand states of each household, where a demand state measures either the demand – level, variation, or peaks. A generalized data-driven methodology based on clustering of historic consumption data from each household is designed for a local computation of the demand states. This methodology captures the temporal dynamics of demand and can be used to identify target consumers for DR programs. Further, an online self-regulation model for the adjustment of demands by targeted consumers is proposed. The selection criteria is governed by four temporal metrics, *viz.*, transition probability, temporal membership, temporal adaptability and temporal similarity. TDR is evaluated and validated using data from a real-world SG project consisting of more than 4,000 households [12].

- S. N. A. U. Nambi, E. Pournaras, and R. V. Prasad, *Temporal Self-Regulation of Energy Demand*, in IEEE Transactions on Industrial Informatics, vol. 12, no. 3, pp. 1196-1205, 2016.

Techno-Social Smart Grids – Chapter 7 Prevalent SG deployments and programs have found to be lacking in consumer awareness and engagement. Understanding what consumers want and how they behave is fundamental for developing sustainable SG services. This chapter fills the gap by modeling and analyzing the social context of consumers along with the energy-usage information. To this end, we propose ideas toward the development of a Techno-Social framework for Smart Grids (TSSG). The technological and social aspects of the consumers in SG are modeled and analyzed to develop consumer-centric SG services. We illustrate the benefits of modeling the techno-social aspects by forming communities directed towards particular goals. The novelty in formation of communities lies in fusing the technological and social data. These communities can *now* be targeted to promote energy awareness, provide tailored recommendations and community-specific tariff-rates.

- S. N. A. U. Nambi and R. V. Prasad, *Toward the Development of a Techno-Social Smart Grid*, in IEEE Communications Magazine, vol. 54, no. 11, pp. 1196-1205, 2016.

Data processing architectures – Chapter 8 While the deployment of smart meters is growing, the lack of adoption of energy services has hindered large-scale smart-grid deployments. This chapter explores how to cope with the overwhelming data generated from smart meters towards design and development of sustainable, smart energy systems. Hitherto several mechanisms have been proposed to tackle a specific architectural aspect, like communication, storage, processing requirement, etc. Currently, there is no comprehensive way to determine which data processing architectures suit best for which SG service. To address this, we investigate four data-processing architectures – centralized, decentralized, distributed, and hybrid – that best satisfy certain information management requirements, such as the accuracy and granularity of collected data, or the privacy level. We considered

realistic SG deployments in both dense (i.e., urban areas with $1.6M$ households) and sparse (i.e., rural areas with $476K$ households) environments. A detailed cost-benefit analysis of the proposed architectures is presented, which SG designers can use to discern the architecture that best fits their system requirements.

- S. N. A. U. Nambi, M. Vasirani, R. V. Prasad and K. Aberer, *A Cost-Benefit Analysis of Data Processing Architectures for the Smart Grid*, in WiMobCity workshop in conjunction with the International Symposium on Mobile Ad Hoc Networking and Computing, MobiHoc, 2014.
- S. N. A. U. Nambi, M. Vasirani, R. V. Prasad and K. Aberer, *Performance Analysis of Data Processing Architectures for the Smart Grid*, in 5th IEEE PES Innovative Smart Grid Technologies Conference, ISGT, 2014.

I

Fine-grained data collection

2

Energy Disaggregation

Worldwide total energy consumption in residential and commercial buildings is estimated to be 30-40% of generation [3] and is expected to rise due to increased use of appliances and electronic devices. A significant part of this could be reduced with better real-time information of appliance-level consumption statistics. With the help of appliance-level energy-usage information, one can provide personalized recommendations by identifying which appliances could most effectively reduce the total energy usage in a household. Furthermore, fine-grained appliance information can also be used to identify faulty or malfunctioning appliances that consume more energy than they should. Consequently, occupants know where the energy is being wasted.

The most common way of obtaining appliance-level information is by deploying sensors for each appliance. Such a deployment is intrusive, cumbersome to maintain and has a high cost. Alternatively, *non-intrusive load monitoring* (NILM) algorithms aim to break down a household's aggregate energy consumption into individual appliances [63]. NILM techniques are gaining popularity due to low-cost sensors for measuring energy usage, large-scale smart-meter deployments to obtain a household's aggregate energy consumption and inference algorithms proposed for energy disaggregation [56, 63, 85].

There still exist several challenges preventing NILM techniques from being widely adopted in households: (i) Most of the proposed mechanisms consider only a subset of appliances – a few high energy-consuming appliances – for disaggregation. This is due to the exponential computation complexity associated with the number of appliances, hence tractable only for a small number of appliances [29]. (ii) Several appliances with similar energy-consumption profiles may exist and moreover, each appliance may have multiple states. Thus modeling and inferring the states of appliances is not trivial. (iii) NILM is often performed in a centralized manner with third-party services or utilities having privacy-sensitive information of consumers. Commercially available NILM systems are required to send smart-meter data to a cloud service for energy disaggregation (for example, Bidgely, PlotWatt). This approach raises several issues related to scalability and privacy. (iv) Lastly, only a few NILM systems manage to provide near real-time energy disaggregation. The ones that do so require detailed information of the household and its occupants and generally utilize cloud services.

To tackle these problems, we present the *Location-aware Energy Disaggregation framework* (LocED) that utilizes user-occupancy information and aggregated energy data to derive accurate appliance-level information. The key idea is that an appliance usage in a household highly depends on the interaction between the occupants and appliance. The motivation for using occupancy/location information is threefold. *First*, by utilizing location information of occupants, the NILM algorithms can reduce the number of potential appliances considered for energy disaggregation. *Second*, by reducing the state explosion, the processing power and storage capacity required for disaggregation are also reduced, making NILM algorithms tractable and implementable. *Third*, with the large-scale proliferation of smartphones and wearables, it is now possible to monitor the location of the occupants (indoor localization) in a non-intrusive and cost-effective manner. LocED performs energy disaggregation at the household on a low-cost embedded system such as Raspberry Pi, due to which consumers' privacy-sensitive data is stored and processed locally.

We have instrumented a household in The Netherlands with several appliance-level sensors and a smart meter to monitor energy consumption. We have released the collected dataset – **DRED** (**D**utch **R**esidential **E**nergy **D**ataset) – which can be used to test the performance of disaggregation algorithms, derive appliance usage behavior and analyze demand response algorithms. The DRED dataset[1] and the LocED framework are made publicly available for the community to support additional analysis.

Contributions. The main contributions of this chapter are:

- We propose a novel real-time location-aware energy-disaggregation framework (LocED) to derive appliance-level information with lower computation complexity.
- We provide our data set – *DRED* (Dutch Residential Energy Dataset) – that contains appliance level and aggregated energy data from a household. The dataset also includes occupancy information and several ambient parameters.
- We propose several accuracy metrics to determine the efficacy of LocED both at house level and at appliance level. LocED was empirically evaluated across several publicly-available datasets.

2.1. RELATED WORK

Several NILM algorithms have been proposed in the literature to derive fine-grained appliance-level information. These algorithms rely on various techniques (supervised, semi-supervised or unsupervised) and also additional data [122]. We first provide details of the existing algorithms and then describe how our approach enhances the current state-of-the-art NILM algorithms.

NILM TECHNIQUES

Unsupervised NILM techniques use no prior knowledge of the appliances, but often require appliances to be manually labeled, and work on low frequency (i.e., 1 Hz) data. These techniques typically rely on accurate detection and modeling of the state change in the aggregate consumption data [26, 63, 68]. Several variants of factorial hidden markov models (FHMMs) to model the states of the appliances are proposed in [63, 68]. Furthermore, other machine learning approaches such as artificial neural networks (ANNs) and genetic

[1] `http://www.st.ewi.tudelft.nl/~akshay/dred/`

algorithms are also used [26]. These approaches are computationally intensive and exact inference from models with large number of hideen markov models is intractable.

Supervised NILM techniques assume that ground-truth appliance-level data is available to train and develop appliance models prior to performing disaggregation. Hart's algorithm identifies step changes in the aggregate electricity consumption and matches them with the appliance signature database to learn the states of the appliance [56]. Other approaches employ both real and reactive power measurements for energy disaggregation [118]. These algorithms require extensive training on appliance-level data to model the states.

Semi-supervised NILM techniques avoid the need to intrusively install sensors for deriving appliance signatures [85, 88]. Nambi et al. [85] propose a semi-intrusive approach to determine the most-optimal number of appliances to be monitored for accurate energy disaggregation. They explore the trade-off between monitoring accuracy and cost, and find the minimum subset of plug-level meters that maximize energy disaggregation accuracy. Parson et al. [88] utilize prior models of general appliance types, which are tuned to specific appliance instances using signatures extracted from the aggregate load. In general, due to the computational complexity involved in training and inference, these algorithms require systems with high processing power for energy disaggregation and, hence, are not suitable for low-power embedded systems.

ADDITIONAL DATA CONSIDERED IN NILM

NILM algorithms use additional information (either energy related or contextual data) to simplify energy disaggregation and enhance its accuracy. Recent algorithms use information on how loads are distributed across different power phases in a household [28, 118] or use transient and harmonic information with very high frequency sampling [55]. However, sampling at high frequency requires expensive hardware and determining appliance distribution across different phases is not trivial. Algorithms described in [69, 102] employ information provided by other sensors as additional input for energy disaggregation. Rowe et al. [102] propose an event detector to determine the state change by sensing the electromagnetic field (EMF) in the surrounding. Kim et al. [69] utilize signals from inexpensive sensors such as light and sound sensor placed near appliances to estimate power consumption. While the aforementioned approaches improve NILM accuracy, they also require additional deployment and maintenance of these sensors. Moreover, algorithms developed using additional data are generally constrained to a particular dataset or a household; consequently, making it nearly impossible to employ the algorithm with other datasets.

2.2. LOCATION-AWARE ENERGY DISAGGREGATION FRAMEWORK

In this section, we describe the usage of occupancy information to derive accurate appliance state information. Fig. 2.1 shows the block diagram of location-aware energy disaggregation.

2.2.1. USER OCCUPANCY MODELING

Occupancy information is generally used to develop efficient energy management systems for smart homes [77]. For example, occupancy information can be used to control the HVAC system efficiently or turn off appliances (lights) when user has left the room. We employ user-occupancy information to improve NILM algorithms by considering only those appliances that are in the current user location for disaggregation. Several direct and indirect ap-

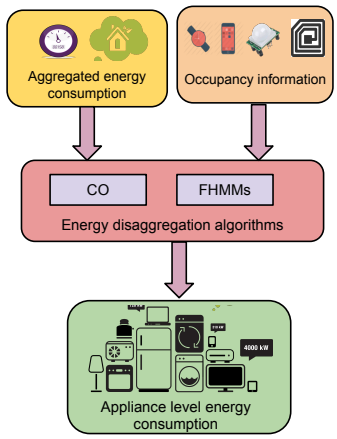


Figure 2.1: Location-aware energy disaggregation.

proaches have been proposed in the literature to derive user occupancy information [77]. Direct approaches employ low-cost sensors such as passive infrared (PIR), reed switches, RFID tags to determine room-level occupancy information. Even-though these approaches are cost-effective, they are cumbersome to maintain and intrusive in residential settings.

In this chapter, we employ an indirect approach for deriving occupancy information with the help of smartphones/wearables. Indirect approaches do not use additional hardware deployment, but rely on existing infrastructure for localization. Smartphones and wearables enable collection of received signal strength (RSS) from WiFi and/or Bluetooth (BT) radios in an indoor environment. In our DRED dataset (see Section. 2.3), we collected both Bluetooth (BT) and WiFi RSS information using occupant's mobile phones to infer user location. To save battery and also to derive accurate location, a radio scan is performed only upon detection of a user movement (i.e., change in accelerometer data or step detection).

The data stream from a radio scan includes the list of all visible access points (APs) and their RSS values along with the timestamp information. In case of a WiFi scan, the list of APs indicates the access points from the neighboring houses, whereas the BT scan indicates the Bluetooth beacons available in the house. Currently there exist several Bluetooth enabled devices in a household such as laptops, mobile phones, speakers, etc. Furthermore, in the near future most of the household appliances will be Bluetooth enabled [2]. Bluetooth enabled devices can now determine accurately indoor location information of the occupants. Classification techniques such as Bayesian, Support Vector Machines, K-nearest neighbor, decision trees, etc., have been proposed in the literature to derive room-level occupancy using RSS information. Our localization algorithm is based on Bayesian classification technique and has two phases *viz.*, training and testing phase as shown in Fig. 2.2. During the training phase, data is collected in each room to build a classifier model. In testing phase, new data from the scan is evaluated using the classifier model built to obtain the room-level occupancy information. Note that other localization algorithms can be employed in LocED framework to obtain occupants' location information.

[2] http://www.bluetooth.com/Pages/Smart-Home-Market.aspx

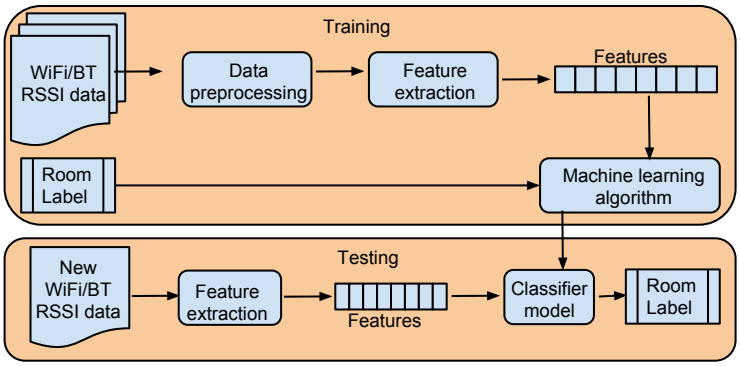


Figure 2.2: Indoor localization using WiFi/BT RSSI.

2.2.2. AGGREGATE ENERGY CONSUMPTION MODELING

We provide a brief description of the CO algorithm for energy disaggregation [56] and then, propose a modified CO algorithm used in our LocED framework.

Combinatorial Optimization (CO): The goal of an energy disaggregation algorithm is to provide estimates of actual energy consumed by each appliance from the aggregate energy consumption data. Let $\hat{y}_t^{(n)}$ be the estimated energy consumed and $y_t^{(n)}$ be the actual energy demand of each appliance n at time t. $\overline{y}_t$ represents the aggregate energy reading of the household. The ground truth state of an appliance is represented by $x_t^{(n)} \in Z \geq 0$ and $\hat{x}_t^{(n)}$ represents the appliance state estimated by the disaggregation algorithm. CO finds the optimal combination of appliance states, which minimizes the difference between the sum of predicted appliance power and the observed aggregate power. It is given by,

$$\hat{x}_t^{(n)} = \underset{\hat{x}_t^{(n)}}{\arg\min} \left| \overline{y}_t - \sum_{n=1}^{N} \hat{y}_t^{(n)} \right| \tag{2.1}$$

where N is the set of all appliances in the household and t is the current time period. The predicted energy consumption of an appliance $\hat{y}_t^{(n)}$ is then mapped to the closest appliance state $x_t^{(n)}$. This approach requires an appliance model, which includes power consumption details for each state of the appliance. This is further used during inference to predict the current state of the appliance. The computational complexity of disaggregation for T time periods is $O(TS^N)$, where S is the number of appliance states and N is the set of all appliances.

CO algorithm has several drawbacks. Firstly, this optimization problem resembles subset sum problem and is NP-complete. Furthermore, the computation complexity in CO increases exponentially with the number of appliances. Secondly, this algorithm does not differentiate between appliances with similar power consumption and appliances with similar states. Third, this algorithm assumes all the appliances in the household are being monitored and assigns some portion of energy to appliances even if they are not currently used, resulting in low disaggregation accuracy.

Modified Combinatorial Optimization: We propose a modified CO algorithm to overcome some of the drawbacks of original CO. Our modified CO algorithm, constrains the number of appliances considered for disaggregation based on the current location of the occupants.

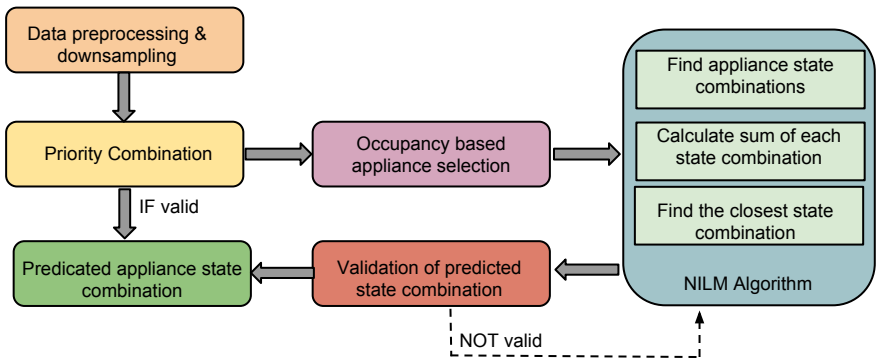


Figure 2.3: An overview of LocED Framework.

This results in exponential reduction in state space for disaggregation. Furthermore, we employ a crowd-sourced generic appliance model from the power consumption database. For example, the power consumption database provides crowd-sourced information on maximum and idle power for a wide range of loads indexed by type, manufacturer, and model number[3]. This information can be obtained *a priori* based on the appliances in the household from the manufacturers datasheet or crowd-sourced data, thus eliminating appliance level energy modeling. Furthermore, our modified CO algorithm requires to know the number of appliances and their location in the household. This metadata information is collected once during the deployment and, except from a few appliances like vacuum cleaner, hair dryer, the location of the appliances is generally static. Fig. 2.3 shows an overview of the proposed LocED framework.

1. Data preprocessing and downsampling: Our framework can handle various data sampling rates and is designed to work with several datasets. In general, during data collection there might be gaps in the data due to sensor malfunction, network connectivity, etc. Hence, it is important to preprocess these gaps either by removing them or using statistical models such as smoothing, interpolation, forward filling, etc. Furthermore, different datasets include different sampling intervals typically from 1 second to 15 minutes. LocED applies a downsampling mechanism similar to NILMTK [29], to filter transients that occur due to high starting current of an appliance.

2. Priority combination: In original CO, at each time period the algorithm tries to find the set of appliances, which are closest to the current aggregated energy consumption. This may result in different set of appliances being used in each time period. For example, at time period 't', CO may determine appliance TV and microwave are being currently used and at time period '$t+1$' it may select fan and microwave. This is due to the fact that TV and fan may have similar energy consumption profiles. This result would mean TV is switched ON in one minute and switched OFF the next minute and so on. Hence, it is necessary to preserve consistency in selection of appliances during consecutive state estimations. LocED defines a *priority combination* that is the set of appliances which are assumed to be currently running. This information can be retrieved from the last iteration of NILM algorithm. At each time period, LocED first evaluates the priority combination to check whether the

[3]The Power consumption database. [Online] http://www.tpcdb.com/

sum of all appliances in the priority combination matches the current aggregated value. If the difference between the sum of priority combination and the aggregated energy is within a threshold δ, then the current priority combination is retained as the predicted set. LocED evaluates the following equation to determine whether the current priority combination of appliances are still valid or not, $[|\overline{y}_t - \sum_{n=1}^{K} \hat{y}_t^{(n)}| \leq \delta]$, where K is the set of appliances present in the priority combination and δ is the variation threshold. The variation threshold parameter ensures small fluctuations in aggregate power has minimal effect. Since these fluctuations vary for different appliances based on their power rating, the δ value needs to be adaptive. The δ value can be obtained by analyzing the energy consumption profiles of the appliances. However, when the difference between current priority combination and aggregate consumption is greater than δ, LocED finds the new set of appliances that are used.

3. Occupancy based appliance selection: When the current priority combination does not match the aggregate energy consumption, LocED estimates the set of appliances that could be currently used. This stage identifies the set of appliances which are present in the current user location. For example, if the current location information of all occupants includes Kitchen and Living room, only appliances present in these locations are considered valid during that time period for energy disaggregation. In general, the appliances considered for evaluation at a particular time period include, (i) appliances present in the current location of the occupants; (ii) appliances that are already “ON"; (iii) appliances that are always “ON”, these are autonomous appliances such as Refrigerator; and (iv) appliances that can be remotely controlled such as lights and other smart appliances. We refer to these appliances as “constrained set of appliances”. LocED uses this constrained set for energy disaggregation rather than the complete set of appliances present in the household. If for a time period, there is no occupancy information available all appliances present in the household are considered for evaluation.

4. CO based NILM algorithm: In this chapter, we employ modified CO algorithm to find the optimal combination of appliance states. We calculate the sum of all possible state combinations from the constrained set and select the closest combination of appliances that match the aggregated energy consumption. The computational complexity of disaggregation for T time periods in LocED is $O(TS^{N_c})$, where S is the number of appliance states, N_c is the constrained set of appliances and $N_c \leq N$. This reduced computational complexity enables LocED to determine the state of appliances in real-time. As mentioned earlier, other NILM algorithms can be used at this stage to infer the state of the appliances from the constrained set. For example, in case of FHMMs the constrained list of appliances can be used during decoding the HMM state sequence.

5. Validation: We now validate the set of appliances predicted in the previous stage. Using occupancy based appliance selection, LocED ensures we do not turn “ON" an appliance when user is not present in that location. However, validation stage ensures not to turn “OFF" an already “ON" appliance when the appliance location is different than the current user location (except remotely controllable appliances). Moreover, this depends on the type of the appliance. In this chapter, we broadly classify the set of appliances into: (i) User dependent appliances – appliances that require user interaction to turn “OFF", for example, TV, fan, etc., and (ii) User independent appliances – appliances that can turn “OFF" themselves and require no user interaction, for example, microwave, washing machine, dishwasher, etc. If the set of appliances selected in the previous stage involves one or more user

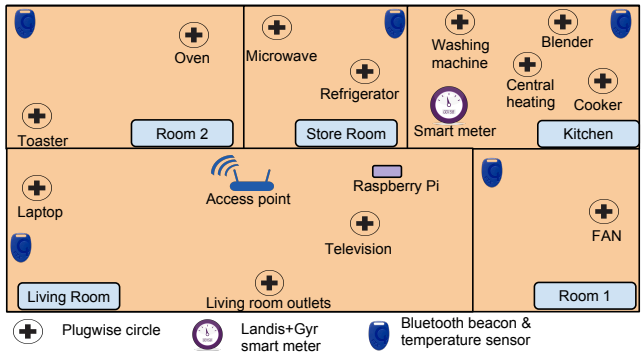


Figure 2.4: Deployment setup in DRED.

dependent appliances being turned "OFF" and even if the occupants location differs from the appliance location, validation stage eliminates this combination of appliances. LocED then selects the second closest combination from the previous stage and re-validates.

2.3. THE DRED DATASET

The dataset consists of several sensors measuring electricity, occupancy and ambient parameters in a household for over 8 months. The sensors were carefully installed to avoid any inconvenience for the occupants. We refer to this dataset as **DRED** (**D**utch **R**esidential **E**nergy **D**ataset). Fig. 2.4 shows the layout of our deployment along with location of the sensors and appliances in the household.

Electricity monitoring: We used off-the-shelf sensors to monitor energy consumption at 1 Hz sampling frequency.

(i) *Mains level*: We installed a smart electricity meter from Landis+Gyr E350 to measure the aggregate energy-consumption information of a household. The data from the smart meter was retrieved using Plugwise Smile[4].

(ii) *Appliance level*: We used off-the-shelf smart plugs from Plugwise circle[5] to collect appliance-level energy-consumption data. 12 smart plugs were installed to monitor the appliances across the household, *viz.*, (1) Refrigerator, (2) Washing Machine, (3) Central Heating, (4) Microwave, (5) Oven, (6) Cooker, (7) Blender, (8) Toaster, (9) TV, (10) Fan, (11) Living room outlets, and (12) Laptop.

The plugs installed in the household communicate via Zigbee protocol by forming a mesh network. We use an open source library python-plugwise to query the data from the plugs at 1 Hz frequency. A Raspberry Pi was deployed locally to generate periodic queries and to store the data. Furthermore, this data is also sent to a server for making it available for the research community.

Ambient monitoring: Apart from collecting energy related data, in our deployment we also collected room level indoor temperature, outside temperature, wind speed, precipitation and humidity. We deployed low-cost Bluetooth beacons from Gimbal[6] with in-built temper-

[4]Smile:`https://www.plugwise.com/smile-p1`
[5]Circle:`https://www.plugwise.com/circle`
[6]https://store.gimbal.com/collections/beacons/products/s10

ature sensor for each room and also one outside the house. These beacons have a battery lifetime of 4 to 5 months. A smartphone and smartwatch application in Android was developed to read the data from these beacons every 1 minute. The wind speed, precipitation and humidity data was collected from the publicly available Royal Netherlands Meteorological Institute (KNMI) website every hour[7].

Occupancy monitoring: In our deployment, we scan both visible WiFi access points and the Bluetooth beacons present in the household for indoor localization every 1 minute. This data is further used with different machine learning algorithms to determine the indoor room level location of occupants. The room-level location inferred from the localization algorithm is also made available. For WiFi based localization, no additional infrastructure is deployed, however, for BT based localization we deployed the BT beacons, which could be further replaced by the smart Bluetooth enabled devices.

Household metadata: Our dataset also includes household metadata such as number of occupants, house layout, mapping between appliance and location. This metadata is generally useful for NILM algorithms. Further details on the metadata can be found in [2].

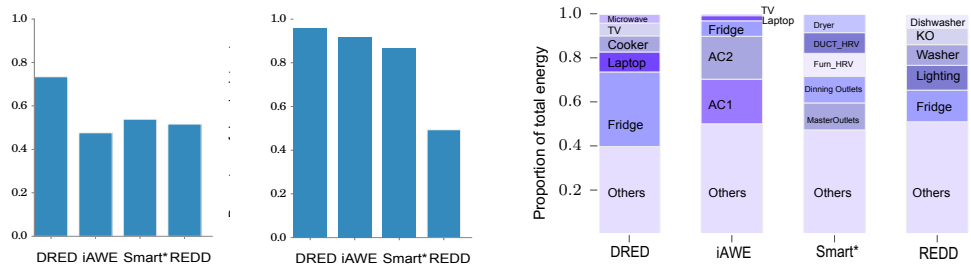
DRED dataset extends the current publicly available datasets: (i) In DRED, almost all appliances are monitored and has very constant baseline consumption. Baseline consumption includes appliances which are occasionally used (guest devices) or not monitored. Popular datasets such as REDD, Smart*, iAWE and ECO has very high and varying baseline consumption. This variation significantly hinders the performance of NILM algorithms. (ii) DRED dataset has less than 5% dropout rate in energy data. Dropout rate indicates the missing data due to communication issue or sensor faults. Most of the other datasets have around 10-20% dropout rate apart from Smart*. (iii) Even though ECO, Smart*, iAWE datasets include occupancy data, they have large gaps and missing data. However, DRED uses an indirect sensing approach for obtaining room-level occupancy information and has high data availability rate.

2.4. EVALUATION

We provide performance evaluation results of the proposed framework across multiple datasets to support wide-adoption and also to validate our work. Our framework imports data from DRED dataset and also other popular publicly available datasets such as REDD (House 1), Smart* and iAWE. Hence, we show the performance results across four datasets collected in different countries.

Dataset Statistics: Each dataset includes data from different set of appliances and for varying time duration. Fig. 2.5a shows the percentage of total energy measured at the appliance level for all days in the dataset. Most of the datasets do not monitor all the appliances in the household, leading to large (sometimes more than 50%) unaccounted energy in the aggregated consumption data. Furthermore, the variation of this unaccounted energy data significantly reduces the accuracy of disaggregation algorithms. DRED has around 75% of energy sub metered and all other datasets have around 45% of energy measured at the appliance level. Furthermore, DRED dataset has around 95% data availability rate at the mains level. The other datasets have 90%, 86% and 50% mains data availability rates for all the days (see Fig. 2.5b).

[7] KNMI:`http://www.knmi.nl/climatology/daily_data/selection.cgi`



(a) Energy submetered. (b) Mains availability. (c) Top-5 appliances.

Figure 2.5: Data characteristics across different datasets.

In general, only a few appliances constitute the majority of power consumed in a household. Hence, it is necessary to derive accurate information of these high power consuming appliances during energy disaggregation. Fig. 2.5c shows the proportion of energy consumed by top-5 appliances and other appliances present in the household across datasets. It is interesting to see the variation of top-5 appliances across datasets, indicating the varying preference of appliance usage in different countries. The top-5 appliances in DRED cover around 60% of total energy consumed.

Finally, since LocED relies on the occupancy information collected, it is important to find the occupancy data availability rate. The occupancy availability rate is the ratio of total number of occupancy data recorded over the total number of expected occupancy data. DRED, iAWE and Smart* has occupancy rate of 81%, 76% and 36% respectively.

Accuracy Metrics

Several accuracy metrics both at house level and at appliance level are considered for evaluation of LocED.

1. Fraction of total energy assigned correctly (FTE): It measures the fraction of energy correctly assigned to an appliance and is one of the common accuracy metrics for NILM algorithms [29, 72]. FTE is the overlap between the actual fraction of energy consumed by each appliance and the fraction of energy assigned to each appliance. It is defined as,

$$FTE = \sum_n min\left(\frac{\sum_n y_t^{(n)}}{\sum_{n,t} y_t^{(n)}}, \frac{\sum_n \hat{y}_t^{(n)}}{\sum_{n,t} \hat{y}_t^{(n)}}\right), \tag{2.2}$$

where $n \in \{1,..,N\}$ and N is the total number of appliances. Also $t \in \{1,..,T\}$ and T is the total time period considered.

2. Total disaggregation error (T_e): Total disaggregation error is the difference between the total energy consumed by all appliances and the actual energy consumed by the appliances, normalized by the total energy consumed. It is given by,

$$T_e = \frac{\sum_{n,t} |y_t^{(n)} - \hat{y}_t^{(n)}|}{\sum_{n,t} y_t^{(n)}} \tag{2.3}$$

We employed the functions provided in NILMTK for calculating FTE and T_e metrics.

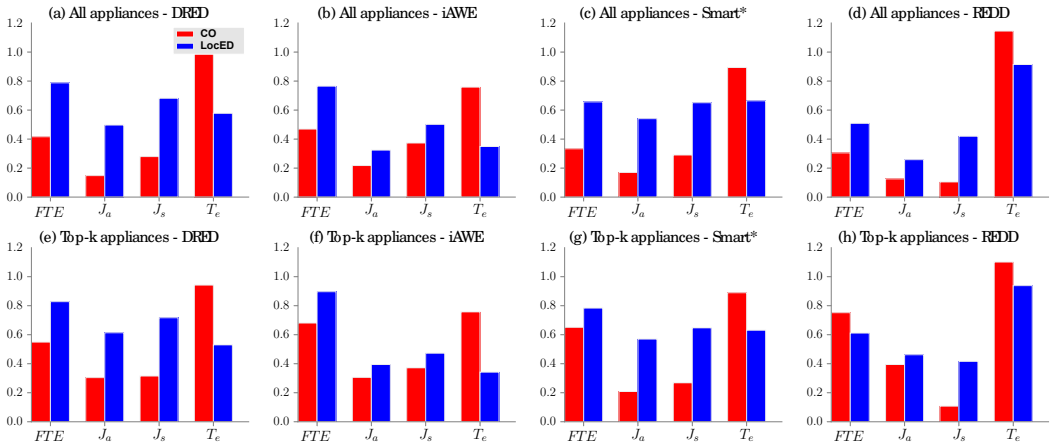


Figure 2.6: Disaggregation performance of CO and LocED across datasets (1 week).

3. Number of appliances identified correctly (J_a): Jaccard similarity coefficient is used to measure the similarity between the predicted set of appliances (J_a^p) and the actual set of appliances (J_a^a) used over a time period. J_a measures the percentage of appliances correctly identified by the disaggregation algorithm. It is given by,

$$J_a = \frac{|J_a^p \cap J_a^a|}{|J_a^p \cup J_a^a|} \tag{2.4}$$

4. Number of appliance states identified correctly (J_s): It measures the similarity between the predicted set of appliance states (J_s^p) and the actual set of appliance states (J_s^a). It is given by,

$$J_s = \frac{|J_s^p \cap J_s^a|}{|J_s^p \cup J_s^a|} \tag{2.5}$$

2.5. RESULTS

We now show the performance of LocED and original CO algorithm. To ensure fair comparison, both LocED and CO utilize the same appliance model from the crowd-sourced database as described in Section 2.2.2. Since the model and make of an appliance varies from one dataset to another due to the geo-location of data collected in these datasets, applying a generic model across all datasets is challenging. LocED uses a crowd-sourced appliance model from the power consumption database based on the manufacturer and model number of an appliance. In our evaluation, we used data obtained using direct sensing (PIR sensors) in Smart*, iAWE datasets and also data from indirect sensing in DRED dataset. Furthermore, we used an adaptive δ value for determining the priority combination. The values of δ was determined based on the appliance type. From our experimentation, we found that appliances with low power consumption have lower noise and smaller variation in their energy consumption and appliances with high power consumption have large variation due to the noise associated. Furthermore, the δ value can be also used to ac-

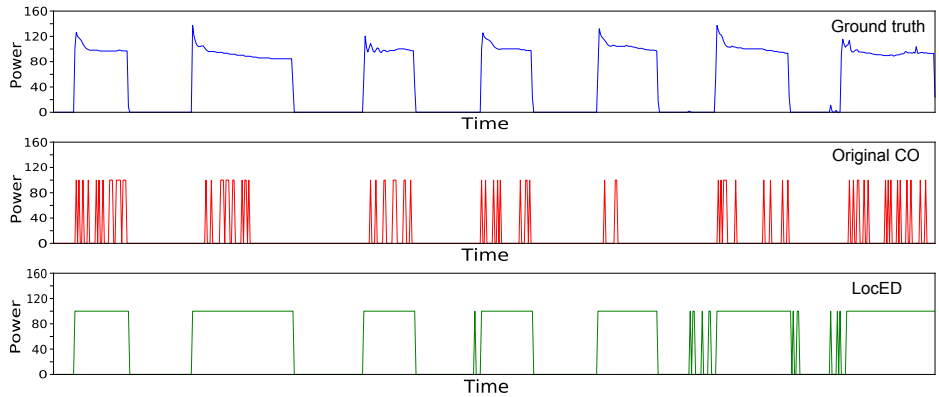


Figure 2.7: Original and disaggregated energy profile of refrigerator using CO and LocED.

count for unmonitored appliances or guest appliances by modeling the historic household energy consumption.

Fig. 2.6 shows the disaggregation performance of CO and LocED across the house level accuracy metrics. We considered one week of data (the week with highest data availability rate) across all the four datasets. In general, FTE, J_a and J_s can vary between 0 and 1, and T_e can take any non-negative value. It can be seen that, LocED performs significantly better across all the datasets for all the metrics. LocED performs better than CO mainly due to two reasons, (i) LocED ensures that the predicted set of appliances does not vary significantly for consecutive time periods, thanks to *priority combination*. (ii) LocED constrains the number of appliances considered to disaggregate based on occupancy information ensuring similar appliances from a different location are not selected.

Fig. 2.6(a) shows that in DRED, LocED correctly assigns up to 80% of energy to all appliances, which is 40% more compared to CO. Furthermore, it determines more than 25% of correct appliances and states than original CO. Fig. 2.6(c) shows more than 30% improvement across all metrics for Smart* dataset and similar trends can be seen in iAWE and REDD datasets. LocED also has much lower T_e across all datasets compared to CO. Fig. 2.6(e),(f),(g),(h) show the disaggregation performance of CO and LocED for top-k (k=5) appliances. As mentioned previously, disaggregating accurately top energy consuming appliances would be very beneficial to reduce cost and manage energy efficiently. It can be seen that LocED correctly assigns upto 89% of energy to the top-k appliances in iAWE and around 80% in DRED and Smart* datasets. Furthermore, the number of appliances and states identified is also higher compared to original CO. In DRED and Smart* datasets the number of appliances and states determined is more than 30% compared to original CO.

Fig. 2.7 shows the original power (in W) consumed by the refrigerator (top) and the resulting disaggregation output using the original CO (middle) and LocED (bottom). CO has a interrupted load profile due to its sensitivity to small changes in aggregated power, however, LocED overcomes this with the help of priority combination and the δ parameter described in Section 2.2.2.

Table. 2.1 shows the percentage increase in disaggregation accuracy of LocED compared to CO for all the days across the datasets with all appliances and top-k appliances. It can be

Dataset	All Appliances				Top-k Appliances			
	FTE	J_a	J_s	T_e	FTE	J_a	J_s	T_e
DRED	30.5	28.7	36.6	-37.9	22.4	23.3	36.5	-41.2
iAWE	8.5	3.2	2.2	-7.9	14.8	3.3	4.5	-9.6
Smart*	29.3	28.3	28.6	-14.4	1.9	28.1	30.4	-18.6
REDD	11.4	12.7	27.6	-13.6	-22.1	5.3	27.3	-8.8

Table 2.1: Percentage increase in performance of LocED over CO (all days).

seen that FTE improvement of 30%, 9%, 30%, and 12% is obtained for all days considered in DRED, iAWE, Smart* and REDD datasets respectively. Similarly, number of appliances correctly identified improves over 30% for all days considered in DRED and Smart* datasets respectively. In DRED, FTE improvement of 22% was achieved for top-k appliances and number of appliances and states identified improved by 23% and 36% respectively. The negative T_e shows the percentage reduction in total error achieved by LocED. The FTE for top-k appliances in REDD dataset is lower for LocED. This is likely due to wrong inference of locations from appliance ground truth data.

Even with additional location information there are errors associated with disaggregation due to several factors. In most of the datasets, due to lack of knowledge on number of appliances and lack of monitoring of all appliances in the household, there exists a significant amount of unaccounted energy in the aggregate consumption. Only DRED dataset monitors almost all appliances and has a very low variation in baseline consumption. Moreover, the percentage of occupancy information available plays an important role in improving the accuracy. Only DRED and iAWE have more than 70% of occupancy data available. Furthermore, if the occupants are spread out across the building or if all the appliances are close to one another, then the benefits of using location information for disaggregation is less. However, in residential settings as seen from the above datasets these cases arise occasionally. In our evaluation, we showed that even with very less location information, LocED was still significantly able to improve the disaggregation accuracy. Furthermore, the framework proposed can include other contextual information such as room temperature, number of users, etc. to further improve energy disaggregation accuracy.

Finally, we also computed the average number of state combinations evaluated in each dataset by CO and LocED to disaggregate. Original CO has a fixed number of state combinations depending upon the number of appliances and its states. However, for LocED the number of appliances considered varies and is determined based on the constrained set of appliances. In iAWE and Smart* the average state combinations to be evaluated for disaggregating a value is 59049 and 8192 for CO and it is 162 and 60 for LocED. Similarly, in DRED 104976 combinations was evaluated by CO and LocED evaluated only 10 combinations on average. It can be seen that across all datasets the average number of state combinations evaluated by LocED is drastically reduced, consequently, decreasing the computation complexity for real-time disaggregation.

2.6. CONCLUSIONS

In this chapter, we proposed a *novel* location-aware energy-disaggregation framework (LocED) to derive accurate appliance-level data from aggregated household energy data. We de-

veloped a modified combinatorial-optimization algorithm to infer the state of the appliances accurately. We also presented a comprehensive dataset (DRED), that can be used to test the performance of energy-disaggregation algorithms, derive appliance-usage behavior and analyze demand-response algorithms. We evaluated LocED across multiple publicly available datasets such as DRED, iAWE and Smart*. Our evaluation shows that around 80% disaggregation accuracy can be achieved for all appliances on the DRED and iAWE datasets. Furthermore, up to 90% accuracy is achieved when only the top-5 appliances are considered for disaggregation in DRED and iAWE.

LocED forms the basis to derive appliance-level information for various personalized energy services described in subsequent chapters. In Chapter 3, LocED is used to identify per-occupant energy-consumption information and in Chapter 4, it is used to effectively schedule appliances to maximize user comfort and minimize energy cost.

3

Energy Apportioning

The energy disaggregation techniques presented in Chapter 2 provide real-time feedback on appliance-level energy consumption. While these techniques help in understanding energy consumption in a building, they lack the ability to provide energy footprint of individuals. A recent study highlighted the advantages of providing energy-consumption information to individuals of a household to raise awareness, and to promote energy savings. This has the potential to reduce up to 20% of the total energy consumption [31, 54]. In shared spaces – such as student house, office environments, multi-occupant household – lack of individual occupant consumption data necessitates even-splitting of energy cost. This results in inefficient energy usage where occupants maximize their energy usage by taking advantage of others [53, 54]. Hence there is a need for *personalized energy disaggregation* systems moving from appliance level towards user-level information. The disaggregated individual energy-usage information can be used to develop better energy-management techniques apart from raising awareness.

The current literature focuses mainly on energy consumption of buildings or households as a whole. Aggregated energy information of households fails to answer questions such as,*"how much energy each occupant has used today?", "who amongst us burns most of the energy?", "which occupant is more energy efficient in the household?"*. Currently there is no comprehensive means of providing information to individual occupants on their energy consumption. Hence, it is required to design an energy-apportioning system that disaggregates household total consumption into per-occupant, per-appliance level.

Recently, there have been few research efforts [37, 57, 75, 98, 104] that aim to study per-occupant energy footprints in various settings. However, providing energy-consumption information of individuals, in shared spaces, is a challenging task due to the following requirements: (i) Collection of fine-grained information at various levels. This fine-grained information includes *which* occupant performed *what* activity and *where* in the household. (ii) Deployment of additional sensors to determine the current location of occupants and which occupant is using a particular appliance. This involves additional cost for deployment and is cumbersome to maintain. (iii) Identifying in real time the correct occupant using an appliance when multiple occupants are present in the same location. Current approaches apportion the energy equally to all the occupants in that location [57, 104]. (iv)

Most of the techniques proposed hitherto are centralized with either third-party services or energy utilities collecting privacy-sensitive information of consumers. This raises several issues related to scalability and privacy. (v) The resolution of apportioning may vary depending upon the environment i.e., residential household – where some of the shared appliances are for the total family (e.g., kitchen utilities and fridge usage during cooking); student house – where a shared appliance is used by only one occupant at a time (e.g., kitchen utilities and fridge usage during cooking by a single occupant). In the former case, one might not be interested in apportioning energy usage during cooking event, however, for the latter it may be important to apportion the energy spent during cooking. Hence, identifying the resolution of apportioning is an important issue.

In this chapter, we present the **P**ersonalized **E**nergy **A**pportioning **T**oolkit (PEAT) that combines readily available data from the ubiquitous sensors present in the household to derive fine-grained occupant-level energy-consumption information. Specifically, we use a single smart meter's energy data to derive fine-grained appliance-level energy information. The Modified Combinatorial Optimization (ModCO) algorithm presented in Chapter 2 is used for energy disaggregation. Furthermore, Internet of Things (IoT) devices such as smartphones and smartwatches with WiFi radios are used for indoor localization and to determine the activities performed by the occupants. A plethora of sensors are now embedded within wearable devices making it possible to gather information about users and their activities. In this chapter, data from these sensors are used to detect the *micro-activities* performed by occupants. Micro-activities refer to the activities performed by occupants such as opening a microwave door, opening the refrigerator door, using a laptop, switching on/off an appliance. PEAT combines NILM techniques with WiFi-based localization and activity monitoring to determine *when* an appliance is being used, and *which* occupant is currently using the appliance. Our system was extensively evaluated in two real-world multi-occupant buildings, *viz.*, (i) a student house and (ii) a residential house.

Contributions. The main contributions of this chapter are:

- We present a personalized energy-apportioning toolkit (PEAT) to derive real-time per-occupant energy footprints in shared spaces.
- We describe our inference algorithm, which models the association between appliances and users to determine the dynamics of appliance usage.
- We provide an extensive experimental evaluation of PEAT from two multi-occupant buildings, *viz.*, a student house and a residential house. PEAT was empirically evaluated in both the settings for over two months.

3.1. RELATED WORK

Energy apportioning in shared spaces covers a broad range of research areas from appliance monitoring, user monitoring to personalized energy monitoring. Please refer to Section 2.1 for related work on appliance-level monitoring. We now describe in detail the state-of-the-art techniques proposed for user and personal energy monitoring in shared spaces.

User monitoring

User-occupancy detection is a crucial element in developing user-centric energy management services [43]. Several direct and indirect approaches have been proposed to de-

Work	Environment	Architecture	Appliance monitoring	User monitoring	Apportionment policies	Additional sensor
[57]	Office	Centralized	Appliance level	Security access logs	Cannot distinguish between occupants	NA
[75]	Residential	Centralized	Appliance level	room level localization using doorway sensors	Evenly apportions to all occupants present in the room for shared appliances	Doorway sensors
[37]	Office	Centralized	Appliance level	proximity based using magnetic inductance	Assigns usage to the nearest occupant	Temperature, special wearables
[104]	Residential	Centralized	Audio based appliance detection	WiFi based room level localization	Evenly apportions to all occupants present in the room for shared appliances	Microphone for appliance detection
[98]	Residential	Centralized	Appliance level with unique sensors	RFID anklets and RFID antennas	Applies heuristics to determine per-occupant footprint	RFID antennas
PEAT	Office & residential	Distributed	Aggregate data using NILM	WiFi based localization, smartwatch for activity recognition	Apportions energy based on historic consumption pattern & activities performed	None

Table 3.1: Comparison of various efforts towards energy apportioning.

rive user-occupancy information [23, 99]. *Direct* approaches employ low-cost sensors such as passive infrared (PIR), reed switches, and RFID tags to determine occupancy information [99]. Even though these approaches are cost effective, installing and maintaining these sensors in a household is intrusive and cumbersome. Furthermore, these techniques do not distinguish between different occupants, when the occupants have similar characteristics. *Indirect* occupancy-monitoring approaches employ WiFi or Bluetooth (BT) fingerprinting using smartphones to derive room-level occupancy [23]. In this chapter, we use the existing infrastructure like WiFi access points (AP) deployed in a household, along with smartphones to derive the current locations of the occupants. We employ simple classification techniques that run locally on smartphones to model and train the data collected from the WiFi scans.

Hitherto, several research efforts used smartphone sensors to determine user activities such as walking, climbing stairs, running, and jumping. However, according to a recent forecast [4] smart wearables are expected to grow from 9.7 M in 2013 to 135 M in 2018. The forecast also predicts that around 87% of these wearables account for smartwatches and wrist bands. Motivated by the large-scale penetration of smartwatches and their increasing sensing capabilities, we employ smartwatches to derive user activities in the household. Smartwatches provide an opportunity to accurately identify the micro-activities performed by the occupant such as opening a microwave door, opening the refrigerator door, and switching on/off an appliance. Unlike existing activity-monitoring techniques that require additional sensors carried by occupants, we argue that sensors present in the smartwatch are sufficient to determine micro-activities. Moreover, with both WiFi and BT radios available on smartwatches, one can use them for indoor localization.

Personalized energy monitoring

Lack of per-occupant energy footprints has resulted in even splitting of energy costs and negligent energy usage in shared spaces [31]. Table. 3.1 provides a concise overview of state-of-the-art techniques proposed for energy apportioning. Hay et al. [57] investigate energy apportioning in an office building. They propose static and dynamic policies to apportion shared energy usage. However, these policies assign energy evenly by determining the number of people inside the building. In residential settings, Lee et al. [75] propose a personalized energy auditor to apportion energy with the help of smartphones and doorway sensors. Doorway sensors were used to determine which occupant is present in a room. They classified the appliances into “personal" and “shared". The policy for apportioning assigns the personal appliance usage to that individual, whereas the shared appliance usage is evenly split across all occupants present in the room. The setup and installation of doorway sensors is cumbersome. Furthermore, metadata information such as occupants characteristics (e.g., height) and intrusive appliance-level metering was considered during apportioning. Cheng et al. [37] present a model to determine the association between human activities and observed energy consumption. They use additional sensors such as light and sound sensors to determine user movements and location. Saha et al. [104] propose mechanisms to combine smartphone data with electricity data for accurate activity detection and energy apportioning. They use WiFi-based localization to determine the locations of occupants and collect audio samples from the microphone continuously to determine which appliance is being used. The proposed models require extensive data collection and training for each appliance, hindering the applicability of the work in other households. Ranjan et al. [98] map energy apportioning to a *fixture assignment* problem to determine per-occupant energy usage. Fixture assignment is done by determining the unambiguous assignments and then learning usage patterns. However, they use custom-made RFID anklets and RFID antennas for indoor location tracking and deploy sensors for appliance monitoring.

3.2. PEAT

Considering the limitations of the state-of-the-art, we propose an energy apportioning toolkit that integrates smartphones, smartwatches and smart-meter data with minimal user intervention to derive real-time per-occupant energy footprint. Fig. 3.1 shows the system overview of PEAT. The toolkit consists of four major components: (i) *Appliance monitoring*, (ii) *User monitoring*, (iii) *Appliance-User modeling* and (iv) *Online evaluation*. We describe each component in detail below.

3.2.1. APPLIANCE MONITORING

Appliance monitoring component with the help of NILM algorithm determines the state of each appliance. A change in state of an appliance from “OFF” to “ON" is considered as an event trigger. Event triggers represent the appliances, which are currently being used by the occupants in the households. In Section 2.2 we described our Modified CO algorithm. Since location of occupants are not known yet in the system, PEAT employs Modified CO without Step 3 i.e., occupancy based appliance selection. This work distinguishes the appliances in the household into three categories: (i) *Personal appliances*: Personal laptop, hair dryer, smartphones, etc. (ii) *Shared appliances*: Television (TV), kitchen utilities, microwave, etc. (iii) *Baseline appliances*: appliances that are always ON – modems, routers, refrigerator, etc.

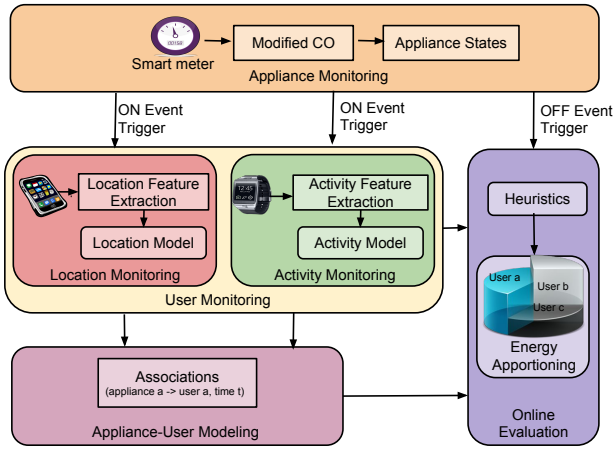


Figure 3.1: System overview of PEAT.

3.2.2. User monitoring

User monitoring component is activated upon the reception of an event trigger from appliance-monitoring component. User monitoring determines the current location and activities performed by all the occupants. Smartphones are used for indoor room-level localization and smartwatches are used for micro-activity recognition. Note that, only a single smartwatch can be used for both localization and activity recognition.

1. Location monitoring

This work focuses on smartphones/watches to determine indoor location of occupants due to the following reasons: (i) smartphone/watch is personally associated and carried by a user, (ii) change in sensor information such as accelerometer can be used to detect user movements and (iii) localization techniques can use WiFi and/or Bluetooth radios to identify user location. The event trigger from the appliance monitoring component initiates the data collection for indoor localization. The data stream includes a scan of visible WiFi access points (APs) and their Received Signal Strength (RSS) along with the timestamp. The list of APs indicates the access points from the neighboring houses. To save battery and also to derive accurate location, a scan is performed only upon detection of a user movement (i.e., change in accelerometer data or step detection).

Naïve Bayesian classifier for localization: Classification techniques such as Bayesian, Support Vector Machines, K-nearest neighbor and decision trees, have been proposed in the literature to derive room level occupancy using RSS information. Our localization algorithm is based on Bayesian classification technique[1] and has two phases *viz.*, training and testing phase. During the training phase, data is collected at each room to build a classifier model. This phase is also called the fingerprinting stage, where data from WiFi scans are used to learn the available APs and their RSS at different locations.

Feature extraction: The collected data from WiFi scan is then used to derive features for the classifier model. Feature vectors are derived by using multiple scans performed. In this work, we use four WiFi scans to derive feature vectors such as max, min, mean, standard

[1] Note that other classification techniques can also be used to derive location information.

deviation of the signal strength for each available AP. Feature vector **l** for k access points is represented as,

$$l_t = < rss_t^{max}(1), rss_t^{min}(1), rss_t^{mean}(1), rss_t^{std}(1), ..., \\ rss_t^{max}(k), rss_t^{min}(k), rss_t^{mean}(k), rss_t^{std}(k) > \tag{3.1}$$

Building classifier model: Feature vectors obtained are provided as input to the classifier algorithm to derive the class labels. The classifier model generates a probability distribution function (PDF), which is further used in testing phase to determine the class label (room location).

2. Activity monitoring

With the large-scale penetration of wearable devices, it is now possible to analyze user activities accurately. These devices provide an opportunity to identify precisely the micro-activities performed by the occupant such as opening a microwave door, opening the refrigerator door, using laptop, switching on/off an appliance. Smartwatch of an occupant is used to determine the activity performed. The different sensors used for acquiring relevant activity information include accelerometer, gyroscope, magnetometer, tilt, rotation and linear acceleration. The activity feature vector, **a**, includes data from both time and frequency domain. Features considered are mean, standard deviation, variance from x, y, z axes along with magnitude, fundamental frequency, zero crossing and step counter.

The event trigger initiates the data collection for activity monitoring. Similar to location monitoring, the activity monitoring also has training and testing phase. During the training phase, multiple samples of activity feature are collected for each micro-activity. The activity feature vectors are then used by the classifier to model and determine the class-label associated with each activity feature. The class label indicates the micro-activity performed by the occupant. In the testing phase, each activity feature is evaluated by the activity model to determine the micro-activity performed. Location and activity monitoring determine *where* the occupant is and *what* activities are performed by the occupant when an event trigger is received.

3.2.3. APPLIANCE-USER MODELING

The appliance-user modeling studies the user association with the appliance. The objective of the appliance-user modeling is to determine the occupants that are currently associated with the appliance being used. If the current appliance being used is personal, then it can be assigned to the relevant user. However, some appliances are shared by all the occupants at different time periods (e.g., Microwave) and some appliances are used by all the occupants at the same time period (e.g., TV). Hence it is important to not just determine where the occupant currently is but also to determine the activities performed by the occupant. PEAT utilizes location and activity information to determine the user association with the appliance.

To study the association between the appliances and users we first determine the event type. PEAT distinguishes the event triggers from appliance monitoring into two *viz.*, (i) unambiguous events and (ii) ambiguous events. *Unambiguous events* are those when there is a total certainty that a single occupant is using the appliance. These events occur when there is only one occupant in the household at that time or when a personal appliance is used. By filtering the occupants who are outside the household, the model determines the occupant currently using the appliance. Unambiguous events can help to determine char-

acteristics of occupants such as, which appliances are used by only one occupant at an instance? Which appliances are commonly used by a specific occupant? Furthermore, for an *ambiguous event* there is more than one occupant associated with the usage of an appliance. This is typically the case in multi-occupant households, where cooking, watching TV, and using lights are usually group activities.

In case of an ambiguous event, PEAT first determines the location of the occupant by evaluating the location feature vector. All the occupants whose predicted location is different from the location where the appliance event occurred are discarded. Furthermore, for all occupants who are in the same location as the appliance being used, the system identifies the occupants who are associated with a smartwatch. The activity information from these occupants is evaluated to identify the micro-activities performed. The model determines if the activity performed resembles the activity related to the appliance usage and associates the occupant accordingly. For example, if the activity of an occupant inferred is “opening microwave door" and the appliance event is from microwave, then PEAT assigns the usage of microwave to that occupant. Energy is apportioned with a high certainty to the occupant whose activity matches with that of appliance being used. *Note that, not all users in the household may have a smartwatch/phone.* Hence, there could be an ambiguous event that is not resolved. Most of the energy apportioning systems hitherto, divide evenly the energy consumption if they cannot resolve the ambiguous events. To this end, PEAT utilizes several heuristics that relies on historical data of the occupant's association with the appliances.

- *Number of times used* ($\mathbf{H}_1$) assigns higher association probability to the occupant who has used the appliance more number of times.

- *Recently used* ($\mathbf{H}_2$) assigns higher association probability to the occupant who recently used the appliance.

- *Average usage duration* ($\mathbf{H}_3$) assigns higher association probability to the occupant whose average appliance usage duration matches the current appliance usage duration.

These heuristics are used to determine the percentage of energy that needs to be distributed among the occupants when an ambiguous event is not resolved. For example, if the appliance used is in a Bedroom, H_1 assigns higher probability to the user of that bedroom. Similarly, if the appliance used is Refrigerator, then H_2 assigns higher probability to the occupant who recently used. This is generally the case in the kitchen. If an Occupant-A watches TV for approximately an hour and Occupant-B watches TV generally for 30 m, then the heuristic H_3 assigns higher probability to Occupant-A when the current TV usage exceeds 30 m. Characteristics such as number of times an appliance is used, its average usage duration, are learned over time by analyzing the energy consumption pattern and location information of occupants. Finally, after applying all the heuristics, the event is assigned to a single occupant or group of occupants who are more likely to have used the appliance. If the association probability to one or more occupants has similar values, then the event is assigned to all those occupants and the energy is equally apportioned. The association probabilities and usage characteristics of occupants are stored in Raspberry Pi for deriving per-occupant statistics and to adapt the heuristics over time.

3.2.4. ONLINE EVALUATION

This component evaluates the energy to be apportioned to each occupant in the household in real-time. The evaluation starts when the “OFF" event trigger is obtained from the appliance. The event is then classified to be either ambiguous or unambiguous based on the location information and appliance under consideration. PEAT then evaluates the location and activity information obtained from the user monitoring component. The location accuracy may be inaccurate in some scenarios due to misclassification or the user may not have carried his/her phone. To overcome this PEAT applies a simple location correction mechanism. From the metadata collected, we know the location of each appliance in the household. If there is only one occupant and his location is other than the location of the appliance being used, we then use the appliance location as the corrected location. This corrected location information is then used by the appliance-user modeling component. The association probability derived from the appliance-user modeling is used to apportion the energy among occupants. This information is further sent to all the occupants with individual and shared energy consumption details.

3.3. EVALUATION

Deployment details:

To evaluate PEAT in real-world, the complete system was deployed in two multi-occupant settings *viz.,* student house and a residential household. The *student house* is a two-bedroom apartment with *four* locations *viz.,* Kitchen, Living room, Bedroom 1 and Bedroom 2. All locations apart from the bedrooms are shared by the occupants. The appliances include microwave, refrigerator, grill, coffee machine, laptops and television (TV). Three occupants were present in the student house during our experimentation. The *residential household* contains 12 appliances spread across 5 rooms as described in Chapter 2.3. Two occupants were present in the household during our experimentation.

Ground Truth: To validate results from PEAT, ground truth about the use of an appliance is required. Hence, we deployed NFC tags to collect this information. Each tag is pre-programmed with the appliance name and location. Upon the initialization of an NFC tag, an event is logged into the system with the occupant ID. We use this information only for comparing the results derived from PEAT.

Methods:

Appliance monitoring component was evaluated with both original CO and the proposed modified CO. We evaluated location and activity feature vectors across three online classifiers *viz.,* Decision trees (J48), Naïve Bayesian (NB) and K-Nearest Neighbors (KNN). Finally, we studied the trade-off between energy apportioning accuracy and the number of devices (smartphones/smartwatches). The following methods were employed to derive energy apportioning accuracy:

- M_1: One user with the smartwatch, all other users with the smartphones and with heuristics H_1, H_2 and H_3.
- M_2: No smartwatch, all users with the smartphones and with heuristics H_1, H_2 and H_3.
- M_3: All users with the smartwatches, smartphones and with heuristics H_1, H_2 and H_3.

- M_4: One user with smartwatch, all other users with the smartphones and no heuristics.

- M_5: No smartwatch, all users with the smartphones and no heuristics.

Metrics:

Several accuracy metrics are considered here to evaluate the components of the toolkit. Different metrics considered for appliance monitoring are Fraction of total energy assigned correctly (FTE), Total disaggregation error (T_e), Number of appliances identified correctly (J_a), Number of appliance states identified correctly (J_s), which are described in Chapter 2 Section 3.2.4.

The set of metrics used to evaluate the classifier models obtained for location and activity are:

1. Precision: It is the ratio of number of correctly identified instances over total number of identified instances. Let t_p and f_p indicate the true positives and false positives respectively and precision is defined as,

$$precision = \frac{t_p}{t_p + f_p} \tag{3.2}$$

2. Recall: It is the ratio of number of correctly identified instances over total number of instances. It is given by,

$$recall = \frac{t_p}{t_p + f_n}, \tag{3.3}$$

where f_n represents the number of false negatives.

3. F_1 Score: It is a measure of accuracy and is defined as the harmonic mean of precision and recall.

$$F_1 = 2.\frac{precision \cdot recall}{precision + recall} \tag{3.4}$$

4. Energy apportioning accuracy (E_a): Energy apportioning accuracy is the ratio of estimated energy utilized by an occupant and the actual energy utilized by that occupant. It measures the total percentage of energy correctly apportioned to an occupant and is given by,

$$E_a = \left(\frac{\text{Estimated energy per occupant}}{\text{Actual energy per occupant}}\right) \times 100 \tag{3.5}$$

System architecture:

The system architecture consists of several clients (occupants' devices) communicating with a server (local raspberry PI). *Server-side* includes a Raspberry PI with WiFi connectivity that acts as the local server in each household. Raspberry PI receives the energy consumption data from the smart meter. In our setup we used Plugwise Smile-P1 to retrieve the data from smart meter and send it to the Raspberry PI via WiFi. Raspberry PI runs the proposed ModCO energy disaggregation algorithm to derive fine-grained appliance usage information. Upon detection of appliance ON event, Raspberry PI sends out a push notification (trigger) to all the clients (occupants devices i.e., smartphones/watches) to start scanning (i) WiFi RSSI samples for indoor localization on smartphones and (ii) data collection of inertial sensors on smartwatches of users. Further, when an appliance OFF event is detected, Raspberry PI sends out another push notification to stop the data collection at the client devices. On the *client-side*, an application is developed for smartphones and watches of users. There could be multiple occupants in a household and hence during the initial phase each

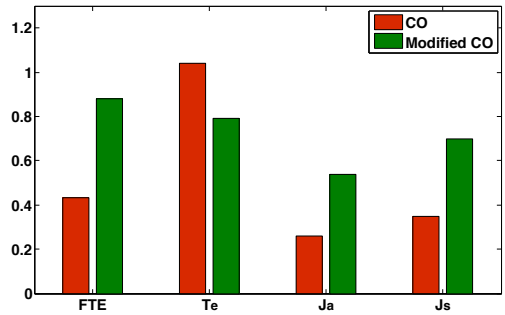


Figure 3.2: Disaggregation performance of CO and modified CO in residential household.

occupant is assigned a unique ID along with their devices. To save energy on the client devices, we do not start data collection until a trigger is received from the local server. Upon reception of the trigger on smartphones, the application starts scanning for WiFi RSSI signals and then sends out a notification to the smartwatch associated for collecting activity related information. Further, when an OFF event trigger is received, the application utilizes an online version of Bayesian classification to derive the room level occupancy and the micro-activity performed.

Further, Raspberry PI receives the inferred location and micro-activity performed for ON-OFF event of each appliance. This information is used to develop an appliance-user modeling and for apportioning energy to individual occupants. The proposed system architecture is distributed, wherein, localization and activity recognition is performed on smartphones/smartwatches, and energy disaggregation and apportioning is performed on the Raspberry PI. Note that all the devices, Plugwise Smile-P1, Raspberry PI and smartphones, are connected to the same access point in the household.

3.4. RESULTS

In this section, we present our experimental results in determining state change of an appliance, room level occupancy and activities performed by the occupants. Furthermore, we show the energy apportioning accuracy across different real-world multi-occupant settings and its trade-off with respect to the number of devices used.

Appliance detection accuracy:

PEAT employs modified CO to determine the state of the appliances in real-time. Accurate energy disaggregation is a critical component for unambiguous energy apportioning. To ensure fair comparison, both original and ModCO utilize the same appliance model from the crowd-sourced database as described in Section 3.2.1. Furthermore, comparison with other NILM algorithms (FHMM [85]) require additional training such as prior probability and state transition matrix. Hence we restrict the comparison of proposed ModCO with the original CO algorithm.

Fig. 3.2 shows the disaggregation performance of CO and modified CO across *four* accuracy metrics in the residential household. We used over 2 months of aggregated energy consumption data of the household. FTE, J_a and J_s can vary between 0 and 1, and T_e can take any non-negative value. ModCO assigns (FTE) more than 85% of the aggregate energy accurately. Furthermore, around 75% of state changes (J_s) are estimated correctly as compared to 35% by original CO. Similarly, when a student house was considered, 90% of

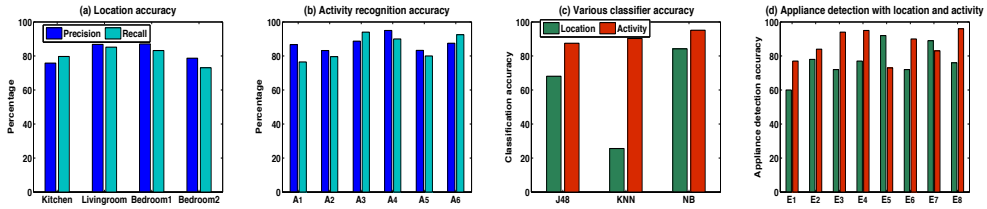


Figure 3.3: Evaluation of user monitoring component in student house.

total energy was accurately assigned and 87% of the states were identified correctly. The increase in FTE and appliance detection in student house compared to residential household is mainly due to reduced number of appliances considered. In both the settings, T_e in modified CO is much lower than original CO, indicating better disaggregation performance.

In general, the improvement in disaggregation accuracy for ModCO is due to the fact that the predicted set of appliances does not vary significantly for consecutive time periods. Our ModCO takes approximately 0.12 s and 0.05 s to determine the state of the appliances in the residential household and student house, respectively.

Indoor location accuracy:

An Android application was installed on smartphones to scan for visible WiFi APs when an event trigger is received. Feature vectors were computed as described in Section 3.2.2. During our experimentation, we considered two approaches for building the classifier model *viz.*, supervised and unsupervised.

Supervised method requires a training phase where RSS values at each location is collected and labeled. During testing phase, each feature vector is evaluated with the classifier model obtained in the training stage to derive the class label (room level location). *Unsupervised method* does not know the class labels *a priori* and learns the label of the location based on occupancy and the appliance metadata. For example, when only one occupant is present in the household and if the appliance trigger was from “Coffee Machine" then the location of the appliance (i.e., Kitchen) can be obtained from metadata information. Consequently, the algorithm learns this label and assigns it to the current location of the occupant. Furthermore, this iterative approach continues until all the class labels are determined. However, this method has several drawbacks such as works when only one user is present in the household and longer delay in developing accurate location models. Recently, several algorithms such as Zee [97] and EchoTag [114] employ crowd-sourced data collection to eliminate the tedious training phase. These approaches could also be used in PEAT. In this chapter, to show the effectiveness of PEAT, we use a standard online classification model.

We employed Naïve Bayesian (NB) classifier model to derive class labels for each new feature vector obtained. Fig. 3.3(a) shows the precision and recall for each location in the student house. High values of precision and recall at each location indicate the good performance of the classifier model. Moreover, F_1 score of 84% was achieved for room level localization using NB. Furthermore, we compared the classification results with two other well-known classifiers *viz.*, J48 Decision trees and KNN with 10-fold stratified cross validation. Fig. 3.3(c) shows that NB performs much better than J48 and KNN, with KNN having the least classification accuracy. Finally, F_1 score of 78% was achieved using NB for room level localization in the residential household. NB does not over-fit the data as compared to J48 and KNN.

Methods	$\mathbf{M}_1$	$\mathbf{M}_2$	$\mathbf{M}_3$	$\mathbf{M}_4$	$\mathbf{M}_5$
Energy apportioning accuracy	92.6	90.7	95.4	87.1	80.6

Table 3.2: Percentage of energy correctly apportioned for different methods in student house.

Appliance ID	$\mathbf{E}_1$	$\mathbf{E}_2$	$\mathbf{E}_3$	$\mathbf{E}_4$	$\mathbf{E}_5$	$\mathbf{E}_6$	$\mathbf{E}_7$	$\mathbf{E}_8$
Apportioning accuracy	66.6	100.0	88.9	88.9	100.0	88.9	80.7	77.8

Table 3.3: Percentage of energy correctly apportioned for each appliance among all the occupants.

Activity detection accuracy

In this chapter we considered *six* micro-activities *viz.*, (i) microwave usage (A_1), (ii) laptop usage (A_2), (iii) refrigerator usage (A_3), (iv) TV usage (A_4), (v) grill usage (A_5) and (vi) coffee machine usage (A_6). When there is an "ON" event from appliance monitoring, activity information is collected from occupants using the smartwatch as described in Section 3.2.2. In the training phase, each activity is labeled and the features in both time and frequency domain are collected. Through exhaustive experiments we found that 6 s of samples at 100 Hz sampling frequency as optimal to detect the activities accurately. Note that a large sampling duration may include additional information, which may not be relevant and having a small sampling duration may not capture the relevant features of an activity. Hence, identifying the right sampling duration is crucial.

Similar to location monitoring we evaluated the activity feature vector across three classifiers (see Fig. 3.3(c)). NB has a classification accuracy of 95% where as J48 and KNN has around 87.5% and 90% respectively. Fig. 3.3(b) shows the precision and recall for each micro-activity performed. It can be seen that the precision of all the activities are higher than 85% and the overall accuracy of identifying the activities is around 95%. Even though the activities A_1 and A_3 are quite similar, the classifier was still able to identify them correctly. This is attributed to the correct identification of the sampling duration.

Furthermore, we conducted several experiments to understand the effectiveness of the location and activity features with respect to appliance detection. Accuracy of detecting an appliance usage with either location or activity features is shown in Fig. 3.3(d). x-axis indicates the various appliances in the student house and y-axis represents the appliance detection accuracy. In general, the activity features can determine the associated appliances more accurately than location features. This is due to the identification of micro-activities using the smartwatches. It can be seen that for some appliances such as Refrigerator and Grill, location features have higher accuracy than activity features. This is attributed to the placement of appliances in different rooms and their distinctive consumption profile. Moreover, for both location and activity information the average accuracy of identifying the associated appliance is around 82%. This information can also be used with ModCO to improve the accuracy of appliance detection.

Energy apportioning accuracy

We considered over 2 months of data to evaluate PEAT in both student house and residential household. All the occupants were equipped with their personal smartphone and smartwatch. Furthermore, the proposed toolkit can also be applied to other shared spaces such as office environments. Note that the level of apportioning required in these spaces

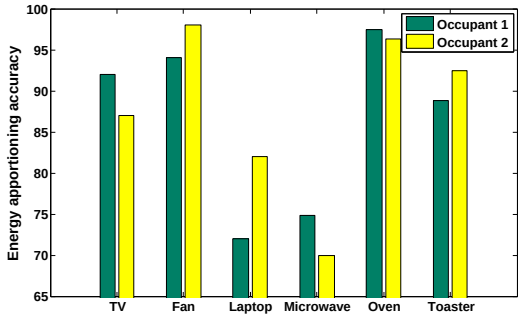


Figure 3.4: Energy apportioned correctly for each occupant in the residential household.

varies. In student house and residential household, PEAT apportions energy of shared and personal appliances. The baseline appliances such as refrigerator, routers, air conditioning and central heating are not apportioned to individual occupant. Moreover, the toolkit can be extended to support apportioning of baseline appliances with additional training and micro-activity recognition (e.g., open refrigerator door and take/keep items).

To study the efficacy of PEAT, we evaluated energy apportioning accuracy for varying number of devices as mentioned in Section 3.3. Table. 3.2 shows the total percentage of energy correctly apportioned for a week in student house with varying number of user devices. When there is one occupant with smartwatch and other occupants have only their smartphones i.e., method M_1 with heuristics, PEAT achieves around 92% apportioning accuracy per-occupant. This accuracy reduces to 87% when no heuristics are considered i.e., M_4. Furthermore, when all the occupants have only their smartphone but no smartwatches, PEAT still achieves 90.7% accuracy with heuristics (M_2) and it is 80% when no heuristics are applied (M_5). Finally, if all the occupants had both smartphone and smartwatch, then the energy apportioning accuracy achieved by PEAT is around 95% (M_3). Clearly, having more user devices increase the apportioning accuracy.

We now illustrate the energy apportioning accuracy for the method M_1 among all occupants on per-appliance basis. Table. 3.3 shows appliance level energy that was correctly apportioned to all the occupants in the student house. It can be seen that when an event was associated with a personal appliance such as Laptops, the apportioning accuracy was 100%. However, with the shared appliances the apportioning accuracy reduces to 80% on average. The average energy apportioning accuracy for all the appliances is close to 90% and per-occupant average apportioning accuracy was around 92% in the student house.

Fig. 3.4 shows the energy apportioned to each occupant across appliances in the residential setting when only smartphones were used (M_2). PEAT achieves on average 87% apportioning accuracy for all occupants when only location information was used in residential household. The apportioning accuracy increases to 92% when all the occupants in the household have their smartphones and smartwatches. Furthermore, for baseline appliance "refrigerator" an apportioning accuracy of 88.9% is obtained for both the occupants.

Discussions:

While PEAT takes the first few steps towards effective user-level energy disaggregation, there are some limitations:

(i) Resolution of apportioning: The level of apportioning may vary depending on the envi-

ronment, for example, in residential settings it may be more useful to apportion top energy consuming appliances than all household appliances. In shared spaces such as student house, shared appliance usage such as TV, washing machine, microwave and oven needs to be apportioned to raise awareness among occupants. In PEAT, users can select appliances for which energy needs to be apportioned to individual occupants during the setup.
(ii) HVAC apportioning: HVAC usage is the primary energy guzzler in households. Hitherto, HVAC consumption was equally shared among all occupants. PEAT with the help of energy disaggregation can identify when HVAC was turned on/off and using location monitoring can determine when and where each occupants are in the household. PEAT attributes the HVAC consumption to the users depending on the time they spent in the household. However, this works only for centralized HVAC systems and currently there is no way to determine individual room HVAC consumption without any additional sensors.
(iii) Shared appliances: PEAT with the help of activity and location monitoring can determine which occupant is using the appliance when multiple occupants are in the same location. However, when an occupant is not carrying his device (smartphone/smartwatch) the event cannot be resolved to an occupant accurately. While we propose three heuristics to apportion energy in such cases, further enhancement of appliance-user modeling is required for accurate apportioning.
(iv) Moveable appliances: PEAT requires location of appliances known *a priori*. Hence it cannot accurately apportion energy usage of moveable appliances such as hair dryer, when there are multiple occupants present in the room.
(v) Localization and activity monitoring algorithms: PEAT achieves 90% apportioning accuracy by using standard classification algorithms. We believe PEAT can be more effective and robust by incorporating other crowd-sourced training free algorithms [97, 114].

3.5. CONCLUSIONS

In this chapter, we proposed a novel Personalized Energy Apportioning Toolkit (PEAT) to accurately apportion energy amongst occupants in shared spaces. Inferring energy footprints of occupants with minimal user intervention and no additional sensor deployment is a challenging task. We showed that PEAT can accurately determine *which* occupant performed *what* activity and *where* in the household. Furthermore, it combines online techniques with minimal training for accurate energy apportioning. We proposed several accuracy metrics to study the performance of each component of PEAT. We specifically deployed the system in two multi-occupant settings – *viz.,* a student house and a residential household – to evaluate PEAT in real-world settings. With only the location information an energy apportioning accuracy of 87% and 92% was achieved for all the occupants in the residential household and student house, respectively. The apportioning accuracy increases to 92% and 95% when both location and activity information are available in the residential household and student house. We demonstrated that PEAT is highly scalable and privacy preserving since privacy-sensitive data of occupants are stored and processed locally.

The techniques presented in this part provide real-time energy-consumption information to occupants to understand their energy-usage behavior and encourage them to optimize their energy usage. LocED and PEAT take the first steps towards providing personalized energy services at both appliance and user level, respectively. In the next part we describe the usage of fine-grained energy information for various demand-regulation programs such as demand shifting and reduction.

II

Demand Regulation in Smart Homes and Buildings

4

DEMAND SHIFTING

WITH the rapid advancements in embedded systems and wireless technologies, numerous appliances such as vacuum cleaners, washing machines, ovens and refrigerators are becoming more intelligent and can be controlled remotely in a smart home. In Smart Grids (SG), utilities are allowed to dynamically adjust the electricity prices in order to control demand. The real-time pricing information communicated to the smart homes can be used to control and adjust the demands at the customer premises.

Demand regulation (DR) is a key technique that can control and influence energy demand at the consumer-end to reduce the overall peak demand, reshaping demand profiles and increasing the robustness of SG [90]. Several demand-regulation techniques are proposed in the literature for load shifting [35], peak clipping and valley filling [76]. In this chapter, we limit the DR to load shifting. *Load shifting*, involves shifting loads from peak to off-peak hours, without significantly influencing the average load over time. Load shifting ensures the energy consumed by the household at any time period does not overload the grid by altering the demand pattern of the household.

Each household is assumed to be equipped with an *information system* that collects real-time demand measurements from smart meters and also controls energy consumption. Existing energy management systems (EMS)/information systems are mainly designed to improve energy efficiency and comfort, i.e., turning off appliances when not in use, changing HVAC/air-condition set points to minimize energy consumption [77, 106], etc. Recent EMS aim to reduce electricity cost by scheduling the demand of a household based on the electricity prices (real-time or day-ahead). These scheduling algorithms utilize either (i) fine-grained energy-consumption information from the appliances or (ii) aggregate energy consumption of the household for load shifting. In the case of *fine-grained information*, energy consumption of each appliance in the household is analyzed for deriving appliance-level schedules. These approaches require detailed user and appliance information to schedule loads effectively. In case of *aggregate consumption*, the scheduler aims to determine the energy that needs to be shifted from the total consumption of the household at a given time period. This approach requires consumers to decide which appliance needs to be turned-on/off to match the energy that needs to be shifted.

There still exists several challenges hindering the applicability of load shifting in residential households. We enlist some of the important ones here: (i) Most of the approaches presented [35, 62, 116] require detailed user and appliance-level information to schedule loads effectively. This either requires additional deployment of sensors or significant consumer involvement. (ii) Approaches based on aggregate energy consumption, either select a demand pattern based on historic data or require the consumer to shift energy. Quite often, consumers have no real knowledge on appliance-level energy information. (iii) Most demand-scheduling algorithms do not consider the heterogeneity of appliances in the household, flexibility in appliance usage, and appliance dependencies during scheduling. Hence the resultant schedule is either infeasible or the user comfort is severely hampered. (iv) Load scheduling or shifting is quite often performed in a centralized manner where energy consumption information is sent to the utilities or an aggregator to determine the optimal schedule. (v) Lastly, the scheduling algorithms are generally evaluated using simulations or numerical analysis, which may not reflect the reality very well [32, 36, 62, 123]. Furthermore, the applicability of scheduling algorithms across different households has never been considered in the prevalent research [105, 115].

To overcome the above limitations, we propose a decentralized demand-scheduling scheme. The proposed demand-scheduling algorithm can, (i) determine appliance-level information and user preferences for appliance usage, using only aggregated energy consumption from the smart meters and (ii) propose a demand-scheduling algorithm that minimizes the user discomfort and electricity cost based on day-ahead hourly pricing. We derive fine-grained appliance information and user preferences from the aggregated energy information using a low-complexity energy-disaggregation algorithm (ModCO described in Chapter 2). The proposed demand-scheduling algorithm is evaluated with our real-world deployment Dutch Residential Energy Dataset (DRED) in the Netherlands [2]. Furthermore, we show the applicability of the scheduler on another household in the USA using the open dataset REDD [72]. The scheduler implementation is made publicly available for the community to support additional analysis [2].

Contributions. The main contributions are:

- We propose a novel decentralized demand-scheduling algorithm that minimizes user discomfort and electricity cost of a household.
- We describe three coefficients to analyze user preferences and appliance usage patterns using historic aggregated energy consumption.
- We provide a detailed empirical evaluation of the proposed algorithm using real-world deployment and publicly open datasets.

4.1. RELATED WORK

Numerous DR programs [32, 36, 62, 105, 123] have been proposed to motivate changes in the consumers's power consumption with the objective to either (i) minimize the electricity cost, (ii) maximize the social welfare, (iii) minimize the aggregated power consumption, or (iv) any combination of the above [45].

Table. 4.1 provides a concise overview of the state-of-the-art approaches against the proposed scheme. The columns indicate whether the scheduling algorithm is centralized or decentralized, if consumer preferences are considered or not, whether the evaluation was

Work	Method	Consumer preference	Study type	Schedule level	implem -entation
[62]	Centralized	Partial	Numerical	Appliance	No
[123]	Decentralized	No	Simulation	Aggregate	No
[35]	Centralized	Partial	Simulation	Aggregate	No
[32]	Centralized	No	Simulation	Appliance	Partial
[36]	Centralized	Partial	Simulation	Appliance	No
Proposed	Decentralized	Yes	Data-driven	Appliance	Yes

Table 4.1: Comparison of state-of-the-art approaches.

based on simulation or data-driven, if the scheduling algorithm is for appliance or aggregated level, and if the scheduling algorithm can be implemented on an embedded system such as a Raspberry Pi.

In [62] an optimized day-ahead pricing scheme is proposed by considering the flexibility of appliance scheduling. A cost-minimization problem is formulated to reduce electricity cost. However this approach does not consider consumer preferences and authors show only numerical analysis of the proposed scheduling algorithm. A genetic algorithm to derive optimal power schedule of a household is proposed in [123]. The genetic algorithm runs at the household to minimize electricity cost and to reduce the delay in usage of appliances. Similar to [123], Chen et al. [35] propose a task-scheduling algorithm that considers per-appliance delay and also long-term average delay to minimize the electricity cost. Contrary to the above approaches in this chapter, we not only consider delay in appliance usage but we also consider the flexibility, appliance dependencies and consumer preferences to schedule the appliance usage. In [32] an integer linear programming technique for online load scheduling is proposed to minimize energy cost. In [36], a scheduling technique is proposed by modeling energy consumption and user preferences as a stochastic variable. Appliance-level schedules derived are evaluated using simulation results. The state-of-the-art techniques do not completely capture consumer preferences and appliance usage patterns. Furthermore, these techniques cannot be applied across households.

4.2. SYSTEM MODEL

Each household is assumed to have an information system (i.e., Raspberry PI or Arduino) connected with the smart meter to balance energy demand by applying demand regulation techniques. Fig. 4.1 shows the system model of the proposed decentralized demand scheduling system in smart homes.

The energy utilities send the day-ahead hourly pricing to all its consumer base. The information system at the household then derives day-ahead schedules to minimize electricity cost. To derive day-ahead schedules, the aggregated energy demand data from the smart meter is given to the energy disaggregation block. Energy disaggregation block employs a Modified Combinatorial Optimization (ModCO) algorithm to infer per-appliance energy consumption information (more details on ModCO can be found in Chapter 2.2. This information is also used to derive consumer preferences such as, which appliances are used and its duration, usage patterns in weekdays and weekends, etc. The appliance level energy information along with consumer preferences are used by the demand scheduler

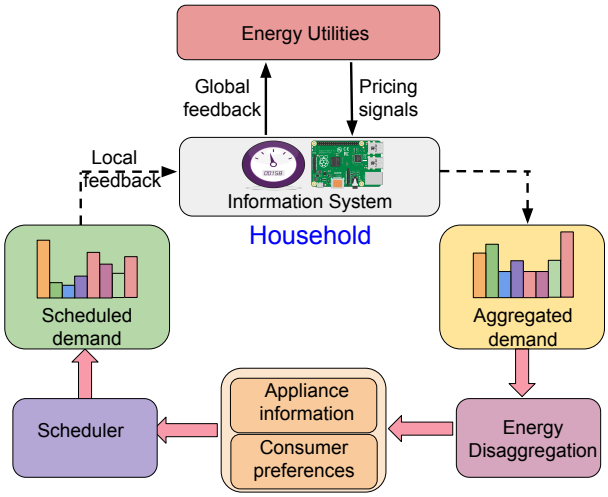


Figure 4.1: System model of decentralized demand scheduling.

to derive day-ahead schedules. The demand scheduler utilizes several coefficients to minimize the electricity cost and consumer discomfort. The proposed day-ahead schedule is then communicated to the household/occupants via the information system. This we call *local feedback*, which can be used to understand the effectiveness of the proposed schedule or how the occupants are adapting. Furthermore, the information system communicates the proposed schedule to the energy utilities, we call this *global feedback*. The global feedback allows utility to plan the energy purchase and also to balance energy at a larger scale like neighborhood and township.

In this chapter, we distinguish the appliances in smart homes as *non-schedulable* and *schedulable*. The former represents appliances that require fixed energy requirement and are not subjected to scheduling decisions. These appliances include television, refrigerators, modems, etc. Schedulable appliances allow appliance usage to be shifted in time and has a direct relation with consumer preferences and behavior. These appliances include dishwashers, washing machines and clothes dryers. The distinction between the loads can be automatically done by analyzing the appliance usage patterns.

4.3. DAY-AHEAD DEMAND SCHEDULING ALGORITHM

We now describe an algorithm that generates day-ahead demand schedule for a household, which minimizes electricity cost and consumer discomfort. Discomfort refers to the inconvenience experienced by the consumers during load shifting. The derived schedule is communicated to the occupants via information system to execute it the next day. Our algorithm is agnostic to time granularity, i.e., it can be applied for an entire day, during peak time periods, hourly, etc.

We formulate a cost minimization problem at the consumer-end by effectively scheduling loads based on day-ahead hourly pricing. Appliance usage patterns and consumer preferences are derived from the disaggregated energy data. The hypothesis considered here is that the proposed day-ahead schedule should resemble to the historic energy consumption

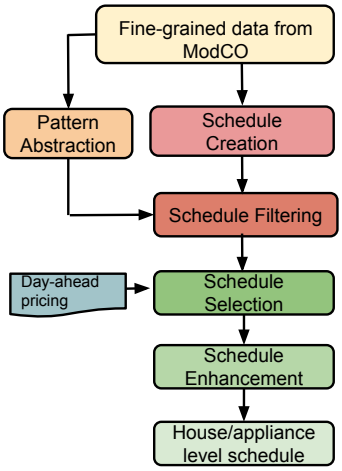


Figure 4.2: Overview of schedule generation algorithm.

pattern of the consumer and it has minimal discomfort since the consumers have executed them previously. However, consumer preferences may change over time. Hence, our algorithm creates schedule not only based on historic demand patterns of consumers, but also by determining several coefficients that define consumer preferences and appliance usage patterns.

Fig. 4.2 shows an overview of the schedule generation algorithm. Our algorithm has five modules *viz.*, schedule creation, pattern abstraction, schedule filtering, schedule selection and schedule enhancement.

Schedule creation:

The first step is to find the set of possible demand schedules of consumers from their historic demand data. These schedules represent the energy usage behavior of consumers in the past. We then group these schedules at different granularities i.e., either weekdays or weekends, day of the week, etc. The past schedules generated are feasible and have minimal impact on consumers daily routine since they have already executed them at some point in the past.

A feasible schedule could be chosen from the past schedules that coincides with the type of day in consideration. For example, a schedule for Saturday may choose only schedules of past Saturdays or weekends. In such a way, the final proposed schedule retains a greater resemblance to what the consumer typically does on that day. It may be possible that some of these schedules do not match user preferences either due to huge variation in demand profile on that day or due to arrival of guests in the household, etc. Hence it is necessary to determine representative schedules that accurately depict consumer preferences from the past schedules.

Pattern abstraction:

We propose three energy usage coefficients to analyze appliance usage patterns and consumer preferences.

(i) *Flexibility coefficient* represents the average usage duration of an appliance in each hour of the day. This indicates the time periods when an appliance was used previously and how much time it was used. Fig. 4.3 shows flexibility coefficients heatmap of appliances

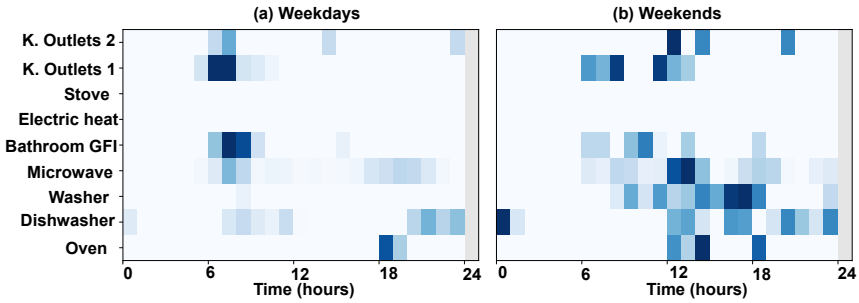


Figure 4.3: Usage patterns of appliances in REDD.

during weekdays and weekends of a household from REDD dataset [72]. It can be seen that, appliance usage is high in mornings (7-10AM) and evenings (6-8PM) on weekdays. However, during the weekends the appliance usage is spread across the day. Flexibility coefficient indicates the most preferable time period of appliance usage by the users.

(ii) *Sensitivity coefficient* indicates the preferred time delay in usage of an appliance by consumers. Some appliances can tolerate longer delays compared to others. For example, the coffee machine might allow shorter delays than the washing machine as the user always prepares coffee within a specific (and shorter) time period.

(iii) *Dependency coefficient* indicates the appliance dependencies, associations and usage sequence. In general, the occupants have a daily routine making it possible to use an appliance in a sequence. For example, TV is always associated with a home theater.

Schedule filtering:

Schedule filtering employs the energy usage coefficients described previously to filter and select schedules that most accurately represent consumer preferences. We select the subset of schedules that adhere to the derived usage patterns and discard schedules that occurred only a few times or that are not representative of a typical day. Fig. 4.4 shows the difference between a representative and non-representative schedule based on per-appliance usage time. It can be seen that, a representative schedule has most of the appliances adhering to the appliance usage time periods (flexibility) of a typical day. Moreover, in a non-representative schedule only few appliances adhere to the average usage duration. The filtering of schedules is done in combination with all the three energy usage coefficients.

The schedule filtering is based on individual appliances and to derive a representative schedule for a household, a minimum number of appliances need to adhere to the requirements derived. This setting is adjustable by the consumer or the utility or after negotiation. It represents the harshness in schedule filtering and can be used to identify the discomfort. For example, a requirement of low number of appliances to adhere to the coefficients may result in selection of a schedule not matching the user preferences, leading to high discomfort. The bounds on the coefficient values are derived based on consumer preferences. Finally, the filtered schedules are the representative schedules for that household.

Schedule selection:

From the set of representative schedules derived, we find the schedule that minimizes the electricity cost. Day-ahead hourly pricing information from the utilities[1] is obtained

[1] Day-ahead hourly prices: `http://www.powersmartpricing.org/pricing-table/`

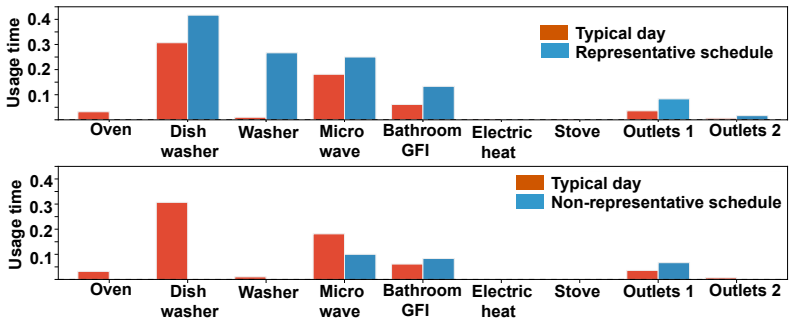


Figure 4.4: Appliance time usage duration in REDD for a representative and non-representative schedule..

to identify the schedule that results in minimal electricity cost. The scheduler selects the schedule with least electricity cost by solving the following cost minimization problem,

$$\begin{aligned} \text{minimize} \quad & \sum_{i=1}^{N} \sum_{t=1}^{24} C_t \, D_t^{(i)}, \\ \text{subject to} \quad & 0 \le D_t \le D_{max}, \forall t, \end{aligned} \tag{4.1}$$

where N is the total number of representative schedules, C_t is the hourly electricity cost and D_t is the hourly energy demand of a representative schedule i. D_{max} represents the maximum hourly energy demand of the household.

Schedule enhancement:

Finally, we try to enhance the cost-optimal schedule derived previously. Enhancements are typically appliance load shifting based on the flexibility and sensitivity coefficients, to further reduce the cost and discomfort associated. Hence the optimization problem in (4.1) can be re-written as,

$$\begin{aligned} \text{minimize} \quad & \sum_{t=1}^{24} C_t \, D_t \\ \text{subject to} \quad & 0 \le D_t \le D_{max}, \forall t, \\ & l_f^a \le f(d_t^a) \le u_f^a, \;\; s(d_t^a) \in (l_s^a, u_s^a), \forall a \in A, \end{aligned} \tag{4.2}$$

where C_t is the hourly cost, D_t is the cost-optimal energy demand, d_t^a is the appliance energy demand, f^a and s^a are the flexibility and sensitivity coefficients for each appliance, $a \in A$ the set of appliances, and l_f^a, u_f^a, l_s^a, u_s^a are the corresponding lower and upper bounds of flexibility and sensitivity coefficients.

We propose an iterative method to solve (4.2) where each appliance usage is either retained at the same time period (if cost is lower) or shifted within the flexibility and sensitivity range derived using energy data. As mentioned previously, the former indicates the average usage time period of an appliance in a hour and the latter indicates the time delay the appliance can tolerate. The iterative method generates a sequence of improving approximate solutions that adheres to these coefficients. Furthermore, the scheduler needs to ensure, (i) an appliance usage event should not be subdivided into smaller events to avoid expensive hours and (ii) an appliance event duration should not be altered i.e., neither stretching nor shrinking of an event is allowed. Our algorithm ensures the above conditions are met and shifts the appliance usage accordingly.

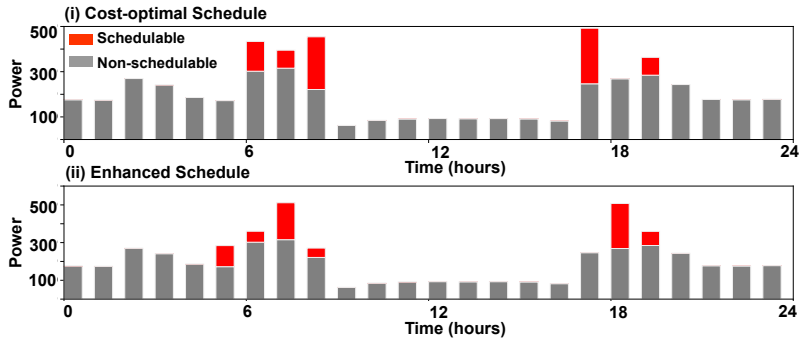


Figure 4.5: Cost-optimal and enhanced schedule in REDD.

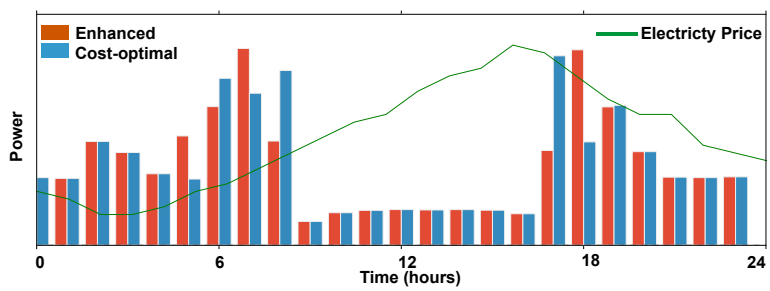


Figure 4.6: Comparison of cost-optimal and enhanced schedule in REDD.

Since, the iterative method is applied only on the cost-optimal schedule the computational complexity associated in load shifting is minimal. Fig. 4.5 shows the demand shifting of schedulable loads for cost-optimal and enhanced schedule. It can be seen that from the cost-optimal schedule appliance usage are shifted to obtain the enhanced schedule. The shifting of appliance usage is based on the electricity price and consumer preferences derived using the energy usage coefficients. The comparison of cost-optimal and the enhanced schedule with day-ahead electricity price is shown in Fig. 4.6.

Furthermore, the proposed scheduler can incorporate renewable energy sources such as solar and wind to balance the energy demand and supply. In this case, D_t in (4.1) and (4.2) can be replaced with $\hat{D}_t$, where $\hat{D}_t$=D_t - D_{RES}. Here, D_t is the demand required by the household, D_{RES} is the demand generated from renewables such as solar and wind, and $\hat{D}_t$ is the energy demand borrowed/requested from the grid.

4.4. RESULTS

We provide performance evaluation of the proposed scheduler across multiple datasets. We first describe the datasets employed for empirical evaluation. We then provide detailed results on energy disaggregation and efficacy of the scheduler.

4.4.1. DATASETS

The energy disaggregation and demand scheduler was evaluated using our deployment Dutch Residential Energy Dataset (DRED) in the Netherlands [2]. Furthermore, the pro-

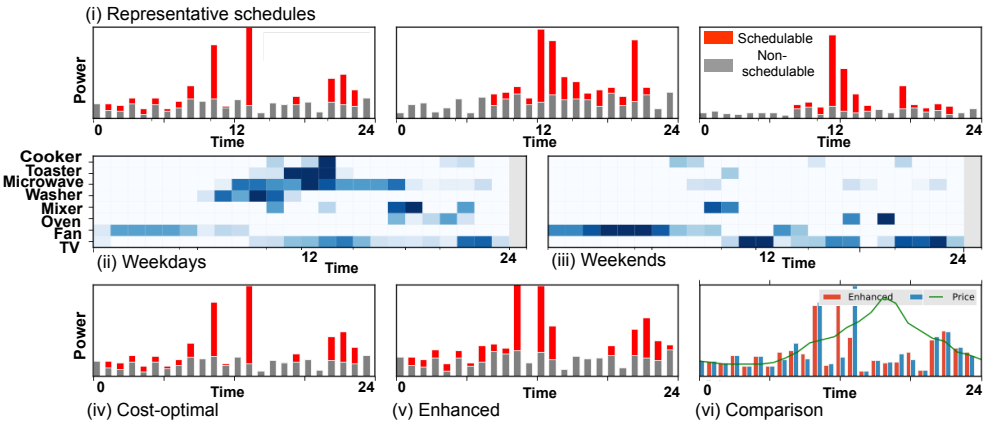


Figure 4.7: Optimal schedule for a day in our DRED dataset.

posed models were also evaluated with a publicly available energy dataset called REDD (Reference Energy Disaggregation Dataset) [72].

(i) DRED: Our deployment consists of several sensors measuring power, occupancy and activities of occupants. The sensors were carefully installed to avoid any inconvenience to the occupants. We collected data at both aggregated and appliance level using smart meters such as plugwise sensors[2] at 1 Hz sampling frequency for over 6 months. The dataset is made public and more details about the deployment can be found in [2].

(ii) REDD: It is one of the first publicly available dataset with both appliance and aggregated energy consumption data. The dataset includes data from 6 households in the USA. Each household has more than 15 appliances and the data was collected at 1 Hz sampling frequency. In our evaluation, we use House-1 data and more details can be found in [72].

4.4.2. DEMAND SCHEDULING

The accuracy of energy disaggregation obtained from ModCO has been previously described in Chapter 2.5 and Chapter 3.4. We now describe the schedule derived using the proposed scheduler in DRED dataset.

We implemented our algorithm described in Section. 4.3 on a Raspberry Pi to determine day-ahead appliance schedule. Fig. 4.7 shows the results from each step of the scheduler in DRED dataset. From all possible schedules, Fig. 4.7(i) shows the filtered representative schedules with schedulable and non-schedulable loads. The grey color indicates the non-schedulable loads such as refrigerator, modem, etc., and the red color shows the schedulable loads washing machine, dishwasher, etc. Fig. 4.7(ii), (iii) shows the appliance usage pattern in weekdays and weekends derived from disaggregated data.

Fig. 4.7(iv) show the cost effective schedule obtained based on the day-ahead pricing using (4.1). The cost-effective schedule shows the schedule executed previously by the household. The scheduler algorithm adapts the cost-effective schedule iteratively to further minimize user discomfort based on the energy usage coefficients proposed. Fig. 4.7(v) shows the derived enhanced schedule using flexibility and sensitivity coefficients. Finally, Fig. 4.7(vi)

[2]Plugwise energy monitoring: `https://www.plugwise.com/smile-p1`

shows the enhanced and cost-optimal schedule along with the day-ahead price. On this particular day in DRED, around 70% of schedulable load was shifted to achieve minimal cost and discomfort. Fig. 4.6 shows the comparison of enhanced and cost-optimal schedules for a household in REDD dataset. On the average monthly electricity cost reduction of 25% and 30% can be seen in DRED and REDD households using the proposed scheduler. The proposed scheduler can be adapted to incorporate renewable energy sources and battery storage. Furthermore, since all the data is stored and processed locally the proposed decentralized demand scheduler is highly scalable. Moreover, the information system at each household can negotiate in a distributed fashion to further minimize the total aggregate load on the grid.

4.5. CONCLUSIONS

In this chapter, we took the first step towards utilizing the fine-grained consumption data (described in Part I) to develop personalized energy services. We presented a decentralized algorithm to derive optimal day-ahead schedules considering consumer preferences and appliance-usage patterns. Our algorithm derives schedules that minimizes the electricity cost and also associated consumer discomfort at the same time. The proposed algorithm was empirically evaluated across multiple datasets such as DRED and REDD. Cost savings of up to 25% and 30% can be achieved in DRED and REDD datasets. Indeed this is the first time actual energy consumption datasets are used to evaluate load shifting at the consumer side. The proposed day-ahead algorithm can also be applied to other variations of electricity pricing such as real-time pricing, critical-peak pricing and time-of-use pricing.

5

DEMAND REDUCTION

IN Chapter 4 we described how to shift energy usage such that user discomfort and electricity cost is minimized. While the scheduling algorithm proposed ensures the energy is utilized in cost-optimal manner it does not reduce the total energy consumption. In this chapter we look at services that provides recommendations to occupants to reduce overall energy consumption.

HVAC (heating, ventilation, and air conditioning) and artificial-lighting systems account for about 25-40% of electricity usage in residential and commercial buildings [19]. Thus efficient usage of the HVAC and lighting is one major step towards reducing energy consumption. Automatic control of HVAC and artificial lights has been one of the popular methods for achieving energy efficiency in buildings. The current systems use fixed set-point controls, which are decided based on a conservative approach. Further, the lighting systems require additional sensor deployment to cope with the continuous intensity fluctuation of natural light.

With many occupants in smart buildings, there is a tradeoff in achieving the preferred comfort levels of users and yet achieving energy efficiency. In this chapter, we present the **i**ndoor **L**ight and **T**emperature **C**ontroler (**iLTC**), which is a smart system that achieves a fine balance between energy efficiency and user comfort. A brief overview of the system is shown in Fig. 5.1. Unlike traditional building energy management systems (BEMS), iLTC employs a room-level controller to decide on an energy optimal operating set-point for the actuators while trying to make all the co-occupants feel comfortable with respect to their preferences. This could be a simple addendum to the main controller of the HVAC in the building. Moreover, the feeling of thermal and lighting comfort is not a single temperature or light intensity value for a person, but a range of values within which a user can feel "equally" comfortable. Thus, iLTC needs to learn in detail about the thermal and visual comfort preferences of each individual.

Implementing iLTC is highly challenging for the following reasons: *First*, an operating set-point for the actuators is a mere number. It is not easy to correlate comfort levels of humans with a certain light intensity and temperature value. A tangible scale of comfort levels needs to be mapped onto the numbers that can decide the set-points. *Second*, thermal comfort varies significantly from person to person. Complete information about the

comfortable temperature range of each co-occupant is required to decide on a common temperature set-point while keeping HVAC energy consumption as low as possible. *Third*, lighting comfort also varies significantly from person to person [50]. Further, light intensity also varies significantly at various locations inside a room. Thus, we need a mechanism to identify natural light intensity at a desired location (e.g., work-desk of a user). An Android application was developed to collect preferences of users. We collected user preference data from 21 participants housed in different rooms to create individual temperature and lighting comfort functions. A detailed evaluation of iLTC is performed based on these comfort functions to measure energy savings. Furthermore, the proposed iLTC system was tested and evaluated by many users.

Contributions. The main contributions of this chapter are:

- We develop a layered design for iLTC that can offer room-level control for lighting and HVAC systems.
- Our system employs a non-intrusive mechanism to derive comfort preferences with minimal user intervention and training. We provide comprehensive mapping functions for a person, which can indicate comfort level for any given light and temperature value.
- We estimate the natural light intensity at the work-desk of users by utilizing light sensor available in their smartphones. We derive a relationship between the light intensity measured by a smartphone sensor and the outdoor light intensity measured with a single sensor. This eliminates the huge cost of deployment of additional sensors and their management.
- We propose an algorithm to determine the most energy-optimal operating set-point for HVAC and lighting systems while making all the co-occupants comfortable.

5.1. RELATED WORK

A significant amount of energy is wasted by the HVAC and artificial lighting systems in a building due to their inefficient usage. Thus, a large body of current works focuses on the efficient usage of these systems from various aspects. Simple solutions proposed to save energy elicit turning off the actuators automatically when there are no occupants [50, 70]. The occupancy is detected using some sensor-based mechanism. ThermoCoach [92] provides a personalized thermostat recommendation exploiting occupancy patterns. However, the technique used cannot be applied for shared spaces where there are multiple occupants.

Another set of work focuses on the reduction in energy consumption and stable operation of the HVAC systems. While some of the research efforts have been to optimize operational efficiency of the HVAC for a given set-point temperature [66, 73], a significant number of works also emphasize on selecting a suitable set-point temperature for the HVAC in order to fulfill thermal comfort of the occupants. There are two main methods for determining thermal comfort, the heat-balanced way and the adaptive approach. The predicted mean value (PMV) approach is a heat-balanced method and depends on six parameters: metabolic rate, clothing insulation, air temperature, radiant temperature, air speed and humidity. Most of these parameters are difficult to obtain from typical sensors in BEMS and hence PMV proves to be impractical. The adaptive approach focuses on the adaptation

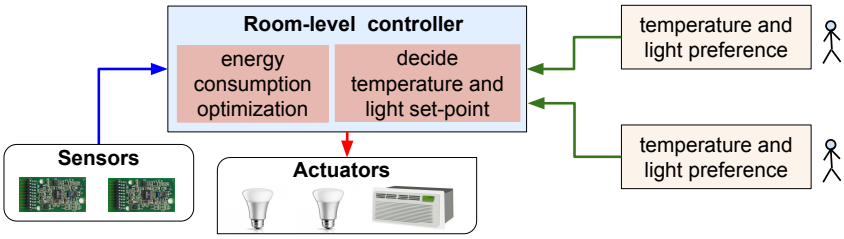


Figure 5.1: iLTC overview.

of human psychological and physiological behavior [41]. Rather than using the traditional PMV model, which assigns a static comfort level to a user, recent studies have shown that participatory-based approaches can be used to optimize user thermal comfort and consequently reducing energy costs [51, 74, 86].

Participatory approaches allow occupants to give feedback based on their comfort level. From the feedback a consensus about a common comfort value can be derived. One major challenge for such a system is that the set-point is resilient to outliers and thus a mechanism is required to cope with outliers as, for example, proposed by Zhang *et. al* [78]. Another challenge is to find the balance between intrusiveness and user involvement in order for the occupants to maintain their incentive to participate. Most of the existing works require additional sensor deployment or detailed user information (preferences, demographics, etc.) and generate a fixed optimal set-point for each occupant given a room, which provides limited flexibility to decide a common set-point temperature. Erickson *et. al* [44] have used a participatory sensing approach to customize HVAC settings. However, their approach does not consider shared spaces with multiple co-occupants. Moreover, it requires a significant amount of user participation to learn about their comfort preferences.

With respect to reducing energy consumption by the lighting systems, the basic idea is to use artificial lights only when it is necessary. To do this light intensity inside a room need to be known. As the natural-light level can change across days without any fixed pattern and also the light intensity varies at different locations within a room, measurements need to be done continuously at every location of interest. Thus many researchers have deployed a number of sensor nodes to monitor the fluctuation in natural light level [27, 61, 89]. However these approaches are intrusive and cumbersome.

5.2. SYSTEM MODEL

A brief overview of iLTC is shown in Fig. 5.1. A room-level controller sets an energy optimal operating point for the lighting and HVAC systems considering the comfort preferences of all the occupants.

There are two major building blocks in iLTC – 'user daemon' and 'room daemon'. A user daemon is associated with each user and it is hosted on his/her smartphone. The proposed system employed a smart-phone based *App* to learn individual temperature and lighting comfort levels for various temperature and light intensity values. The data collected through the *App* is utilized to create comprehensive comfort function. On the other hand, any activity associated with a room is handled by the room daemon. For each room in a building, a separate instance of room daemon is created for their independent operations.

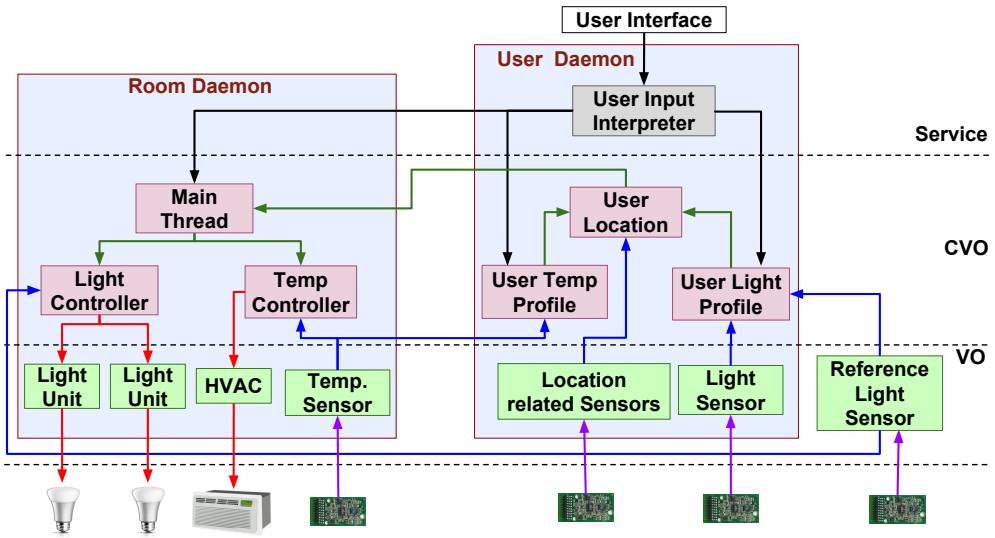


Figure 5.2: Layered design of iLTC.

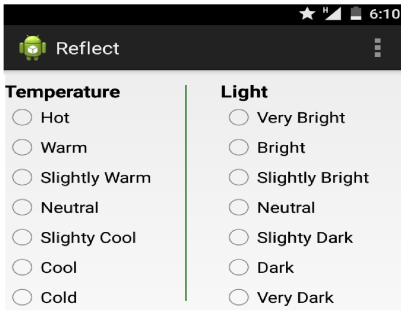
These daemons can be hosted on a centralized server at the building or at a room-level embedded device. The virtual representations of the actuators are hosted by the associated room daemon. Additionally, a temperature and humidity sensor is required for every room to get thermal feedback. As the modern HVAC systems can maintain a near homogeneous temperature within a room, it eliminates various thermal zones within a room. Thus, a single temperature sensor for a room is sufficient.

The proposed system uses artificial lights only when the received natural light at the work-desk fall below the comfort level. The natural light intensity at various positions inside a room varies significantly, and it keeps on changing. Thus, a trivial solution is to deploy light sensor at every work-desks. We use a smart technique to avoid deployment of multiple light sensors inside a room. Rather we use a single light sensor for the whole building, and the effective natural light intensity at each user desk is estimated using this solitary sensor data. This reference light sensor is used by all users and room daemons, and is not part of either of the user and room daemons. Thus, the associated virtual representation is hosted at a centralized location, which can be accessed by all stakeholders.

5.3. USER DAEMON

The user daemon obtains the preferences through the user interface and builds a comprehensive comfort function for thermal and visual preferences. In this section, we describe how individual user preferences are obtained and modeled. Furthermore, we also discuss how to estimate natural light intensity at a user's work-desk using only a single reference light sensor.

A user daemon hosts three modules – 'light profile', 'temperature profile', and 'location'. While the light and temperature modules learn preferences of users during the training period, the location module identifies whether the user is inside a room or not. This binary classification of user presence is performed using WiFi based indoor localization. We uti-

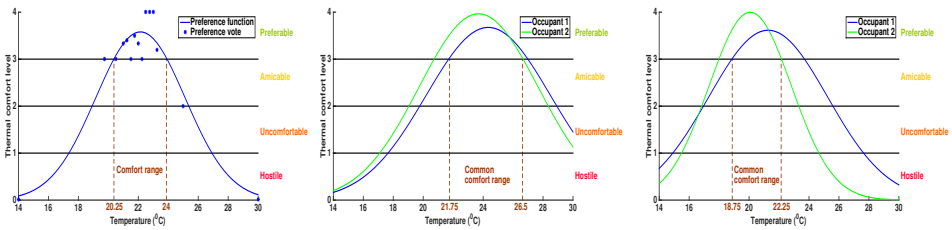


(a) Application screenshot.

comfort indicator	ASHRAE value	iLTC value
hot/very bright	+3	1
warm/bright	+2	2
slightly warm/bright	+1	3
neutral	0	4
slightly cool/dark	-1	3
cool/dark	-2	2
cold/very dark	-3	1

(b) Comfort levels.

Figure 5.3: Screenshot of the App for collecting user preferences and their mapping into a numeric scale.



(a) User comfort function (b) Comfort range: room 1 (c) Comfort range: room 2

Figure 5.4: Comfort function of an individual based on preference voting, and common comfort range of a room based on individual comfort functions of all the co-occupants.

lized existing WiFi access points deployed in the building along with smartphones of occupants. More details on the localization approach is described in Chapter 2.2.

5.3.1. INDIVIDUAL USER PROFILING

The core of iLTC is to build individual comfort profiles for temperature and lighting. For a person, maximum thermal comfort is not a single temperature value rather a range of temperature values. Similarly, visual comfort also spans over a range of light intensities. Most users cannot easily correlate their comfort levels with temperature or light intensity values. Even if they do, there is a chance of significant deviations. To this end, iLTC utilizes a smartphone *App* to learn the preferences, coupled with precise measurements from corresponding sensors. When new users enter the system, we consider a conservative approach with respect to her preferences. Initially, we assign preference values derived from survey, which are then personalized based on the preferences data collected over time.

Collection of user preferences: Thermal comfort of a person for a particular temperature cannot be determined without the feedback from the person. We use both explicit and implicit way of learning the preferences (feedback on comfort levels) – explicitly by asking the users to indicate their comfort levels and implicitly (later while iLTC is in operation) when the user overrides the controller settings. This allows dynamic adaptation of comfort preferences. Furthermore, if the new setting is significantly different, then preferences are adapted again by collecting additional voting data from the user. Thus, by capturing the changes in user preferences overtime, iLTC eliminates any outliers.

A screenshot of the App, which is used to collect feedback, is shown in Fig. 5.3a. To indicate comfort a seven-point scale is used as suggested by the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE). However, we convert the traditional numeric scale of [-3,+3] to [1,4] (see Fig. 5.3b), because indicator level 'neutral' is given the highest preference, whereas indicator level 'cold' and 'hot' are given the lowest preference. To learn the visual comfort a similar scaling is also adapted as indicated in Fig. 5.3. The data collected from the App is used to model the comfort preferences for temperature and light. Note that the explicit data collection is done only in the beginning when a user becomes part of iLTC.

Our goal is to create comprehensive comfort functions that can indicate the comfort level of the person for any given temperature or light intensity. To build such a function, ideally, we should have comfort level indicator for each possible value, which is not a feasible option. Thus we collected a few comfort indicators and then we try to model them. The data collection was conducted from 21 users with 5 different ethnic background, age varying from 24 to 51 years. The data collection is done over several weeks with a dedicated sensor node deployed in each of the rooms of the occupants (participants). Next, we explain the mapping from these measurements to comfort levels.

Modeling user profile – mapping room temperature and luminance to comfort levels: Whenever a person indicates her comfort level, it is registered with its corresponding sensor values. During function creation, we cluster these comfort indicators in multiple equal sized bins. The bin size indicates a small range of sensor values for which user comfort remains unchanged. Different bin sizes were empirically evaluated across all the participants to determine the optimal bin size. More details on selecting appropriate bin size is discussed in Section 5.5. After analyzing user preference data collected from multiple users, we notice that thermal comfort function can be represented using a Gaussian function (eq. 5.1), whereas the light preference function can be represented using a Beta function (eq. 5.2).

$$F_T(\underline{\alpha}, t) = \alpha_1 \exp^{-(\frac{t-\alpha_2}{\alpha_3})^2} \tag{5.1}$$

$$F_L(\underline{\beta}, l) = \beta_1 l^{(\beta_2-1)}(1-l)^{\beta_3-1} \tag{5.2}$$

For every user, the parameters ($\underline{\alpha}$ and $\underline{\beta}$) of these functions differ. Based on the comfort indicators, we derive the individual function parameters using the least square curve fitting. To derive a reflective function from the limited samples, we assume that any temperature beyond 14° C and 30° C will be uncomfortable for any person. Thus, with the existing comfort indicator data set, we add two additional data points of these two extreme temperatures with comfort value 0 using the least square method. Similarly, for light these two extreme values are 0 and 1200 lux.

Fig. 5.4a shows the clustered preference data and the thermal comfort function derived for an individual. From this function, we can conclude that this particular user feels maximum comfort within the range of 20.25° C and 24° C . When multiple users occupy a room, a common comfort range needs to be determined. Fig. 5.4b and 5.4c, shows the common comfort range in two different rooms occupied by different set of people (each room with two occupants). It is clear that common comfort range of two different rooms can be quite different based on the individual comfort ranges of the occupants. Thus, a common temperature set-point for all the room can cause discomfort for some of the occupants or it will expend more energy by the HVAC or lighting systems or both. iLTC exploits the comfort

ranges of all the occupants in a room to determine the most energy optimal HVAC set-point for the room. Moreover, the comfort range for a user is location independent and it can be carried over when user changes her location.

Similarly, the clustered lighting comfort indicators and the fitted comfort function for two users is shown in Fig. 5.5. From these figures, we can conclude that the minimum desirable luminescence varies significantly from person to person. Unlike temperature, a common comfort range for lighting need not be derived for following reasons. First, when the light intensity becomes uncomfortable due to superfluous light, the artificial lighting system cannot be used to reduce the intensity (like a HVAC through cooling). Rather, the lighting systems remain completely off and window blinds can be used to block additional sunlight. Second, the particular light intensity from the lighting systems can illuminate differently at various parts of the room. Thus, a work-desk close to window might get sufficient sunlight, while a work-desk far away from the window might experience light deficiency. As the received light intensity from a light source (natural or artificial) differs from desk-to-desk, only individual visual comfort threshold need to be considered.

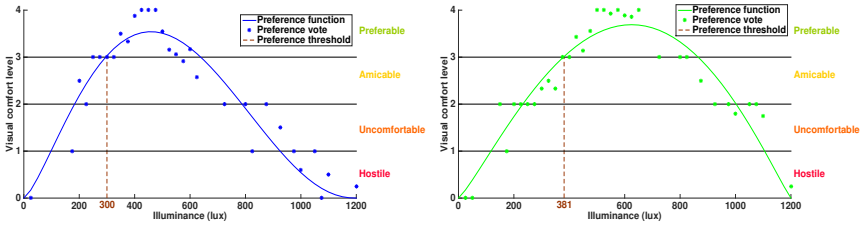
5.3.2. MODELING OF RECEIVED LIGHT AT WORK-DESKS

As the light intensity varies within a room, it is important to measure the amount of natural light reaching each work-desk for a particular outdoor light intensity. Thus a single light sensor is not sufficient to measure the amount of natural light at different locations (in case there are more occupants in a room). Moreover, setting a particular set-point (brightness level) for a light unit does not mean a uniform light intensity in all parts of the room. This necessitates measuring received amount of light at the work-desks of the users from different sources of lights – artificial and natural.

Modeling of received natural light: iLTC employs the smartphone light-sensor of a user to measure the received light intensity at his/her work-desk. However, for *natural light*, one time measurement using the smartphone is not sufficient as the natural light can vary over the days. To resolve this, we use one reference light sensor. Based on the measurement tuple of the reference light sensor and the smartphone light sensor, a relationship is established. This can be used to derive received natural light at the work-desk when there is a particular outdoor light intensity. Our goal is not to measure light intensity at every part of the room, but only at the work-desk of the occupant. Once this measurement is done, the relationship remains the same until the user changes his/her work-desk.

Though the gradient of light intensity can be expressed as, $l_r = \frac{l_s}{4\pi d^2}$, where l_s and l_r are the light intensity at the source and at a distance d from the source respectively, the gradient of light intensity inside a room cannot be described using the same relation. However, a similar form of relationship can be seen between outdoor and indoor light intensity as, $l_u = a_1 * l_o + a_2$, where l_u and l_o are the light intensity at the user work-desk and outdoor respectively. To determine the unknown parameters (a_1 and a_2) for a work-desk, we collected light values for a day after the user becomes part of the iLTC system (using her smartphone). Using the training data set, the parameters are estimated using least square method. Once these parameters are known for the work-desk, the received natural light can be estimated based on the reference light sensor values.

Intuitively, it is clear that once the parameters are found, the same parameters can be used to estimate light values for that location irrespective of the room and window dimension. However, the estimation accuracy suffers significantly if the same set of parameters is



(a) User 1 preference function (b) User 2 preference function

Figure 5.5: Comfort functions of individuals based on preference vote and lower bound of lighting comfort.

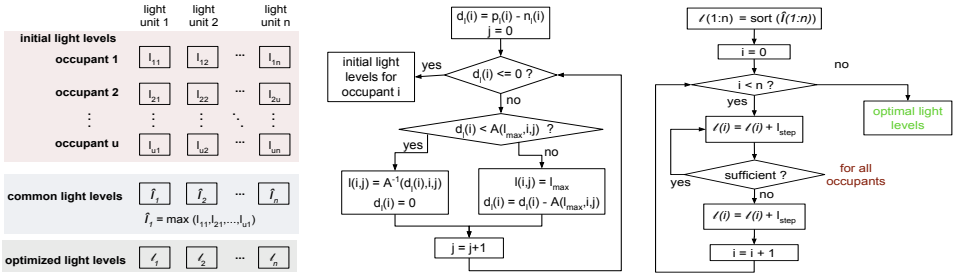
used irrespective of time of the day and weather condition. To improve the estimation accuracy, our approach splits the data set collected into multiple segments based on the light intensity values of the reference sensor. Then for each of these segments, we find the set of parameters.

Another important factor that influences the accuracy of natural light estimation is the visibility of the sun from the window (time of the day). We divide the data into various segments based on the location of the Sun and orientation of the window in the measurement room, and then determine the estimation parameters. Our evaluation shows that dividing the data set into two parts – when the Sun is visible from the window, and when it is not visible from the window – improves estimation accuracy. By visibility of Sun, we mean the Sun's location is within the visible area of the sky through the window. Thus, to estimate the natural light at a work-desk, the right set of estimation parameters are chosen based on Sun's visibility from the window and light intensity segment of the reference light sensor.

Though the visibility of Sun from a window changes drastically throughout the year, it can be easily determined. If the direction of the window and geographical location of the room (latitude and longitude) are known, visibility of Sun can easily be derived from the Sun's azimuth. To derive Sun's azimuth, we have used the algorithm provided by the measurement and instrumentation data center (MIDC) of the national renewable energy laboratory (NREL) [5]. This is a one-time data collection activity to determine the direction of the window and geographical location of the room.

Modeling received artificial light Similar to the natural light, the amount of received artificial light also varies within a room. Thus, it is also necessary to measure the amount of received light from each of the light units at the work-desk of a user. During the training period, we turn on the light units in appropriate steps to measure the received light intensity from each of the light units at the work-desks. The light units were set to full brightness. Using the collected data, the gradient of brightness can be determined. For every light unit, there is a different gradient at different work-desks.

In building iLTC, we made the following two assumptions about modeling of preference: (a) since the temperature preferences is not location dependent, it can be utilized across various rooms, including a common meeting room and home environments; (b) the same is applicable for visual comfort. However, in a different room the gradient of light intensity from Sun and artificial sources varies differently. This necessitates a different set of parameters to model received light intensity for different work-desks. Since a user spends most of his/her time at a particular work-desk, one training campaign is sufficient.



(a) overview of the algorithm (b) initial light level for user i (c) Optimized light level

Figure 5.6: A three step algorithm to decide set-points of the light units inside a room.

5.4. ROOM DAEMON

Most of the real-time activities are handled at the room daemon. Whenever a user enters or leaves the room the set-points for the actuators need to be adjusted. Moreover, if the natural conditions change, that also influences the set-point values of the actuators. Each room daemon hosts three modules – 'main thread', 'light controller', and 'temperature controller'.

5.4.1. MAIN THREAD

The main thread module manages overall execution of the room daemon. When a person enters a room, it sends an 'arrival message' to the room daemon. This message contains the identity of the users and their preferences. Upon detecting arrival of an occupant, the daemon communicates preference values to the controller modules. It also makes a temporary local copy of the user preferences along with marking the presence of the user. It periodically monitors presence of occupants inside the room, and instructs the controller modules to adjust set-points of the actuators if required. On the other hand, the user daemon sends periodic 'hello messages' to indicate the presence. If no hello message is received from a user for significant amount of time, it assumes that the user has left the room, and removes her comfort preferences. If the user daemon itself identifies that the user left the room, it explicitly sends a 'departure message' to the room daemon. Whenever a departure is detected, the controller modules are invoked to adjust the set-points if required.

5.4.2. LIGHT CONTROLLER

The goal of the light controller module is to ensure usage of artificial light only when the natural light is insufficient for lighting comfort. Given that it knows the amount of received natural light at a work-desk (l_n) and the corresponding user preference regarding the lighting comfort (l_p) (see Section 5.3.2), the amount of light deficiency can easily be calculated ($l_d = l_p - l_n$). Using the artificial light modeling described earlier, received amount of light at the work-desk can also be calculated if a particular light unit illuminates at certain brightness. Based on these, the module can decide the minimum brightness for each light unit so that the light deficiency of the user can be supplemented. A minimum amount of energy consumption is ensured since lower brightness means lower energy consumption.

When there are multiple occupants inside a room, a certain brightness level for the light units may not satisfy everyone's lighting preference. Thus, to fulfill light deficiency of all the occupants while maintaining lower energy consumption, a suitable combination of set-

points (brightness level) for each light unit needs to be decided. We formulate this as an optimization problem and it is given below.

$$\min \sum_{i=1}^{n} l(i) \tag{5.3}$$

$$\texttt{subject to:} \quad \sum_{i=1}^{n} A(l(i), i, u) \geq d_l(u), \quad \forall u, \tag{5.4}$$

$$\texttt{where,} \quad A(l, i, j) = a_1(i, j) \times l + a_2(i, j). \tag{5.5}$$

The objective of the optimization problem is to select a combination of set-points for each light unit in a room such that the light deficiency is fulfilled while having minimum energy consumption. If a light unit i sets its brightness to $l(i)$, then the received light amount for user u can be calculated using 5.5. Eventually, the combined received light amount should be at least equal to the light deficiency $l_d(u)$. If there are 4 light units and each light unit has 16 brightness levels, there are a total of 65536 combinations to choose an optimal brightness level. Light intensity can change quickly within a short time period. Thus, the light controller needs to adjust the set-points every now and then. Selecting an optimal set-point out of all possible combinations can incur significant computational cost. Thus, we propose a heuristic algorithm to find the set-point that maximizes user comfort and minimize energy consumed.

A brief overview of the algorithm is shown in Fig. 5.6. First, a combination of optimal set-points for all the light units is derived for each of the occupants based on her light deficiency. The common set-point for a particular light unit is decided by taking the maximum among individual brightness levels required for all the occupants affected by that light unit. This ensures that everyone would receive sufficient amount of light. In the final step of the algorithm, the brightness levels are decreased one step at a time to see whether this new combination can fulfill the deficiency of all the occupants. This iterative process stops, when no further decrease in brightness level is possible. Here the algorithm assumes that the brightness level of the light units can be varied. However, for traditional lighting system with only two states (on/off), we use a similar but simpler technique to decide whether a light unit should be On or Off at any time instant. When there is a quick drop in natural light levels, the brightness levels are not increased immediately, rather a similar iterative approach is considered but with faster rate of change in brightness level. This ensures occupants do not notice any immediate fluctuation in the light units.

5.4.3. TEMPERATURE CONTROLLER

The temperature controller module decides temperature set-point for the HVAC, which maximizes the comfort of all occupants and minimizes HVAC energy consumption. Current HVAC systems are quite efficient in terms of maintaining a room temperature based on the given set point. Furthermore, with the introduction of zone heating, HVAC systems can now eliminate hot or cold spots in a room and maintain a set temperature value across the room. Moreover, several research efforts are conducted to determine the optimal pre-conditioned temperature of a room before an occupant arrives or after she leaves. In this regard, we focus on how to obtain an optimal temperature set-point, which maximizes the comfort of all occupants and minimizes the energy consumed by the HVAC system. Determining a set-point is not trivial when there are multiple occupants present in a room. The

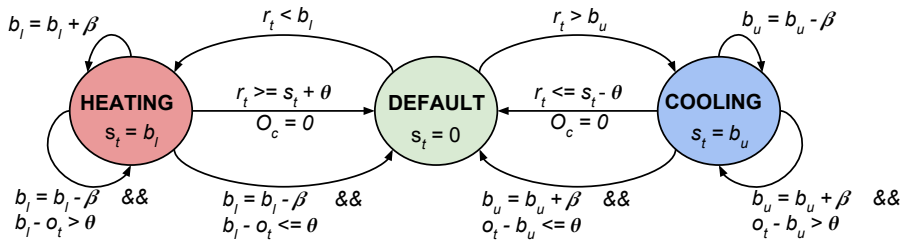


Figure 5.7: State diagram of the HVAC operation and associated temperature set-point (s_t) assignment. Transitions among the states depend on the room temperature (r_t), outside temperature (o_t), number of occupants (O_c), and the common thermal comfort range ($< b_l, b_u >$) of all the occupants.

temperature controller finds the common comfort range of occupants based on the preferences collected previously.

A HVAC can be at three different operating states – default, heating or cooling. Default is a state when the HVAC consumes minimal amount of energy – without loss of generality, it can be the Off state. In heating state, the HVAC blows warm air inside the room such that the room temperature reaches a set-point value. The warmth of the air is decided based on difference between the desired set-point temperature and current temperature, whereas the energy consumption depends on the difference between the desired set-point temperature and outdoor temperature. In the cooling state, HVAC operation is similar.

We now describe how HVAC states are switched and how the set-point temperature is decided (Fig. 5.7) so that all the occupants feel comfortable. In the beginning, the HVAC is in the default state and its set-point is set to zero ($s_t = 0$). When an occupant enters the room, the room daemon receives her thermal comfort function. Then, the temperature controller finds a comfort range for the person, which includes a lower (b_l) and upper (b_u) bound. The module also gets the current room temperature (r_t) from the temperature sensor. If the current room temperature is within this bound, then HVAC continues to remain in the default state. The module periodically checks if the room temperature falls out of this bound, and changes its operating state. This can happen for two reasons: (i) the room temperature changes due to occupants and/or the outdoor temperature; and (ii) the bounds of the comfort range are modified (narrowed) because of a new occupant entering the room. Hence our system periodically monitors the user presence in the room and the current temperature to maximize user comfort.

In case, the room temperature falls below the lower bound of the common comfort range ($r_t < b_l$), the HVAC enters the heating state, and the set-point is set to this lower bound ($s_t = b_l$). This ensures that all the occupants comfort preferences are met. At the heating state, if a new occupant leaves/arrives, the bounds for the common comfort range are modified. If the lower bound increases compared to the previous one ($b_l = b_l + \beta$), that means the heating need to be continued and the set-point is adjusted to the new lower bound. In case the lower bound gets reduced, there is a possibility of decreasing the heating intensity. If the new lower bound is significantly higher than the outside temperature ($b_l - o_t > \theta$), then the set-point is adjusted to the new lower bound and the HVAC continues in this state. Otherwise the HVAC is switched to the default state. At the heating state, if all the occupants leaves the room (O_c) or the room temperature reaches sufficiently higher

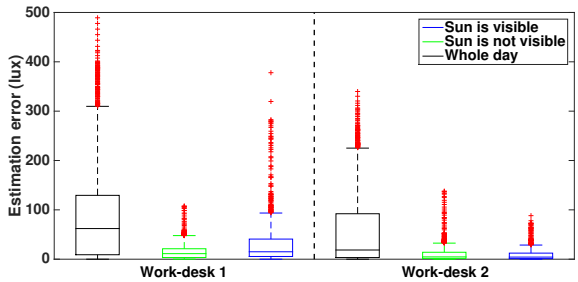


Figure 5.8: Estimation error of the received natural light at two work-desks.

than the set-point temperature ($r_t >= s_t + \theta$), the HVAC enters the default state with set-point being zero. The switching between heating to default state is guarded with a threshold θ to ensure that the state change does not happen frequently. A similar switching happens on the right side of the state diagram when room temperature goes beyond the upper bound of the common comfort range and the HVAC enters the cooling state. As mentioned earlier, this work focus on deciding the optimal set-point and assumes that the current HVAC system is capable of maintaining the set temperature value based on the physical conditions of the room. Thus by constantly monitoring the room condition and occupancy, iLTC adapts the set-points to maximize user comfort and minimize energy consumption of the HVAC.

5.5. RESULTS

In this section, we describe our experimental setup and provide details of all the sensors used during the setup. Further, we present results regarding modeling of received light at the user work-desk and energy savings incurred with the deployment of iLTC system in shared environments.

5.5.1. EXPERIMENTAL SETUP

iLTC system was deployed in an office environment with multiple rooms. The number of lights, window size and room size can vary depending upon the building considered. However, the functioning of the iLTC system is independent of these parameters. The set of devices used for our measurements, actuation and data collection includes: (i) Moteiv tmote-sky sensor nodes measuring temperature and light intensity in indoor and outdoor locations, (ii) Smartphones from different manufacturers for localization and user comfort indicator, (iii) Philips hue light bulbs for indoor lighting, and (iv) Plugwise circles were used to measure energy consumption of the hue bulbs.

Our experimentation considered 21 participants in an office environment from 5 different ethnic background with age varying from 25 to 51. The participant list includes both male and female users and all the users had their smartphone.

5.5.2. iLTC EVALUATION

The goal of iLTC is to decide set-points for the actuators such that (i) all the co-occupants in a room can be provided maximal comfort; and (ii) the energy consumption by the actuators is kept minimal.

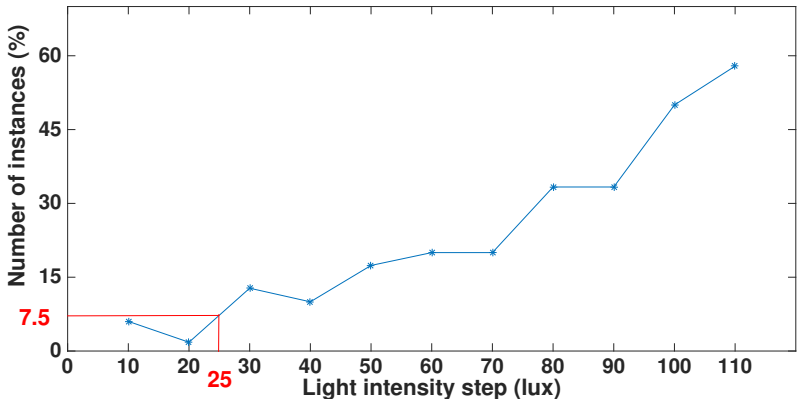
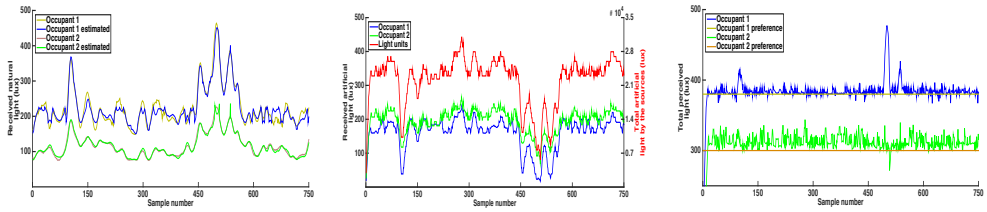


Figure 5.9: Various light intensity change in one step versus the number of instances when people felt annoyed with that sudden change.

Lighting energy reduction. Deciding light unit set-point requires information about deficiency in natural light conditions. Thus, we first show the natural light estimation at the user work-desk using the estimation technique proposed in Section 5.3.2. Fig. 5.8 shows the estimation error of the received natural light at two work-desks. It can be seen that, when the whole day data is considered a high error is associated with the estimation. On the other hand, by splitting the data set based on visibility of sun significantly reduces the estimation error. This is mainly due to the consideration of rate of change of natural light with respect to sun visibility at the user work-desk.

Fig. 5.9 shows the light intensity steps (in lux) and the corresponding number of instances when the occupant noticed the change in light intensity. We experimented this across participants and the average results are shown in Fig. 5.9. It can be seen that, when the combined received light intensity at the work-desk changes in bigger steps more number of users feel annoyed. From our experiments, we determined 25 lux to be the step change, applied when increasing or decreasing the light intensity to prevent user inconvenience. In Section 5.3.1, we discussed that a range of temperature and light values are clustered into bin before deriving the comfort functions. For the light values, a bin size of 25 lux is used for the range of 0 to 1200 lux (This is also clear from Fig. 5.9). For temperature, it starts from 14° C until 30° C with a bin size of 0.25° C , that means any comfort label indicator for the range 18° C to 18.25° C is mapped to 18° C .

The light intensity of the hue bulbs considered in our experimentation are within the range 600 to 16000 lux near the source, and the energy consumption ranges from 0.58 W/s to 5.4 W/s. Unlike the temperature set-point, setting a particular brightness of light units does not guarantee the required level of lighting comfort at the user's work-desk as the light intensity degrades significantly while moving away from the source. Consider two users inside a room with minimum light comfort preference of 300 and 381 lux (Fig. 5.5). Fig. 5.10a shows the amount of natural light received at the work-desks over a time period. The light samples are measured every 10 s. During the real deployment of the system, the measured value of received light intensity will not be available. Thus, we use an estimate of received amount of light using the reference sensor as described earlier. It can be seen that, the



(a) Measure and estimated natural light

(b) Received artificial light

(c) Total perceived light

Figure 5.10: Based on the estimation of the received natural light at two work-desks, and their corresponding light preferences set-points are decided for the light units.

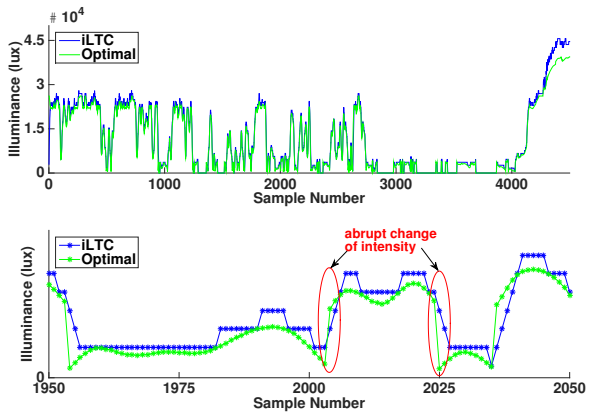


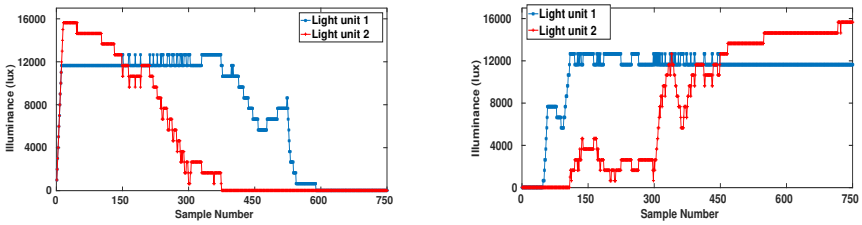
Figure 5.11: Luminance of all light units: a comparison between iLTC and the optimal light level selection.

estimated light intensity closely follows the actual received light intensity. This estimated light is used as input for the light controller module.

As mentioned in Section 5.3.2, we measured the gradient of light intensity from each source at each work-desks. Fig. 5.10b shows the total amount of light seen at the source (with all six lights) and the received light intensity at the user work-desks. The decrease of received light at the work-desks is indeed due to distance from the artificial lights. Furthermore, Fig. 5.10c shows the total perceived light at the work-desk by considering both natural and artificial light. It can be seen that, lighting preferences of both the occupants are always met by adjusting the brightness level of the lighting system when necessary.

As mentioned in Section 5.4.2, the light controller uses a computationally inexpensive algorithm to decide the set-points. From Fig. 5.10c, it is clear that the algorithm serves the purpose of providing lighting comfort to all the occupants. To evaluate efficiency of the algorithm, we derived the set-points using an optimal solution and compared it with iLTC solution. The optimal solution is found by testing all the possible combinations of brightness levels.

Fig. 5.11 shows the combined brightness of all the light units using the optimal and iLTC set-point solutions. As the energy consumption is directly proportional to the brightness



(a) Brightness (set-point) decreases over time for two light units.

(b) Brightness increases over time for two light units.

Figure 5.12: Variation in light intensity to maintain user comfort across two light units.

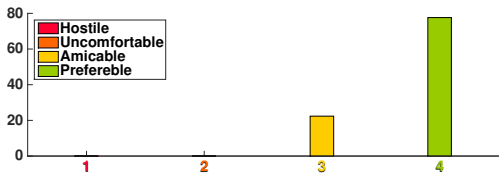


Figure 5.13: Visual comfort feeling of people when the artificial light units are controlled automatically using iLTC.

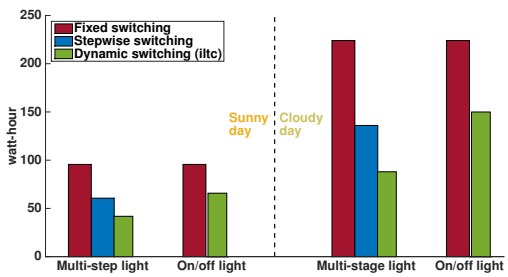


Figure 5.14: Combined energy consumption by all light units in a room for different switching strategies.

level, this also reflects the level of energy consumption. From the figure, we can conclude that the iLTC solution is very close to the optimal solution. To be precise, iLTC uses only 6.92% higher light than the optimal solution. However, the number of iterations to find a suitable brightness levels using iLTC is a mere 0.01% of the optimal solution. There is another significant drawback with the optimal solution. In order to find the least energy consuming brightness levels, the optimal solution can change the brightness levels of the light units too frequently with certain change in natural light level. This may cause annoying experience to the users. On the other hand, iLTC ensures that whenever the brightness level gets decreased, it decrease by only one level of brightness (a closer look at Fig. 5.11 for samples between 1950 and 2050). From our empirical evaluation, we found that a lux difference of 25 is unnoticeable by the users at their work-desk and we used that as the minimum brightness step (see Fig. 5.9). This ensures that the user will hardly notice any change in the brightness when decreasing the light intensity.

Fig. 5.12a shows the reduction in light intensity to maintain user comfort and minimize energy consumption across two lamp units. It can be seen that the light intensity is reduced

	Sunny day		**Cloudy day**	
Switching Method	Multi-step light	On/off light	Multi-step light	On/off light
Fixed	56.25%	31.20%	60.71%	33.06%
Stepwise	31.03%	–	35.29%	–

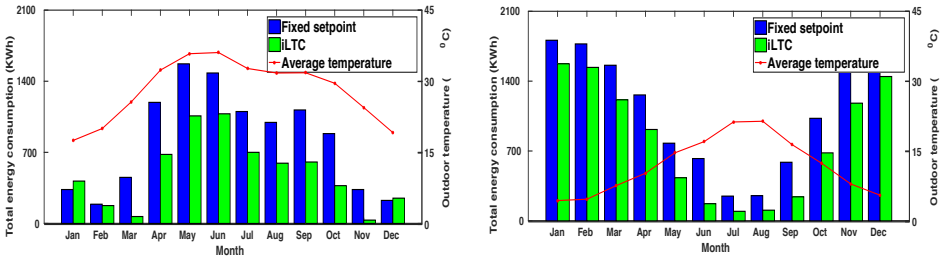
Table 5.1: Reduction of energy consumption by iLTC as compared to fixed and stepwise switching of the light units.

iteratively such that not more than 25 lux difference is perceived at the user's work-desk. This approach ensures sudden fluctuations in natural light do not affect the user comfort. Moreover, when the natural light is not sufficient, artificial lights are turned ON to maintain the user comfort levels. For example, in evening the natural light perceived at the user work-desk might be lower than the comfort preference. During this stage, we follow an iterative approach to increase the lux value, but at a faster rate rather than abrupt change in lux causing inconvenience to the user. Fig. 5.12b shows the rise in lux value to maintain user comfort across two light units. Thus iLTC avoids both abrupt increase and decrease in light intensity and follows a iterative approach to achieve the same without causing too much inconvenience to the users.

We also conducted a post-deployment user evaluation to determine the efficacy of the system (Fig. 5.13). Based on the user preferences collected, the set-points for the lighting system and the brightness level of the light units were constantly adapted. The feedback was collected using a questionnaire available on the smartphone App. The questionnaire comprised of questions related to visual comfort feeling. Each user selects one of the comfort levels *viz.*, (i) hostile, (ii) uncomfortable, (iii) amicable, and (iv) preferable based on the current set-points decided by the iLTC. The post-deployment evaluation was conducted on several days to generalize the outcome. On an average 78% of the feedback from 21 participants indicated preferable comfort feeling when the light intensity was adjusted due to either insufficient light intensity or excess light intensity. 22% feedback received indicated amicable feeling at certain time periods. A closer look revealed that these are due to the variation in comfort preference of the user and also due to sudden fluctuations in natural light perceived at the user work-desks. This change is preference was further considered to adapt the user preference models accordingly.

Finally, to evaluate the efficiency of the light controller, we compared iLTC with two switching mechanisms. (i) Fixed switching – where the light units are turned on to the maximum levels when the outdoor light intensity drops beyond a certain threshold. This threshold is decided when either of the users face light deficiency. (ii) Stepwise switching – where the light units are turned on with varying brightness with the variation of outdoor light intensity. We tested these strategies on the data sets for a sunny day and a cloudy day, where on the cloudy day indoor light intensity was insufficient for the occupants almost throughout the day. The total energy consumption by all the light units is shown in Fig. 5.14 when various switching strategies are used. All the strategies considered include occupancy detection before turning on the lights. iLTC based switching consumes the least energy as compared to other strategies. This is mainly due to the individual control of brightness level at each light unit. Table 5.1 shows the total reduction in energy consumption by iLTC as compared to fixed and stepwise switching mechanisms.

HVAC energy reduction. To compare the energy consumption of the HVAC, we adopted the energy-temperature correlation model $P = |\frac{\lambda}{M}(t_i - t_o)|$ as described in [120], where P is



(a) Delhi, India (b) Delft, Netherlands

Figure 5.15: Comparison of energy consumption by the HVAC for fixed set-point technique and iLTC. Considering the total yearly consumption, iLTC is 39% and 27% less expensive based on the outdoor temperature in two cities.

the amount of energy consumed by the HVAC system in one second, λ is the conductivity of a particular room, M is the efficiency of the HVAC system, and t_i and t_o are the indoor and outdoor temperatures respectively. For a particular room, λ and M are constant. So, the HVAC energy consumption is mainly dependent on the difference in set-point temperature and outdoor temperature. Additional details of the HVAC system such as duct type, radiation/convection, air re-circulation is not considered as it varies from one HVAC system to another and also dependent on the building characteristics. Our objective is to show the potential energy savings by finding the optimal set-point to maximize user comfort and minimize energy consumption. Thus we use a simpler energy model based on difference between indoor set-point and outdoor temperature as described in [120].

For our evaluation, we fixed the values of λ and M to be 70.5 J/s.K and 0.14, respectively (from [120]). We consider two rooms at two different parts of the world – (i) Delft, the Netherlands, and (ii) Delhi, India. We collected the yearly weather data for these two cities from public repositories [6]. Also there were two occupants in these rooms in both the places with comfort range spanning from 21.75° C to 26.5° C (Fig. 5.4b), and from 18.75° C to 22.25° C (Fig. 5.4c). iLTC sets the room temperature based on the common comfort range of the occupants and the outdoor temperature, whereas a fixed temperature set-point strategy selects a fixed temperature for rooms irrespective of the comfort preferences of the occupants. However, we assume that the fixed set-points also vary between 21° C to 23° C from winter to summer months. Both the strategies employed occupancy detection before selecting a set-point.

The total energy consumption of the HVAC on a monthly basis is shown in Fig. 5.15. In most of the cases, iLTC incurs significantly less energy consumption than the fixed set-point strategy. However during the winter season in India, iLTC induces more energy consumption. This is because the lower bound of the common comfort range is 21.75° C , where the fixed-point strategy sets the temperature to 21° C . But, when yearly basis energy consumption is calculated, iLTC outperforms fixed-point strategy in terms of lesser energy consumption. In India, the total yearly energy consumption is 6032 kWh and 9881 kWh for the two methods, respectively. On the other hand, in Netherlands, they are 13129 kWh and 9595 kWh, respectively. Thus, iLTC reduces energy consumption by 39% and 27% in the respective cities.

5.5.3. DISCUSSION

As demonstrated in the previous sections, iLTC system can decide a set-point to reduce energy consumption to maximize the comfort level for all the co-occupants in a shared space. However, there are a few challenges that need to be addressed: (i) The light intensity values collected from smartphone of users may vary due to the heterogeneity of the sensors used by different manufacturers. Hence, the data collected needs to be calibrated to derive accurate user comfort preferences. Data calibration can be easily performed by comparing the sensed data with a baseline sensor data. (ii) In some scenarios, there may not be any common comfort range between the co-occupants in a shared space. It could even be discontinuous when more than two occupants share the space. iLTC then determines a set-point to save energy and also minimize the average discomfort for all the co-occupants. (iii) The efficiency of traditional BEMS can be very low, when there is frequent user movement. Our iterative approach in iLTC for controlling the actuators ensures that frequent movement of users does not affect the overall comfort drastically. (iv) The HVAC model utilized here shows the energy saving considering only the temperature difference between outdoor and indoors, however sophisticated simulation tools can be employed to derive detailed energy savings by considering other building parameters such as duct type, air re-circulation, zone thermal storage, etc. iLTC can take any given model and tries to set the operating point; (v) Since iLTC system measures the current light intensity and temperature at a specific space, it is agnostic with respect to building type and the surrounding spaces. It learns the comfort preferences and decides on energy optimal set-points. Hence iLTC can also be used in large shared spaces with multiple occupants; (vi) User involvement in iLTC is minimal where new users can join and leave the system freely.

5.6. CONCLUSIONS

We developed an indoor environment controlling system called **iLTC** that offers automated HVAC and lighting control at the room level trying to fulfill individual user preferences. Instead of choosing a conservative set-point for the actuators that can provide nominal comfort to the occupants in a shared space, iLTC decides a set-point that can be energy optimal while considering comfort levels of all co-occupants. The system learns the preferences of each individual based on human perception of comfort through the developed smartphone *App*. We developed a comprehensive comfort representation function from a few comfort indicators using the collected data, and reduced explicit human intervention. We leveraged the light sensor in a smartphone to monitor the received light at users' desk in addition to a single reference light sensor for all the users in the building. Thus iLTC reduces the deployment and management costs of multiple light sensors. Results show that iLTC's set-point selection can reduce energy consumption up to 39% and 60% by the HVAC and lighting systems, respectively, compared to the fixed set-point mechanism. We evaluated iLTC with 21 participants housed in multiple rooms and qualitative user evaluation shows over 78% of the participants felt comfortable with the deployed iLTC system.

In chapters 4 and 5 we presented services that shift or reduce energy consumption by considering individual preferences in smart homes/buildings. In the next part, we present services that can identify target households/buildings in a neighborhood for various DR programs.

III

Demand Regulation in Neighborhoods

6

Temporal Demand Regulation

The introduction of smart meters at a large scale offers new opportunities for fine-grained real-time data collection. Energy utilities can now devise targeted demand-response (DR) programs, tailored energy-saving advice, and dedicated electricity tariffs for specific consumers. Energy consumption is highly influenced by consumer behavior and their characteristics. Rather than selecting all the households in a neighborhood for a DR event, an effective DR mechanism should first identify the set of target consumers and then apply the DR technique. However, in a neighborhood or community with thousands of households, heterogeneity in consumer preferences hinders identifying consumers for specific DR programs such as reduction in average energy consumption, reduction in demand peaks, etc.

Recent research works have shown that analyzing the energy consumption data of consumers plays a crucial role in the classification of consumers [38]. Analyzing smart meter data is challenging due to several reasons *viz.*, (i) huge amounts of energy data from smart meters need to be processed effectively, (ii) the granularity of smart meter data and relevant features therein need to be identified across households, (iii) there is no comprehensive method to analyze large-scale smart meter data from households.

In this chapter we present Temporal Demand Regulation (TDR) to analyze and classify households based on their historic energy-consumption data. We describe a new perspective to study energy demand enabling the design of novel mechanisms for decentralized demand-side energy management. Rather than only optimizing the demand levels of each household so that it meets available supply, the concept of computing the demand states of each household and feasible transitions between these states are introduced. The demand states measure one of the following features: (i) demand level, (ii) demand variation and (iii) demand peaks. In contrast to the related work [62, 93, 125], we show that the orchestration of temporal transitions between the demand states can meet a broad range of Smart Grid objectives set by the utility companies. A *generalized data-driven* methodology based on clustering of historic consumption data (time-series) from each household is designed for a local computation of the demand states at different aggregation granularities, e.g., daily, weekly, etc. This methodology captures the temporal dynamics of demand and can be used to identify target consumers for DR programs. The proposed methodology is decentralized, highly scalable and privacy preserving. This can be used to build effective real-time recom-

mendations for the self-regulation of demand. The methodology can be further applied to other time series data, especially resource-consumption data such as water and gas. Further, an online self-regulation model for the adjustment of demands by targeted consumers is proposed. The selection criteria is governed by four temporal metrics, viz., transition probability, temporal membership, temporal adaptability and temporal similarity. TDR is evaluated and validated using data from a real-world Smart Grid project consisting of more than 4,000 households.

The objective of TDR is two-fold, (i) identify target households for DR programs by classifying households based on their historic energy consumption data and (ii) a decentralized self-regulation model for the adjustment of demands by targeted households.

Contributions. The main contributions of this chapter are:

- We present a generalized, domain-independent data-driven model and methodology for the computation of demand states.
- We propose four temporal metrics for measuring and evaluating demand adjustments.
- We present an online self-regulation model for the adjustment of demands by targeted consumers and its validation using survey responses of consumers.
- We describe our evaluation of the methodology using demand data from a real-world Smart Grid project with 4,000 households.

6.1. RELATED WORK

Numerous DR programs [33, 62, 93, 109, 125] have been proposed to motivate changes in the consumers' power consumption. These DR programs can be broadly classified into centralized and decentralized schemes [115]. In centralized schemes, a central controller collects all the demand information from consumers for DR decisions [62, 125]. Decentralized schemes allow consumers to coordinate directly with each other to participate in DR programs [24, 34, 93].

Successful implementation of DR programs relies on the identification and participation of target consumers. The majority of previous efforts on the identification of target consumers relied on customer self-reported data [18, 105]. Large-scale deployments of smart meters has paved the way to analyze real-time energy consumption to provide insights into the energy usage of households [63, 82, 83]. Moss *et al.* investigate the segmentation of consumers into groups based on the similarity of energy usage [82], whereas, Chicco *et al.* study different unsupervised clustering algorithms to classify consumers, based on the load pattern shape [38]. The majority of earlier work does not study the temporal transitions for classification of households. In contrast, we analyze the temporal dynamics of demand by considering multiple features such as average and peak consumption.

A plethora of optimization-based models are designed for DR programs with various objectives. Zhu *et al.* derive optimal power consumption, by taking into account loads that can shift or adjust their consumption in successive time periods [125]. This centralized approach requires consumers to communicate their demand needs and usage patterns for each appliance. In contrast, TDR analyzes the temporal demand of households to derive consumer characteristics such as how often the demand pattern varies and which consumers are willing to participate in DR. Joe-wong *et al.* propose a day-ahead device-specific

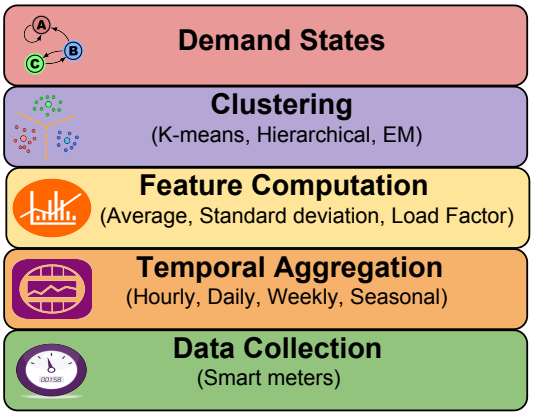


Figure 6.1: Computation of demand states.

scheduler that is based on task schedules, which considers heterogeneity in appliance delay tolerance [62]. This centralized model employs convex optimization to derive demand schedules. However, the main problem is that it requires fine-grained appliance-level energy data and also continuous real-time communication between the energy provider and the consumers. Recent work [93] shows how to manage the energy demand of households by analyzing historic aggregated energy-consumption data. Pournaras *et al.* propose a decentralized approach for demand-side self-management [93], where software agents represent the demand preferences of consumers and control their demand by selecting a plan according to the criteria defined by a selection function. The decentralized approach enforces all the consumers to select a plan that meets the DR objective set by the utility. In contrast, TDR identifies the target consumers who can participate in different DR programs by analyzing the temporal dynamics of demand. Baharlouei *et al.* propose a decentralized scheme along with a fairness index to minimize total generation cost with a smart billing mechanism [24]. This approach assumes all consumers are flexible in participating towards DR. In TDR, the selection of consumers and the discomfort associated with the demand regulation is governed by the threshold parameter θ. Several insights obtained from temporal analysis can be applied to develop more effective consumer-centric DR programs. Majority of the literature is concerned with simulations or numerical analysis compared to the real data employed here.

6.2. MODELING TEMPORAL DYNAMICS OF DEMAND

A generalized data-driven model and methodology (illustrated in Fig. 6.1) for computing local demand adjustments for each household is introduced in this section. An outline of mathematical symbols used in this article is given in Table 6.1. Each household is assumed to be equipped with an information system that collects and stores real-time demand measurements using smart meters [77]. The collected data are aggregated at different granularity levels – daily, weekly or seasonal. The information system manages l samples of historic demand series $\mathbf{d}_1$,...,$\mathbf{d}_l$, with $\mathbf{d}_1$ being the most recent historic time series and $\mathbf{d}_l$ is the earliest. Each demand series consists of $T = |\mathbf{d}_j|$ demand measurements, therefore, T is the number of measurements aggregated for a certain granularity.

The information system serves the DR program of the utility companies by turning the *forecasted* demand series $\mathbf{d}_{l+1}$ to the *regulated* demand series $\hat{\mathbf{d}}_{l+1}$. Such an adjustment is achieved by mining the historic demand series to infer and reason about possible demand changes observed in each household. Utility companies may introduce one or more features for characterizing and assessing the quality of the forecasted and regulated demand. The quality of the demand represents the characteristics extracted from the demand time series. For example, a household demand with low load factor shows occasional high demand peaks resulting in low quality of demand.

The quality $\mathbf{q}_j$ of a demand series $\mathbf{d}_j$ is defined by a set of m measurable features $\mathbf{q}_j = \{p_j^1, ..., p_j^m\}$, where p_j^u is the property of a demand series $\mathbf{d}_j$ according to the feature u at time j. A property p_j^u is defined as $p_j^u = f_u(\mathbf{d}_j)$, where $f_u(\mathbf{d}_j)$ is a function performed over the demand time series.

This work focuses on $m = 3$ quality features of demand: (i) *average* (AVG), (ii) *relative standard deviation* (RSD) and (iii) *load factor* (LF). The average (AVG) feature is defined as,

$$p_j^1 = f_1(\mathbf{d}_j) = \frac{1}{T}\sum_{t=1}^{T} d_j^t, \tag{6.1}$$

where, $d_j^t \in \mathbf{d}_j$ is the demand measured at time t within the demand time series $\mathbf{d}_j$. This feature indicates the aggregate demand over the time period T and does not provide information about how demand is distributed over T. In contrast, relative standard deviation (RSD) feature computes the homogeneity of demand over time period T and is defined as,

$$p_j^2 = f_2(\mathbf{d}_j) = \frac{1}{p_j^1}\sqrt{\frac{1}{T}\sum_{t=1}^{T}(d_j^t - p_j^1)^2}, \tag{6.2}$$

where, $d_j^t \in \mathbf{d}_j$ is the demand measured at time t within the demand time series $\mathbf{d}_j$. Note that the average demand over the time period T in demand time series $\mathbf{d}_j$ is indicated in (6.1) by the property p_j^1. Finally, the load factor (LF) [33] determines the scale of demand peaks and is computed by the ratio of average demand and maximum demand measured over a time period T and is defined as,

$$p_j^3 = f_3(\mathbf{d}_j) = \frac{p_j^1}{\max \mathbf{d}_j}, \tag{6.3}$$

where, the property p_j^1 denotes the average demand over the time period T. The $\max \mathbf{d}_j = \max(d_j^t), \forall t \in \{1, .., T\}$ denotes the maximal element that corresponds to the maximum demand peak during the time period T for the demand time series $\mathbf{d}_j$.

Demand, and its quality features, can be forecasted by analyzing the historic demand time series. For example, the average demand p_{l+1}^1 at time period $l+1$ can be predicted by using the average demand $p_1^1, ..., p_l^1$ during the past l time periods. Although a broad range of data mining and machine learning algorithms can be used for predicting future demands, the main focus here is on clustering because of the following reasons: (i) clustering is an unsupervised method that does not require labeling of the demand data; (ii) future demand predictions can be determined by analyzing the centroids of the clusters and their corresponding sizes [109]; (iii) the possible states, in which a feature of demand may be, can be

extracted via clustering. For example, by clustering the past average demand $p_1^1,...,p_l^1$ into *three* clusters, the centers of the clusters ranked from low to high indicate the low, medium and high demand states of a household; and (iv) clustering provides information about the temporal transitions between different demand states that represent the center of the clusters. In this way, the temporal dynamics of demand are modeled, since clustering reasons about whether or when certain demand transitions are feasible by each household.

Given l demand properties $p_1^u,...,p_l^u$ of a feature u, clustering to k clusters is defined as,

$$\bigcup_{o=1}^{k} \mathbf{c}_u^o = p_1^u,...,p_l^u, \tag{6.4}$$

where, $\mathbf{c}_u^o$ is the cluster o containing demand properties for the feature u. For each cluster $\mathbf{c}_u^o$, the center c_u^o is computed by the centroid or medoid [30]. *Expectation Maximization* (EM) clustering [39] is employed here to determine the number of clusters based on the demand properties.

When a demand property changes its membership from one cluster to another, this is defined as a *transition*. A *demand state* $s_j^u = o \in \{1,..,k\}$ is defined by the cluster index to which the demand property p_j^u belongs. States s_{l+1}^u and $\hat{s}_{l+1}^u$ represent forecasted and regulated demand states, respectively for a feature u. A sequence of z transitions defines a *demand adjustment* observed or triggered at time j and is given by,

$$\mathbf{a}_j^u = \{s_j^u,..,s_{j+z}^u\}, \tag{6.5}$$

where, $\mathbf{a}_j^u$ is a sequence of transitions starting from state s_j^u of feature u at time point j to state s_{j+z}^u with $z = |\mathbf{a}_j^u|$.

6.3. MEASURING DEMAND ADJUSTMENT

This section defines the following four metrics to measure and evaluate demand adjustments, *viz.*, (i) *transition probability*, (ii) *temporal membership*, (iii) *temporal adaptability* and (iv) *temporal similarity*.

Transition probability.

It measures the probability of moving from a certain demand state to another demand state. Given a quality feature u, the average transition probability $\mathsf{T}_{a\to b}^u$ from demand state a to b is defined as,

$$\mathsf{T}_{a\to b}^u = \frac{1}{l-1}\sum_{j=1}^{l-1}\beta_j;\ \ \beta_j = \begin{cases} 1 & \text{if } s_{j+1}^u = b \mid s_j^u = a \\ 0 & \text{if } s_{j+1}^u \neq b \mid s_j^u = a \end{cases} \tag{6.6}$$

where, β_j is a binary variable that equals '1' if a transition from demand state a to b occurs at time j or '0' otherwise. It holds that $\mathsf{T}_{a\to b}^u \in [0,1]$.

Temporal membership.

This metric evaluates the probability of a certain demand state occurring over time. The temporal membership M_o^u of a demand state o for feature u is defined as,

$$\mathsf{M}_o^u = \frac{1}{l}\sum_{j=1}^{l}\gamma_j;\ \ \gamma_j = \begin{cases} 1, & s_j^u = o \\ 0, & s_j^u \neq o \end{cases} \tag{6.7}$$

Table 6.1: List of mathematical symbols used.

Symbols	Description
l	Number of historic demand time series
$\mathbf{d}_j$	A demand time series at time j
T	Number of measurements of a demand series
j	Time index, e.g., jth aggregation time period
$\mathbf{d}_{l+1}$	Forecasted demand series
$\hat{\mathbf{d}}_{l+1}$	Regulated demand series
$\mathbf{q}_j$	Quality of a demand series $\mathbf{d}_j$
u	Quality feature e.g.,, AVG, RSD, LF
m	Number of quality features
p_j^u	Demand property of feature u at time j
$\mathbf{c}_u^o$	Cluster o for feature u
k	Number of clusters computed
s_j^u	Demand state of a feature u at time j
s_{l+1}^u	Forecasted demand state for feature u
$\hat{s}_{l+1}^u$	Regulated demand state for feature u
$\mathbf{a}_j^u$	Demand adjustment for feature u at time j
z	Number of state transitions
$\mathsf{T}_{a\to b}^u$	Transition probability from demand state a to b for feature u
θ	Transition probability threshold
M_o^u	Temporal membership of a demand state o for feature u
$\mathsf{A}(\mathbf{a}_z^u)$	Temporal adaptability on a demand adjustment $\mathbf{a}_z^u$
$\mathsf{S}_{x,y}^u$	Temporal similarity between two consumers x and y for a demand feature u
β, γ, σ	Binary variables to compute temporal metrics
J	Jaccard similarity coefficient

where, γ_j is a binary variable that equals '1' if the demand state s_j^u occurs at time j or '0' otherwise.

Temporal adaptability.

This metric measures the probability of a demand adjustment occurring over time. Temporal adaptability $\mathsf{A}(\mathbf{a}_{l+1}^u)$ of a demand adjustment $\mathbf{a}_{l+1}^u$ for feature u and with size $|\mathbf{a}_{l+1}^u| = z \leq l$ is defined as,

$$\mathsf{A}(\mathbf{a}_{l+1}^u) = \frac{1}{l-z+1} \sum_{j=1}^{l-z+1} \sigma_j; \;\; \sigma_j = \begin{cases} 1, & \mathbf{a}_j^u = \mathbf{a}_{l+1}^u \\ 0, & \mathbf{a}_j^u \neq \mathbf{a}_{l+1}^u \end{cases} \tag{6.8}$$

where, σ_j is a binary variable that equals '1' if the demand adjustment $\mathbf{a}_{l+1}^u$ defines the same sequence of transitions as the sequence of the demand states $s_j^u, .., s_{j+z}^u$. Otherwise it holds $\sigma_j = 0$.

Temporal Similarity.

This metric evaluates the similarity between the demand states of two consumers. Temporal similarity $\mathsf{S}_{x,y}^u$ between the demand states of consumer x and consumer y for a feature u is defined by the Euclidean distance as,

$$\mathsf{S}_{x,y}^u = \sqrt{\sum_{j=1}^{l} (s_{j,x}^u - s_{j,y}^u)^2} \tag{6.9}$$

where, $s_{j,x}^u, s_{j,y}^u$ represent the demand states of two households x and y, respectively.

6.4. ONLINE SELF-REGULATION OF DEMAND

A model that improves the quality of demand by a transition from the forecasted state s^u_{l+1} to the regulated state $\hat{s}^u_{l+1}$ is introduced in this section. Demand quality is improved by adjusting one of the demand properties (see Section 6.2), e.g., performing a transition to a demand state with reduced demand, lower variation in demand or lower demand peaks. A heuristic is presented to select consumers who can perform such a transition. The heuristic employs the temporal adaptability metric to quantify the probability of each consumer to perform such a transition. The criterion for selection of target consumers is governed by the *threshold* θ. For example, if a consumer has $\theta = 0.2$ and $\mathsf{A}(\mathbf{a}^u_{l+1}) = 0.25 > \theta$, the model reasons that this consumer can self-regulate its demand, i.e., it can perform the change to regulated state using the forecasted state. Otherwise, if $\mathsf{A}(\mathbf{a}^u_{l+1}) < \theta$ the consumer remains in the forecasted state. This threshold can be selected by the utility companies, each consumer or it can even be the result of a negotiation between the two parties. For example, utility companies can provide monetary incentives to consumers for lower values of θ so that they increase the likelihood of participation in DR programs in case of a high overload in the power grid.

Algorithm 1 A heuristic for online self-regulation of demand.

Input: Demand properties $p^u_1, ..., p^u_l$, the forecasted state s^u_{l+1} and the threshold θ.

Training phase:
1. Compute the demand states by clustering $p^u_1, ..., p^u_l$ as in (6.4).
2. Compute the transition probability $\mathsf{T}^u_{a\to b}$ for all possible transitions.
3. Compute the transitions from step **2** that satisfy the DR objective.
4. Compute the regulated state $\hat{s}^u_{l+1}$ from the transitions of step **3** with maximum $\mathsf{T}^u_{a\to b} < \theta$.

Testing phase:

if $\mathbf{a}^u_l = \{s^u_l, s^u_{l+1}\}$ satisfies the DR objective **then**
  5. No demand regulation is required.
else
  6. Change from forecasted state s^u_{l+1} to the regulated $\hat{s}^u_{l+1}$.
  7. Compute efficiency: AVG reduction, RSD reduction or increase in LF.
end if

Algorithm 1 illustrates the local heuristic that realizes the online self-regulation model. The heuristic is executed by each household. It consists of a training and testing phase. In the training phase, all possible demand adjustments that satisfy the DR objectives are computed and ranked according to the transition probability metric. The training phase completes with the computation of the regulated state, in case the constraint for a maximum $\mathsf{T}^u_{a\to b} < \theta$ is satisfied. The testing phase checks if the adjustment from the current demand state s^u_l to the forecasted demand state s^u_{l+1} satisfies the DR objective. If the objective is satisfied, no regulation is required, otherwise, the forecasted state is adjusted to the regulated state. Each household is assumed to be equipped with an information system that can translate the forecasted demand state to the regulated demand state [34]. Based on this adjustment, the efficiency of the heuristic can be computed by measuring the AVG reduction, RSD reduction or LF increase, depending on the selected quality feature.

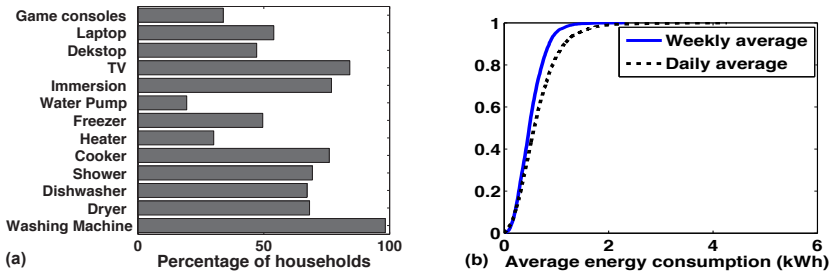


Figure 6.2: Dataset charcteristics.

The self-regulation model is *online* and the training model proposed is adaptive, wherein the temporal metrics are updated after each time period. To regulate the demand, households have to only identify the current and forecasted demand states at each time period.

6.5. RESULTS

This section illustrates the experimental evaluation by employing a dataset [14] of 4,232 residential households to identify target consumers for the DR programs. The performance of the proposed online self-regulation of demand is evaluated empirically.

6

6.5.1. DATASET

CER dataset [14] collected during a smart metering trial in Ireland is used for empirical evaluation. The dataset contains energy consumption measurements from 4,232 households every 30 minutes between July 2009 and December 2010 (75 weeks in total). The objective of the trial was to investigate the effect of feedback on household electricity consumption. Each participating household fills out a questionnaire before and after the trial. The questionnaire contains questions about the socio-economic status of the residential consumer, appliance stock, properties of the dwelling, and the consumption behavior of the occupants. Fig. 6.2(a) shows the distribution of loads across households in the dataset. The x-axis represents the percentage of the households having an appliance. Furthermore, Fig. 6.2(b) shows the distribution of daily and weekly average energy consumption across all households.

Expectation Maximization (EM) clustering [39], [48] is employed to determine number of clusters based on the demand properties. One of the major limitations with clustering algorithms such as k-means clustering is its requirement of prior knowledge on the number of clusters, k. EM clustering iteratively refines an initial clustering model to fit the data based on the principle of maximum likelihood estimation.

The number of clusters found for all the households is 7 and 5 for daily and weekly AVG features, respectively. Similarly, 5 and 4 clusters are found for the RSD feature and, 5 and 5 clusters are found for the LF feature with daily and weekly granularity, respectively. Members of Cluster 1, for the AVG feature indicate households with low average energy consumption. Similarly, members of Cluster 1 for the RSD feature indicate households with low demand variation and members of Cluster 5 for the LF feature indicate households with low demand peaks.

The number of clusters computed with the unsupervised EM approach is validated with two well-known cluster evaluation metrics [101]: Davies-Bouldin Index (DBI) and Silhou-

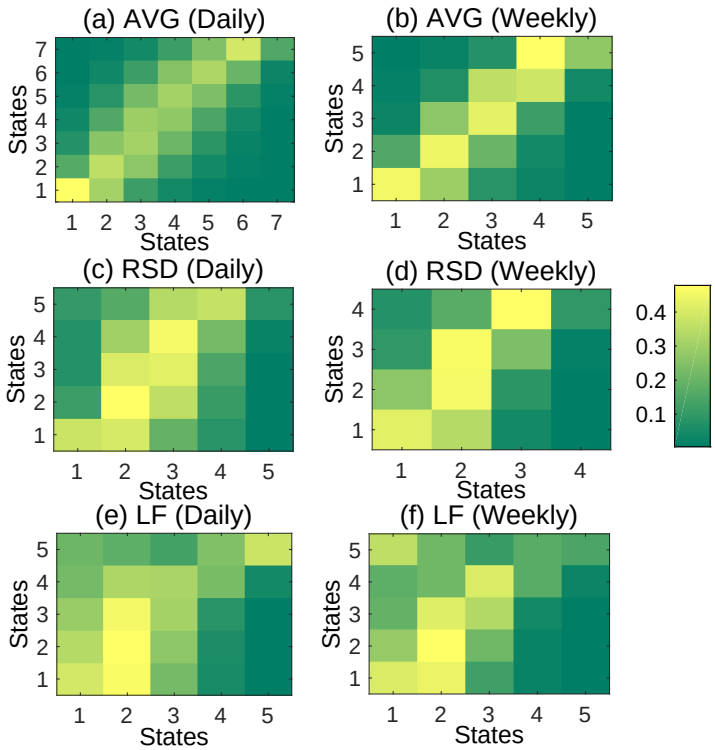


Figure 6.3: Transition probabilities for the quality features – AVG, RSD and LF.

ette. The cluster evaluation metrics verify the number of clusters and confidence of EM method. More details on the cluster evaluation metrics can be found in [84].

Summary: *Clustering identifies the demand state of the households. Cluster evaluation metrics such as DBI and silhouette verify the accuracy of cluster formation.*

6.5.2. TEMPORAL DYNAMICS OF DEMAND

Fig. 6.3 shows the average transition probability of all households for the AVG, RSD, and LF features. The higher the gradient, the higher is the probability of transition from one demand state to another. Households in a certain demand state have higher probability to remain in the same state than transiting to other demand states. This can be seen in Fig. 6.3(a) and 6.3(b), where a household has a high probability to remain in the same demand state, indicating a constant average demand. However, for the RSD and LF features the transitions are more rapid indicating the variations in demand and sudden peaks, respectively. The transition probability from a high RSD state to a low RSD state is low, indicating not so drastic variation in the demand as seen in Fig. 6.3(c) and 6.3(d). Hence, DR programs employed by utilities should consider step-wise reduction matching the variations instead of immediate reduction in demand variation. Fig. 6.3(e) and 6.3(f) show the transition probabilities for the LF feature. The households change their load factor quite often as depicted by the transitions in low demand states.

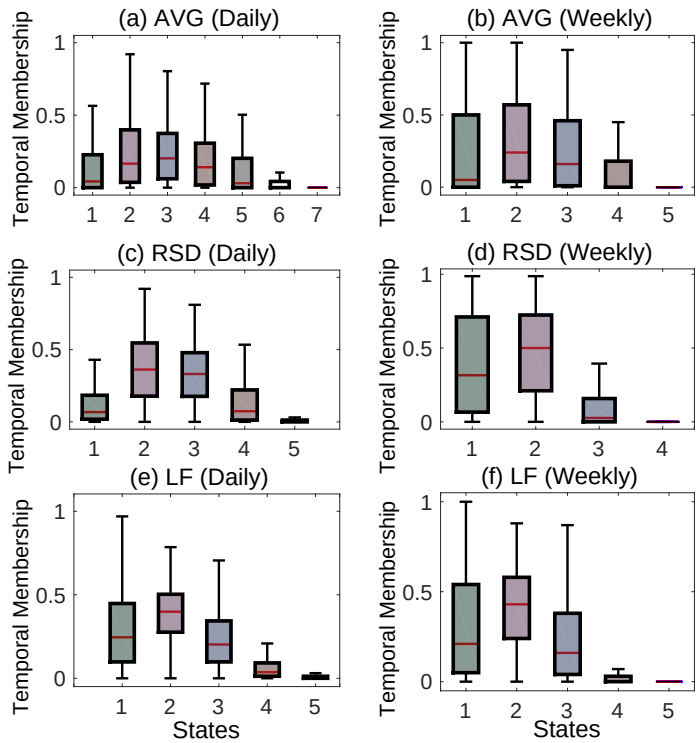


Figure 6.4: Temporal membership for all households.

Summary: *The transition probability illustrates the temporal adjustments of demand. The results show that for the AVG feature transitions are more fixed than the ones of RSD and LF features, where households change their states frequently.*

Fig. 6.4 shows the average temporal membership of all households. The box plots describe the distribution of households for each demand state membership. The lowest line segment indicates the minimum temporal membership value of a household and the top line segment indicates the maximum temporal membership value. The rectangular box indicates the distribution of temporal membership values for different households with the red line segment indicating the median. The majority of the households belong to the intermediate demand states (States 2, 3 and 4) as seen in Fig. 6.4(a) and 6.4(b) for the AVG daily and weekly properties. Indeed, less than 10% of the households belong to low and high AVG state. Temporal membership reveals the most favorable demand state of a household. Utilities can use this information in order to provide tailored recommendations.

Fig. 6.4(c) shows that around 40% of households have high membership probability in state 2 and 3 indicating the majority of the households having moderate demand variations. However, Fig. 6.4(d) shows that around 80% of households have high membership probability in demand state 1 and 2 indicating a low variation in weekly demand. Thus, weekly demand variation of households is more stable compared to the daily variation, which increases the membership probability associated with the weekly properties. Hence, varying the granularity level provides insights on how demand properties change over time.

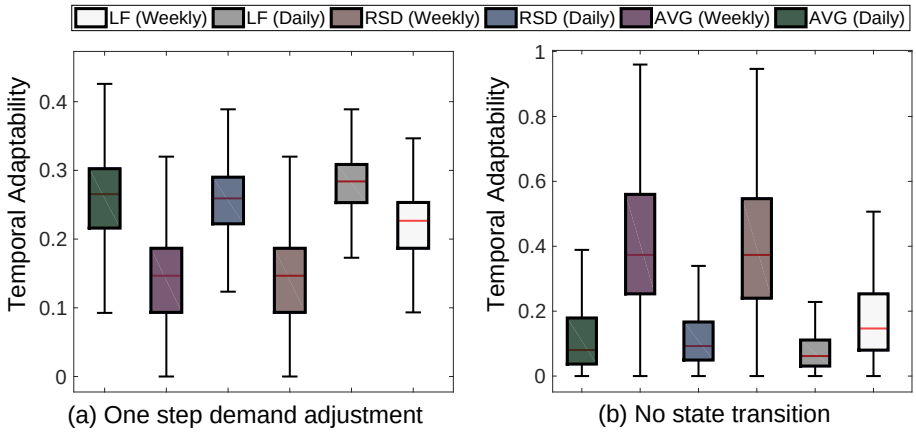


Figure 6.5: Temporal adaptability for all households.

Fig. 6.4(e) and 6.4(f) show the membership of households for the LF feature. The majority of the households are distributed over low demand states, indicating high demand peaks.

Summary: *Temporal membership reveals the most probable demand state of a household. With respect to the AVG feature, only 10% of the households belongs to low demand states indicating that the majority of the households are either moderate or high energy consumers.*

Fig. 6.5(a) show the average temporal adaptability of all households for different quality features with transitions that aim to reduce average energy demand, demand variation and demand peaks. This work considers, (i) one step demand adjustment – transition from one state to another (consecutive or non-consecutive states); (ii) two step demand adjustment – two consecutive transitions from one state to another; and (iii) no transition – self-transitions to the same demand state. An adjustment from a high demand state to a low demand state for the AVG and RSD features indicates the reduction in average demand and variation (e.g., transitions from State 5 to 1 (one step) or State 5 to 3 and then to 1 (two step)). Similarly for the LF feature, demand adjustments from a low LF state to a high LF state indicates reduction in demand peaks. Fig. 6.5(a) shows around 30% of the households can reduce AVG daily demand with one step demand adjustment. For the RSD and LF features, 30% of the households have transitions that can result in reduction of demand variation and demand peaks. The total number of households adaptable for weekly granularity is around 15% for all the quality features. This observation is due to stabilization of demand properties over a week. Fig. 6.5(b) shows the number of households having state transitions to the same state (for example, transition from State 2 to itself). The households containing no/self-transition, indicate the consumers who are not adaptable towards demand regulations. Hence, utilities can use temporal adaptability to identify households that can participate in the DR programs.

Summary: *Temporal adaptability identifies households that are potential target consumers for the DR programs. The results show that around 30% and 15% of the households can participate in the DR programs for daily and weekly granularity, respectively.*

Fig. 6.6 show the average temporal similarity of all households for the AVG, RSD and LF features. Around 25% and 10% of the households have similar demand state transitions for

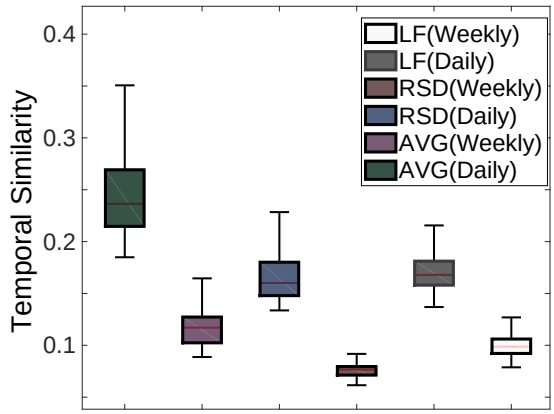


Figure 6.6: Average temporal similarity for all households.

the AVG feature with daily and weekly granularity respectively. Similarly, for the RSD and LF features around 16% and 18% of the households have same transitions for the daily demand. DR programs can use temporal similarity to determine potential households, which have similar demand variation for peak reduction and peak shifting.

Summary: *The results show that, around 25%, 16% and 18% of households have similar demand state transitions among the 4,232 households for daily AVG, RSD and LF features, respectively.*

6.5.3. ONLINE SELF-REGULATION OF DEMAND

The online self-regulation model considers over a year of energy consumption data for the training phase. Since the proposed model is adaptive and online, the duration of training data can be varied. Fig. 6.7 shows the demand regulation for each quality feature with both daily and weekly demand properties. The x-axis represents the threshold value θ indicating the probability of having a demand adjustment that satisfies the DR objective. The y-axis indicates the demand regulation in percentage. The figure also illustrates the number of households participating in the demand regulation.

Fig. 6.7(a) and 6.7(b) show the total energy reduction by all households for daily and weekly AVG demand properties. Each day around 3000 households have demand adjustments that can support energy reduction, resulting in 33% daily average energy reduction (this corresponds to 3.5kW of power) for threshold $\theta = 0.1$. With the increase in θ, the number of households participating in demand reduction decreases. This means that not every household has a demand adjustment with high probability, which can regulate the demand. Moreover, the percentage of energy reduction decreases with the increase in θ. For example, when $\theta > 0.9$, even though around 400 households have state transitions that can regulate the demand, the average energy reduction per day is low. This is because, most of these households selected for $\theta > 0.9$ have low energy consumption. Hence regulating the demand of these households results in low demand reduction. The θ value can be used to select the households, which can participate in demand reduction. Utilities can set a low θ

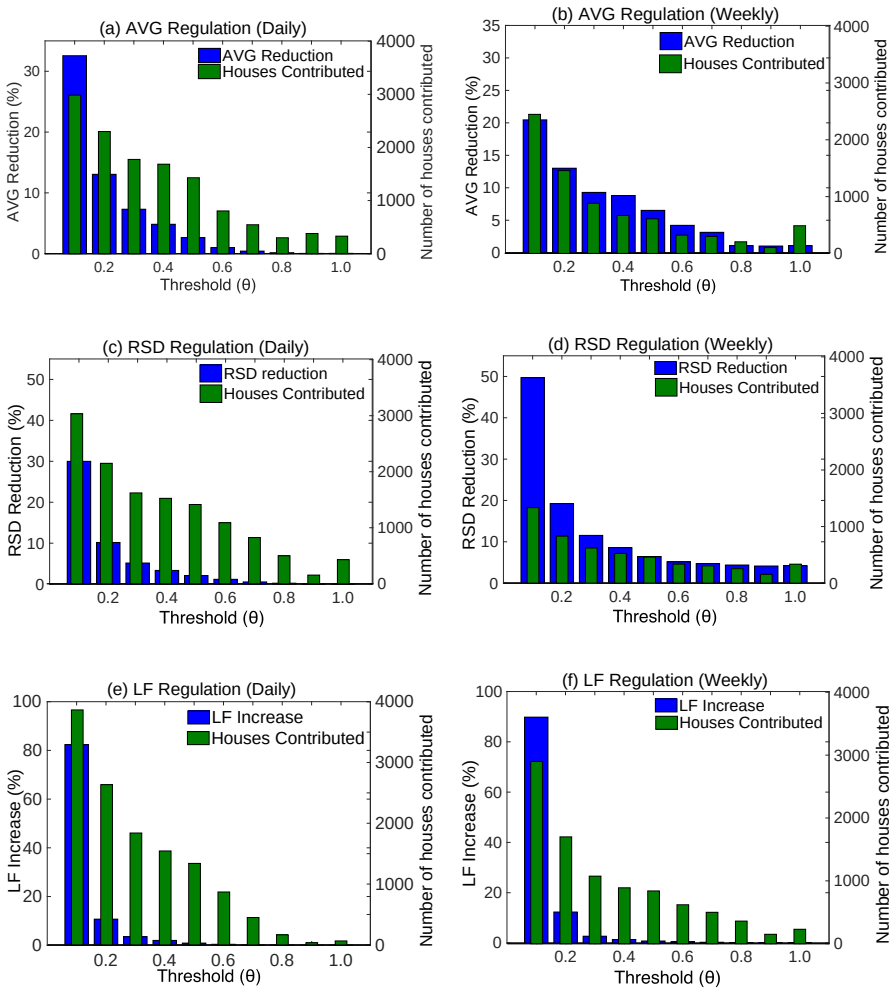


Figure 6.7: AVG, RSD and LF regulation for daily and weekly demand properties.

value during the peak period to select more households for demand regulation and a high θ value during the off-peak period. Fig. 6.7(b) shows the demand reduction for weekly demand properties and it follows a similar trend like daily reduction. For all θ values, reduction of 10% is achieved for daily AVG demand properties.

Fig. 6.7(c) and 6.7(d) show the RSD regulation by all households for daily and weekly demand properties. Demand variation is reduced by 30% and 50% for daily and weekly RSD feature when $\theta = 0.1$. The RSD regulation is higher for the weekly demand than the daily demand. This indicates that households prefer to adjust their demand properties during the week as compared to the daily regulation. For all θ values, the demand variation is reduced by 15% for weekly RSD properties. Fig. 6.7(e) and 6.7(f) show LF regulation by all households for daily and weekly demand properties. Households regulate the LF by reducing the peak demand. Load factor is increased by 80% for both daily and weekly demand properties when $\theta = 0.1$. The number of households participating in demand peak shaving gradually decreases, with the increase in θ. For all θ values, LF increase of 15% is achieved for both daily and weekly demand properties.

The results from the self-regulation model can be used to identify the households that participate in different DR programs. Furthermore, recommendations can be provided to the utilities regarding their DR programs. For example, utilities are encouraged to choose daily AVG demand properties over weekly AVG demand properties for effective demand reduction program. Similarly, for an effective reduction in demand variation, utilities need to select the weekly RSD demand properties over daily RSD demand properties. Utilities can either select daily or weekly LF demand properties for the demand peak shaving as they result in similar LF improvement.

Summary: *The online demand regulation model enables average reduction of 10% in daily average energy demand, 15% in weekly demand variation and 15% in daily demand peak shaving for all θ values.*

Fig. 6.8 illustrates the distribution of households participating in a DR program for all θ values. Fig. 6.8(a) and 6.8(b) show the households that participate either towards (i) reduction in demand (AVG) or (ii) reduction in demand variation (RSD) or (iii) reduction in demand peak (LF). The number of households participating towards demand reduction (AVG) for θ between 0.3 and 0.5 is comparatively higher than for other θ values. This indicates that these households have frequent demand adjustments that regulate the average demand. A large number of households participate in reduction of demand variation when θ is greater than 0.4, indicating that these households have frequent state transitions from low RSD demand state to high RSD demand state. In contrast, more number of households participate in demand peak shaving when θ is lower than 0.5.

Fig. 6.8(c) and 6.8(d) show the households that participate either towards (i) demand reduction (only AVG feature) or (ii) demand variation and demand peak shaving (LF and/or RSD features) or (iii) all the three DR objectives. The number of households participating to all the three features reduce as θ increases and is maximum when $\theta = 0.1$. Utilities can use these insights to choose the appropriate θ value for the selection of households towards the DR program. For example, incentives to consumers with lower values of θ can increase the likelihood of their participation in DR programs. In [14], consumers are incentivized to participate in DR program either based on (i) time of use tariffs, (ii) weekend tariffs and (iii) behavioral change in energy consumption.

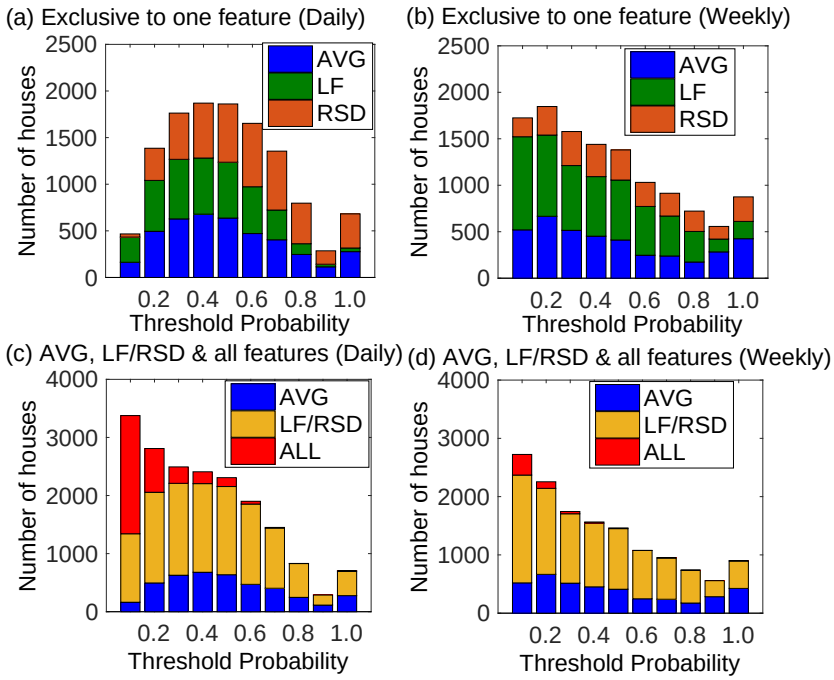


Figure 6.8: Number of consumers participating in DR programs.

Summary: *The results show that the selection of θ plays a crucial role in identifying the target consumers for the DR programs.*

Evaluating the proposed distributed methodology with other related methodologies is a challenge and requires an equivalent context, same dataset and experimental settings. However, we present a constructive empirical comparison with EPOS, the Energy Plan Overlay Self-stabilization system [93]. EPOS is a fully decentralized mechanism for planning and optimizing demand, and employs the same CER [14] dataset for its evaluation. The experimental evaluation settings of EPOS are replicated[1] and compared with the proposed model for different values of θ.

Fig. 6.9 shows the performance of the proposed model against EPOS for θ values in the range 0 and 1. Each colored block indicates the model with the highest performance for the corresponding θ value. The consumer associated with regulation can be managed with a relevant choice of θ. This means that the selection of this parameter is a trade-off and can make the proposed methodology perform higher or lower than other methodologies.

Fig. 6.10 shows the comparison of the proposed online model against EPOS on two specific days viz., 19/01/2010 and 28/05/2010 for all features. Threshold value of $\theta \leq 0.5$ is used to obtain the regulation results. The proposed online self-regulation model has a higher performance than EPOS across all quality features. Demand regulation can be further improved by allowing higher θ values that implies higher consumer tolerance in discomfort. EPOS studies a scenario in which all households participate in the process of demand regu-

[1] Three selection functions of EPOS *viz.*, MIN-Demand, MIN-Relative-Deviations and MAX-Load-Factor are used for comparison. These three functions corresponds to the proposed AVG, RSD and LF regulation.

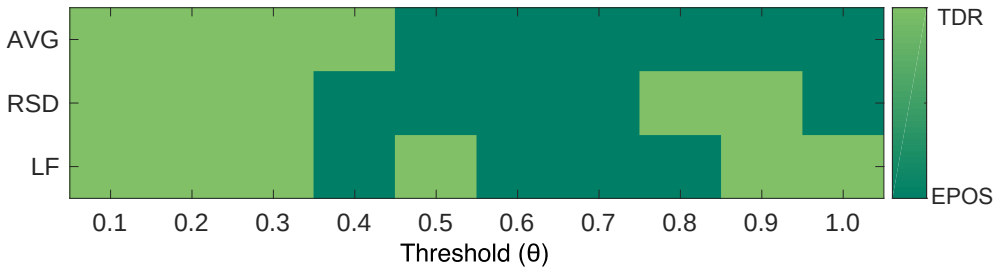


Figure 6.9: Comparison of the proposed TDR against EPOS for θ values in the range 0 and 1.

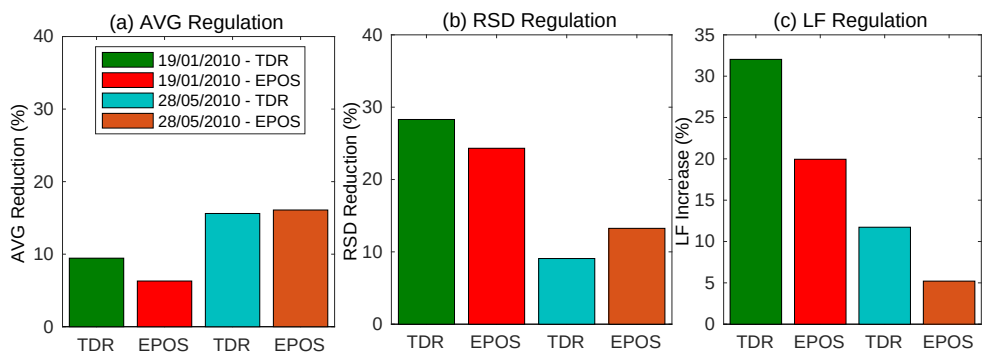


Figure 6.10: Comparison of the proposed model with EPOS.

lation. In contrast, the online self-regulation model identify the households for DR program based on the temporal characteristics of the demand. Consequently, only households that have a valid transitions satisfying the DR objective are selected.

6.5.4. VALIDATION WITH SURVEY DATA

The experimental findings derived are validated with the survey data collected from the trial [14]. Each participant is asked questions regarding the collection of energy data, their attitude towards energy reduction, environment, etc. The objective of this validation is to quantify how close the data-driven analysis is to the survey data. Specifically, the survey responses of consumers are compared with the demand regulation results. The following questions are selected from the survey questionnaire:

- Q1: *I/we am/are interested in changing the way I/we use electricity if it reduces the bill.*
- Q2: *It is too inconvenient to reduce our electricity usage.*
- Q3: *I/we am/are interested in changing the way I/we use electricity if it helps the environment.*

The answers to the above question is in the range [1, 5], where 1 stands for a strong agreement and 5 stands for a strong disagreement. Questions Q1 and Q2 are used to compare the results obtained for the AVG feature and Question Q3 is used to compare results obtained for the RSD and LF features. Consumer survey responses are grouped into two

categories, (i) households which agree towards reduction (survey response: 1,2,3,4) and (ii) households which do not agree towards reduction (survey response: 5). The hypothesis here is that a survey response of strong disagreement (i.e., response 5) means the consumer has no interest towards DR programs. Hence, any other response indicates the willingness towards the DR program. Grouping of consumer survey responses with different combinations is also evaluated, *viz.*, (i) households with survey response (1,2,3) and households with survey response (4,5); and (ii) households with survey response (1,2) and households with survey response (3,4,5). Jaccard similarity coefficient is used to compare the results from the data analysis to the survey results. It is defined as,

$$J(R_s, R_a) = \frac{|R_s \cap R_a|}{|R_s \cup R_a|}, \tag{6.10}$$

where, J(·) is the Jaccard similarity coefficient [113], R_s and R_a are the set of households obtained based on the outcome of the survey response and data analysis respectively. The similarity coefficient is the ratio of intersection and union of these two sets and takes a value [0, 1]. The output of $J(R_s, R_a)$ indicates the percentage of households, which are found both in the survey and analysis results. To evaluate the similarity, the following statistical measures are derived:

- True Positive (TP): The number of households that are present both in survey and analysis. This is similar to Jaccard similarity coefficient.
- False Positive (FP): The number of households that are present in the survey, but are not present in the analysis.
- False Negative (FN): The number of households that are not present in the survey, but are present in the analysis.
- True Negative (TN): The number of households that are not present in both survey and analysis.
- F1-score: The measure of accuracy and is obtained by calculating the harmonic mean of precision and recall [94].

Table 6.2 shows the TP, FP, FN, TN and F1-score for all questions when compared to the data analysis results. The analysis correctly identifies 70% of the consumers in the survey data, who agree with the reduction. FP shows the percentage of consumers who responded positively towards reduction but are not found in the analysis. This observation can be explained by the fact that survey response collected is from only one occupant of a household and this opinion may be different from the other occupants in the household. Similarly, FN indicates the consumers who do not agree towards reduction but are found participating in the analysis. These households could be the potential new target consumers for the utilities. The FN in the dataset for the AVG feature is around 2% (85 households) and for the RSD/LF feature it is around 9% (380 households). Overall, the analysis results are around 87% accurate (F1-score) when compared to the survey data. Due to paucity of space, results from different groupings of consumer survey responses are not shown in detail.[2]

[2]When survey response (1,2,3) are grouped together, the TP for Q1,Q2, Q3-RSD and Q3-LF is 0.64, 0.67, 0.69 and 0.69 respectively. Furthermore, when survey response (1,2) are grouped together, the TP for Q1 is 0.65, Q2 is 0.55, Q3-RSD is 0.65 and Q3-LF is 0.63.

Table 6.2: Comparison of survey data with analysis result.

Questions	TP	FP	FN	TN	F1-score
Q1	0.68	0.12	0.02	0.17	0.90
Q2	0.63	0.20	0.02	0.15	0.85
Q3-RSD	0.69	0.13	0.09	0.09	0.86
Q3-LF	0.68	0.15	0.09	0.08	0.85

Summary: *Validation results show 70% similarity among the consumers identified in the data analysis and survey. New potential target consumers close to 10% are determined for the DR programs, which are not apparent in the survey data.*

6.6. CONCLUSIONS

In this chapter, we presented a promising data-driven methodology for understanding and measuring the temporal dynamics of energy-demand adjustments. Based on this methodology, an online self-regulation model was introduced that identifies consumers who can adjust their demands to meet various DR objectives. Since time-series analysis is used, the approach could be generally applied to any application domain that deals with such data. Experimental evaluation of demand data from real-world Smart Grids shows that around 30% and 15% of the consumers can be incentivized to participate in daily and weekly DR programs. In this case, DR programs achieve 10% reduction in the average daily demand, 15% reduction in the weekly demand variations and 15% reduction in daily peak demand. The data-driven analysis was also validated with the data from the survey.

Several energy services can be designed using the proposed TDR to identify target consumers for DR programs in a neighborhood. In the next chapter, we present techniques that can identify target consumers based on both energy consumption and consumers' social context (beliefs, behavior and ties).

7

TECHNO-SOCIAL SMART GRIDS

IN the previous chapter, we presented energy services that can identify target consumer for various DR programs. However, the effectiveness and adoption of these techniques highly depend on the consumer awareness, their participation and engagement. Prevalent SG deployments and programs have been found to be lacking in consumer awareness and engagement [13]. Hence understanding *what* consumers want and *how they behave* is fundamental for developing a sustainable future-proof energy system. In this chapter, we present a novel techno-social framework that combines both technological aspects of SG and social aspects of consumers to determine target consumers.

Energy utilities need to develop innovative energy services to *understand* how the energy supply is perceived by consumers and *engage* them to actively participate in the functioning of the grid. Current research efforts [64, 112] try to design feedback mechanisms to promote awareness of energy consumption by providing detailed energy breakdowns and appliance-specific energy/cost details. While these mechanisms are necessary and valuable, they are not sufficient to motivate pro-environmental behavior. Hence along with energy-consumption characteristics, consumer preferences (pro-environmental, pro-reduction, pro-behavioral change) and their social context (social -ties, -influences, and -relationships) need to be considered during the development of SG programs.

A growing number of efforts, hitherto, have been applying the principles of social network analysis to home energy management. These mechanisms are aimed at revolutionizing the understanding of energy-usage characteristics and help lessening the impact on the environment[1]. Rich data from social networks capturing human interactions has closed an important loop [52, 71]. This has enabled to model social interactions and to use them in the design of SG services.

Most of the current work does not consider the interplay between the social context of consumers and their energy information in developing consumer-centric services. In contrast, we try to fill the gap by modeling and analyzing the social context of consumers along with the energy network. We propose a techno-social framework to model both technological and social aspects of the SG. The framework determines *how* the social context of consumers can be obtained and *which* social network models can be used along with energy-

[1] C. Chima. http://mashable.com/2011/02/08/smart-grid-social-media/

consumption details to develop consumer-centric services. For example, aspects such as, interests and preferences of consumers can be deduced from social networks through posts, comments, likes, products purchased, etc. These aspects along with consumer energy-usage patterns can be used by utilities to target specific *groups* of consumers in DR programs to either reduce or shift their usage. We also see that the social context of consumers constantly change and evolve, e.g., new friends, change in relationships, preferences, etc. Modeling both the technological and social aspects jointly is a challenging task as the models developed should capture the *dynamics*, not only the energy-consumption patterns but also the social context of consumers.

We propose a ***Techno-Social framework for Smart Grids*** *(TSSG)*, where infrastructure composed of various technologies and social aspects are studied together. A social network overlay is proposed to capture the behavior and preferences of consumers vis-à-vis SG. The framework utilizes traditional social network models [42] to analyze consumers' preferences along with energy-consumption profiles. We illustrate the benefits of modeling the techno-social aspects by forming communities directed towards particular goals. The novelty in formation of communities lies in fusing the technological and social data. These communities can *now* be targeted to promote energy awareness, provide tailored recommendations and community-specific tariff rates. Most of the real-world datasets in smart grids include only energy consumption information of consumers with very minimal user information. While, social network analysis has plenty of user information from various social networking sites, merging these two is non-trivial. Due to lack of real-world datasets containing both the information we choose a dataset with 4000 households, which has energy consumption information and also a survey on consumer preferences. We consider consumer preferences such as pro-environmental, pro-behavioral change, and pro-reduction as key social context information for demonstrating the potential of TSSG. While we use survey to get this information, later we show how consumer preferences can be derived if one had access to consumers social network. This chapter shows the potential of combining technological and social aspects by employing just consumer preferences, we believe it can be extended to provide several novel services if sufficient user data is collected. To the best of our knowledge we are the first to propose the interaction between energy consumption and consumer preferences to determine communities for targeted recommendations.

Contributions. The main contributions of this Chapter are:

- We propose a novel techno-social framework for smart grids (TSSG) to enable effective energy -coordination, -management and -awareness amongst consumers.
- We describe the advantages of having a social network overlay for different SG programs to engage active participation of consumers.
- We present a case study of goal-oriented community formation by fusing the techno-social data from a real SG deployment with more than 4,000 households.

7.1. TSSG FRAMEWORK

The techno-social framework integrates physical (energy network), cyber (ICT network) and social dimensions of the SG (see Fig. 7.1).

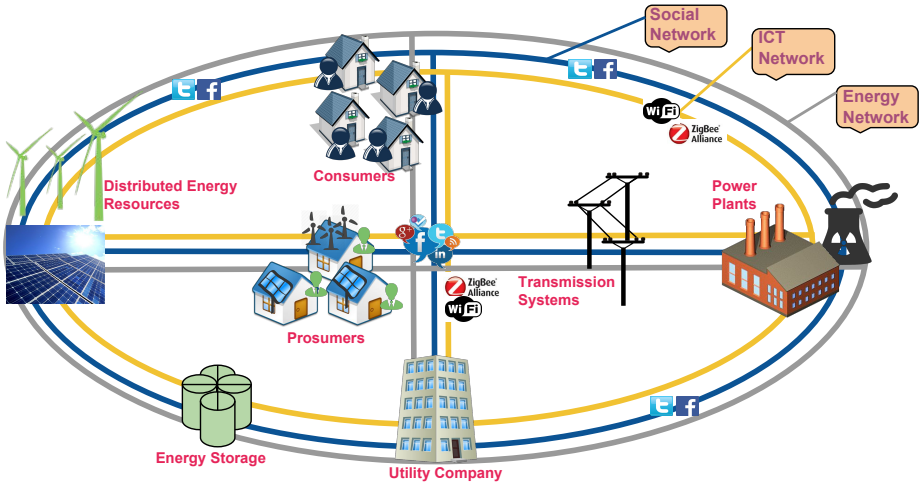


Figure 7.1: Techno-social framework of the SG with its entities.

7.1.1. OVERVIEW

The proposed TSSG framework consists of three layers *viz.*, *physical grid*, *smart grid*, and *social grid* (see Fig. 7.2).

Physical grid is an interconnected network for delivering electricity to consumers. The physical grid consists of several power generation stations, power plants, high-voltage transmission lines to carry power to the consumers and distribution lines to interconnect various entities of the grid. The different actors in physical grid include transmission system operators (TSO) and distribution system operators (DSO) for operating, maintaining and developing the grid.

Smart grid is an intelligent power system that uses ICT to enhance efficiency, reliability and sustainability of power generation and distribution network [45]. SG utilize several mechanisms to curtail and balance load, flatten peak demands, automate load control, apply adaptive pricing, and bring awareness to its consumers.

Social grid promotes interactions among different SG entities to support coordination of energy consumption, to trade energy between prosumers and consumers, and to promote awareness. With the increased growth in deployment of embedded sensors in the environment, the social grid can now accurately monitor consumers in different dimensions [25, 71]. In social grid, third-party services such as Facebook, Twitter, LinkedIn, and Google, along with ubiquitous sensors play a crucial role in collecting consumer social activities.

TSSG framework supports development of several consumer-centric services like personalized recommendation, forming communities with similar consumer characteristics and energy profiles, etc., by modeling and analyzing the data from the techno-social ecosystem. The introduction of social overlay on SG supports two-way information flow: (i) TSSG derives information about consumer norms, preferences, ties and interactions to understand the perception of consumers regarding the energy pricing, demand reduction, etc.; and (ii) TSSG analyzes these preferences to derive best practices towards sustainable behavior and disseminates them using the social overlay. Moreover, it is easy to see that this feedback itself could be used to gather more inference and to understand whether con-

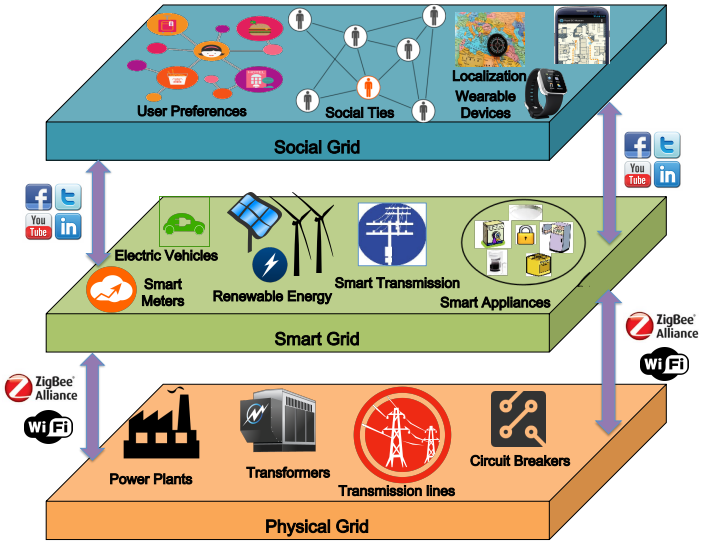


Figure 7.2: Techno-Social framework for SG.

sumers followed certain practices. Further, it is easy to gather which feedback information was useful or followed. This helps in further improvement of information dissemination. Furthermore, the proposed TSSG framework is decentralized, i.e., it is implemented at a neighborhood/community level, where techno-social data from SG and social grid are available.

7

7.1.2. CORE COMPONENTS

To enable interactions between the SG and social grid layer, several challenges need to be addressed in gathering and modeling both energy and consumer data. Some of these challenges are: (i) processing the raw, fragmented, and unstructured data collected from the techno-social ecosystem; (ii) analyzing and modeling multi-dimensional data from different data sources; (iii) adapting to changing and evolving social context (preferences, relationships, and ties) of consumers over time; (iv) striking a trade-off between quality of collected data and privacy; (v) deducing aggregated behavior of consumers to promote collective awareness; and (vi) adjusting to temporal dynamics of the techno-social data with different resolution (hourly, daily, monthly, seasonal). To address these challenges, we define two core components *viz., techno-social sensing* and *mining*. Fig. 7.3 shows the core components of the TSSG framework and their interactions towards development of consumer-centric services.

Techno-social sensing: Techno-social sensing is responsible for collectively harvesting data from the SG and social grid layers. From the *smart grid layer* data regarding energy consumption at households, appliance-specific consumption, power factor, voltage and current are collected. From the *social grid layer* data regarding consumer activities, their social ties, behavior, interactions, preferences and opinion are collected with the help of wearable devices, smartphones, RFID tags, online social networks, GPS logs, and IoT sensors embedded in the environment [83]. Furthermore, several participatory sensing and social

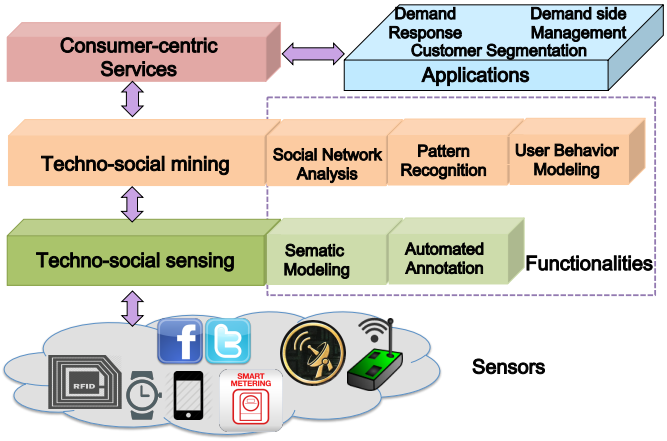


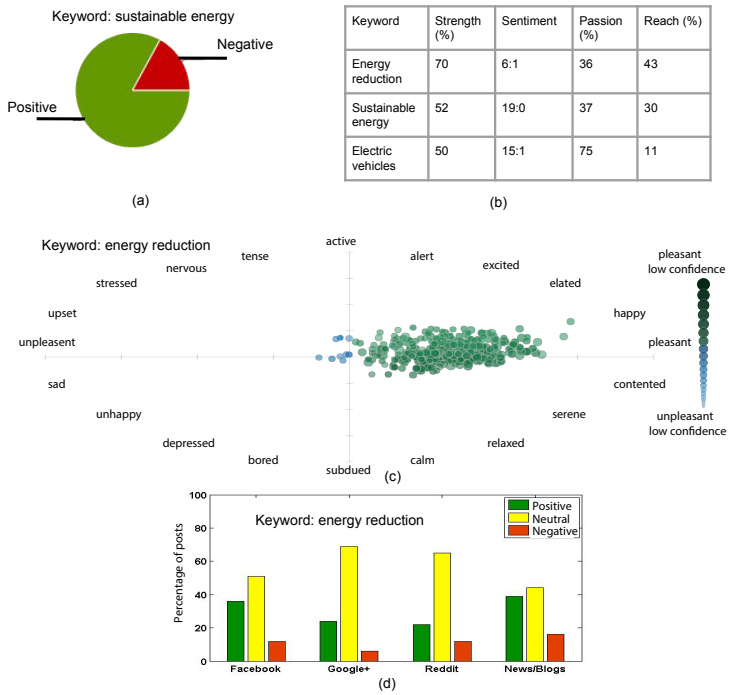
Figure 7.3: Interactions of core components of techno-social SG.

network API's can now be used to search and gather information about consumers. For example, Social media such as Facebook, Twitter, Google+, Pinterest and Youtube, has wealth of information on preferences of consumers. Data collection from these networks is now possible with the help of open-source API[2].

Techno-social mining: Techno-social mining aims to develop models to understand consumer behavior by identifying underlying patterns, rules and beliefs from the data collected by techno-social sensing. At the *smart grid layer*, user-specific energy consumption profiles, appliance energy profiles, cost-aware appliance usage schedule, load shifting strategies can be derived by applying different pattern recognition and machine learning algorithms. At the *social grid layer*, consumer behavior, their beliefs, and social ties can be derived by mining consumers' content on social media using social network analysis. Many ways have been proposed for social mining, e.g., social network analysis, social media mining, and sentiment analysis. Recently, *sentiment analysis* is gaining popularity to analyze consumer sentiments. For example, sentiment analysis on consumers' posts on Facebook and Twitter can help in determining the sentiment of user towards a technology (electric vehicles, energy reduction, sustainable energy usage, etc.). We now illustrate how consumer preferences on some of the energy related topics can be derived using sentiment analysis across social networks. Sentiment analysis aims to identify and extract subjective information from data shared across social networks. We utilize open source sentiment analysis mechanisms to determine positive or negative sentiments with respect to energy related topics across various social networking sites. Fig. 7.4 shows the sentiments of several consumers (more than 1000) over a week towards green energy campaigns.

Fig. 7.4(a) shows the percentage of positive and negative sentiments associated with "sustainable energy" on Twitter for a week. Among the tweets related to sustainable energy, 85% of tweets had positive sentiment and 15% had negative sentiment. Similarly, Fig. 7.4(b) shows the strength, sentiment, passion, and reach for three keywords *viz.*, energy reduction, sustainable energy, and electric vehicles. Electric vehicles have strength of 50% – likelihood of it being discussed in social media; Sentiment of 15:1 – ratio of positive to negative sen-

[2]Social Media API's http://www.programmableweb.com/category/social/apis?category=20087



Keyword	Strength (%)	Sentiment	Passion (%)	Reach (%)
Energy reduction	70	6:1	36	43
Sustainable energy	52	19:0	37	30
Electric vehicles	50	15:1	75	11

Figure 7.4: Sentiments of several consumers over a week towards various energy campaigns.

timents; passion of 75% – likelihood of repetition of the topic; and, reach of 11% – range of influence of the topic in the social media. Fig. 7.4(c) shows the classification of tweets corresponding to various sentiments associated with energy reduction on Twitter. Here a tweet is classified into one of the 20 categories derived from social behavior models. Fig. 7.4(d) shows the percentage of posts with positive, neutral and negative sentiments across social networking sites such as Facebook, Google+, Reddit and news/blogs for energy reduction. Several metrics have been proposed in the literature to derive accurate and reliable sentiments over several days. Techno-social mining uses these to derive preferences of consumers, which can be employed to develop consumer-centric energy services.

7.2. ROLE OF TSSG IN CONSUMER-CENTRIC SERVICES

This section describes how TSSG can be employed towards developing consumer-centric SG programs.

1. Consumer awareness: Smart meter data when analyzed effectively can help consumers to reduce both energy consumption and cost. In-house displays, interactive games, and smartphone applications are being proposed to increase awareness amongst consumers. However recent research efforts show that, this approach fades away after a few initial weeks of deployment [64]. With constant increase in number of consumers using social networks (SN), information propagation on SN is a promising approach to seek responses from consumers; interacting with them and also to disseminate information [25]. The proposed TSSG framework enables smart meters to directly connect to social networks and provide

energy consumption information to online social networks. The role of TSSG is to provide the energy consumption information to consumers using SN in an unobtrusive way. Furthermore, SN models can be utilized to analyze who is the most influential consumer, how information spreads among consumers and consumer-interest based groups in SN.

2. Consumer coordination: The coordination of energy consumption among consumers in a neighborhood helps in balancing the temporal energy usage. Current approaches for active coordination require a central coordinator and prior agreements among consumers. TSSG framework can support completely distributed coordination with the help of SN overlay on SG. TSSG identifies and enables interactions among consumers with similar interests to balance aggregate consumption. These consumers can coordinate among themselves to address their energy demands over SN. Consequently other consumers can schedule their demands such that total energy consumed may not exceed the generated energy. Identifying such a group of consumers is highly challenging due to the evolving nature of social context. TSSG utilizes the data from techno-social ecosystem to derive consumer groups that can effectively coordinate and support energy management programs.

3. Energy trading: With the penetration of electric vehicles and renewable energy sources such as wind turbines, solar panels in residential settings, households not only consume but also produce their own energy ("prosumers"). Concepts such as, vehicle to home, vehicle to grid and energy trading requires active coordination among consumers to enable direct selling of excess energy. Currently there exists no secure, privacy-aware, market place to enable real-time coordination for energy traders. TSSG supports energy trading by allowing consumers to coordinate on SN. Specifically, TSSG aims to enable trading of energy by mining the techno-social data where consumers with similar social contexts and beliefs are allowed to negotiate.

4. Goal-oriented communities: Heterogeneity in consumer characteristics hinders identifying target consumers for specific SG programs. Hitherto, utilities used methods based on sociology, psychology, and behavioral economics to determine target consumers for DR programs. These efforts mainly considered data reported by consumers and did not model the dynamics of energy consumption and social contexts. TSSG supports identification of consumers by deriving *virtual communities* – group of consumers such that intra-community associations are denser than inter-community association. The associations could be derived based on various features explained earlier. Based on the requirement utilities can identify the features and consequently forming communities, we call this as "goal-oriented communities". The goal could be identifying consumers who are pro-environmental, pro-behavioral change or pro-energy reduction. Consequently, utilities can devise different incentives such as community-specific recommendations and tariffs. TSSG supports formation of these communities by applying community detection algorithms based on correlation among energy consumption, social ties and relationships.

The key requirement in determining the techno-social mining techniques that could be used in TSSG, depends on the objective of the SG program. For example, to promote energy awareness and sustainability in energy usage, we need to understand consumer behavior, their preferences and also their energy consumption pattern. Several models, for example, Theory of planned behavior, Health belief model, Social practice theory, and Diffusion of innovation theory [111] exist to understand and model behavior changes. However, Social practice theory (SPT) is increasingly being applied to analyze behavior in the context of energy management, transportation and waste management.

The principle of SPT is that behavior of consumers (vis-à-vis energy consumption) arises from the interactions between three components [111]: (i) norms – individual and shared expectations on comfort levels, social aspirations, etc.; (ii) material culture – physical aspects of a home i.e., building type, heating devices, energy-related technologies; and (iii) energy practices – actions of and processes used by consumers i.e., temperature settings, maintenance of technologies, etc. SPT argues that the focus should not be on individual behavior, rather, (a) on social practices to understand why certain practices are performed; (b) how and why others are prevented from carrying out some tasks; and (c) the evolution of technology with respect to societal behavior. Furthermore, behavior change is most likely due to careful scrutiny of the norms and practices, and then the promotion of the best practices.

Recently, a framework for energy cultures [111] is proposed to understand energy consumption behavior of consumers. This framework utilizes the basic components from SPT and adapts them to understand energy consumption behavior of consumers. For example, space-heating inefficiencies might be the result of ineffective heating technologies (material culture) or inappropriate heat settings (practices) or unrealistic expectations about warmth (norms). The combination of norms, material culture and energy practices can create self-reinforcing habitual patterns. Achieving behavioral change involves altering one or more of these components, noting that a change in one will almost inevitably lead to change in the others.

TSSG extends the current social science models and can provide feedback to the consumers on the best practices to support sustainability. Specifically in TSSG, (i) norms about consumers are derived from techno-social sensing, i.e., deriving consumer preferences and values from social networks such as Facebook, LinkedIn and Twitter as described earlier; and (ii) material culture and current practices can be derived from electricity consumption pattern of the households. Furthermore, TSSG also supports integration of other models from social sciences to analyze behavior of consumers and promote sustainability.

7.3. ILLUSTRATION: GOAL-ORIENTED COMMUNITIES

In TSSG, Social Networks (SN) help us to derive consumer preferences, which can be used in designing better energy management services. Some SG services may be location dependent i.e., energy sharing between neighboring houses, and energy shifting in a neighborhood. However, energy reduction and pro-environmental energy usage apply to a larger group of consumers irrespective of their locations. To show the benefits of collecting preferences of consumers, we develop new goal-oriented virtual communities to promote energy awareness and provide tailored recommendations and community specific tariff-rates. Utilities currently employ techniques that offer the same information to all consumers – such as how to reduce cost and new energy policies – to its entire consumer base irrespective of their preferences and energy consumption profiles. This information may not be valid or mostly redundant to some consumers. Forming goal-oriented communities enables tailored feedback and tariff-rates to consumers with similar preferences and interests. Virtual communities formed by analyzing both the social and energy contexts help in bringing effectiveness in the campaign.

We employ the CER dataset [12] collected during a trial in Ireland. The dataset contains energy consumption measurements from 4,232 households every 30 minutes between July 2009 and December 2010. The objective of the trial was to investigate the effect of feedback

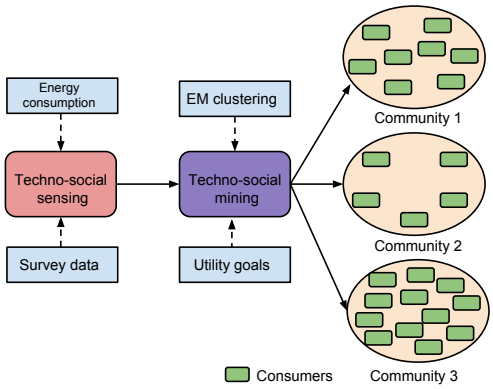


Figure 7.5: Block diagram of goal-oriented community formation.

on energy consumption in households. Each participating household filled a questionnaire before and after the trial. The questionnaire contained questions about the socio-economic status of the consumers, appliances, properties of the dwelling, and the consumption behavior of the occupants.

Fig. 7.5 shows the block diagram of goal-oriented community formation using TSSG core components. The techno-social sensing gathers data about average energy consumption using the smart meter (SM) and consumer preferences, beliefs, opinion and interests using the survey data collected during the trial. The techno-social mining component tries to find structures in the sensed data to detect communities that match the goals set by the utilities. The community detection is performed by applying *unsupervised clustering* technique called Expectation-Maximization (EM) clustering [39]. The EM clustering applies Maximum Likelihood Estimation to determine the optimal number of communities. The techno-social data is used by EM clustering to determine different communities based on the goals defined by the utilities. This article studies three goal-oriented community formation *viz.*, (i) pro-behavioral change, (ii) pro-environment and (iii) pro-energy reduction. Thus utilities now can devise tailored recommendations to the communities formed based on these goals.

Fig. 7.6(a) shows the communities formed by traditional approaches, which utilize only average energy consumption of the consumers. Each community here indicates the group of consumers who have similar average energy consumption. Utilities currently use this mechanism to provide incentives and feedback to each community to reduce or change their usage patterns. However, these recommendations are not tailored based on consumer preferences, which may result in lesser adoption rate of SG programs. Hence, TSSG framework considers data from both energy network and social context of consumers to provide tailored recommendations and to promote awareness. In this article, we deduce consumer preferences, opinion and interests from the survey data. Note that, social context of consumers from other sources such as online SN, mobile phones, and other sensors embedded in the environment can also be utilized. The questions considered from the survey are:

1. Pro-behavioral change: Am/Are I/we interested in changing the way we use electricity if it results in reduction of the bill?

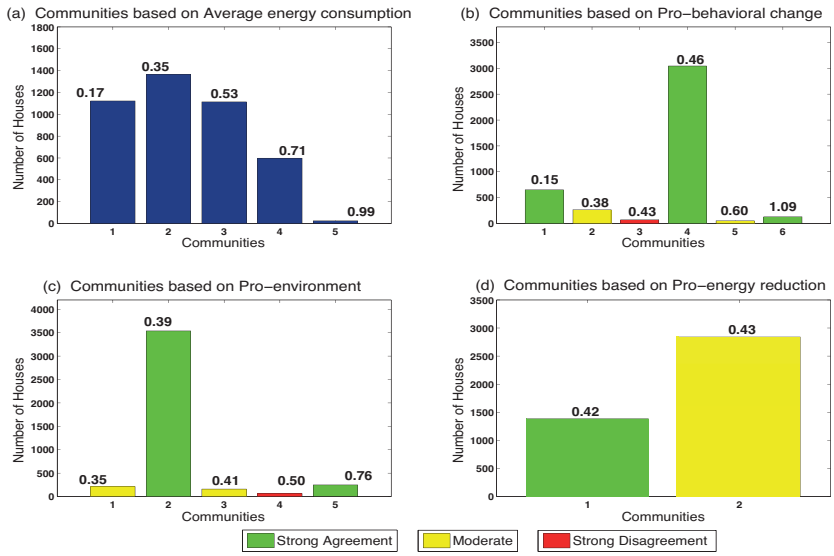


Figure 7.6: Different goal-oriented community formation strategies. (Numbers on the bar indicate average energy consumed per household in kWh)

2. Pro-environment: Am/Are I/we interested in changing the way we use electricity if it helps the environment?

3. Pro-energy reduction: Is it too inconvenient to reduce our usage of electricity?[3].

The answers to the above questions are in the range [1, 5], where 1 means strongly agree and 5 means strongly disagree. The techno-social mining, i.e., EM clustering uses both the average energy consumption and social context data of consumers to derive accurate goal-oriented communities. Fig. 7.6(b), (c), and (d) shows the various communities formed based on the above questions and average energy consumed.

Fig. 7.6(b) shows *six* communities formed based on consumers who are "***pro-behavioral change***". Mean average energy consumed in each community is shown over each bar and the bar color shows the average response by all the consumers in that community. It can be seen that, communities 1, 4 and 6 are more towards behavioral change. Communities 2 and 5 are moderate towards change in their behavior and Community-3 strongly disagrees with changing the usage pattern to reduce bills. Even though consumers in communities 1, 4, and 6 have similar interests their average energy consumption are different. Hence, utilities can now target each of these communities separately with tailored recommendations, feedback and tariff-rates. For example, utilities can target Community-6, whose members are open to behavioral change and consume high energy to change/modify their energy usage pattern. These recommendations and feedback have high potential to reduce consumption. Moreover, the same recommendations may not be applicable to Community-1 as their energy consumption is low, even though they have similar interests as Community-6. Thus TSSG supports utilities to segment consumers to promote sustainable energy usage.

[3] Questions 1, 2 and 3 corresponds to questions 4332, 4333, and 4352, respectively in the residential pre-trial survey.

Similarly, Fig. 7.6(c) shows *five* communities formed based on consumers who are "***pro-environment***". Members of communities 2 and 5 strongly agree with change in usage behavior to help the environment. Communities 1 and 3 are moderate towards environmental impact and members of Community-4 do not worry about the environment. It is interesting to see that, even though average energy consumed in Community-2 is lower than communities 3 and 4, members of Community-2 are more environment-friendly and agree to change energy usage pattern. Members of Community-5 have the highest average energy consumption and are willing to change their usage patterns to become environment-friendly. This can be used by utilities to target members of Community-5 by providing recommendations on different usage patterns to reduce consumption.

Fig. 7.6(d) shows the communities formed based on consumers who are "***pro-energy reduction***". Only two communities are formed based on the answer and the average energy consumption. Both the communities have similar average energy consumption. However members of Community-1 strongly think there is no inconvenience to reduce their energy usage, unlike members of Community-2 who are moderate towards reduction in energy usage. In total, 33% of 4232 consumers think that there is no inconvenience in reducing energy usage and the rest 67% are moderate towards reduction in energy usage. Utilities can provide different tariff-rates to these communities as incentives to reduce their average energy consumption.

Deriving communities with consumers having various levels of social activities is one of the open problems that need to be addressed to guarantee fairness. Recent work by Muhammad et. al [121], proposes several fairness constraints to adapt the clustering and classification algorithms to consider consumers with varying levels of participation in SN. TSSG can incorporate these constraints during community formation to guarantee fairness among consumers with varying levels of social activity. Additionally, data from surveys, face-to-face interviews, and campaigns can be used to guarantee fairness. Furthermore, our community formation strategy can be applied to DR algorithms such as load shifting, load reduction and energy sharing, by using data from physical grid and consumers location.

7.4. CONCLUSIONS

The development of sustainable energy services depends heavily on active participation and engagement of consumers. We proposed a novel Techno-Social framework for Smart Grids (TSSG) where infrastructure composed of various SG technologies interacts with social activities of consumers. The TSSG framework uses traditional social network models to analyze consumer behavior along with energy-consumption information to develop consumer-centric services. The role of TSSG towards various SG services and its benefits were described. Furthermore, we illustrated goal-oriented community formation with data from more than 4,000 households. We also showed how groups of consumers can be targeted differently by considering the heterogeneity in consumer preferences and their energy consumption.

To the best of our knowledge, TSSG is the first framework that employs a holistic view, inclusive of consumers, prosumers, utilities and SG infrastructure. We believe the integration of technological and social aspects lead to the development of sustainable, smart energy systems.

8

DATA PROCESSING ARCHITECTURES

IN the previous chapters, we discussed several personalized energy services that aim to improve user comfort in smart spaces – households, buildings and neighborhood. In this chapter, we explore how smart-meter data can be stored and processed efficiently to support these energy services at city level. While the deployment of smart meters is growing, the lack of adoption of energy services has hindered large-scale smart-grid (SG) deployments. A sustainable, smart energy system should support handling huge amounts of smart-meter data from millions of consumers, protect privacy-sensitive data of consumers, allow designing of energy services at various scales and provide guidelines for implementing a large-scale system for DR programs.

This chapter explores how to cope with the overwhelming data generated from smart meters towards design and development of sustainable, smart energy systems. An estimate from a utility provider indicates 22 gigabytes of data being generated every day from its 2 million customers [108]. The overwhelming data generated by smart meters calls for developing information management mechanisms for large-scale data storage and processing. While there have been deployments of SG (e.g., Grid4EU[1] and SmartWatts[2]) with a few participants, the design of a suitable architecture to support envisaged SG services on a large scale is an important research topic currently [124], [100].

In this chapter, we investigate the cost-benefit analysis of four data processing architectures for various energy services. We introduce several key cost indicators to analyze hierarchical data processing architectures for SG. In our evaluation, we consider realistic deployments for both dense (i.e., urban areas with $1.6M$ households) and sparse (i.e., rural areas with $476K$ households) environments.

Contributions. The main contributions of this chapter are:

- We model different data-processing architectures (centralized, decentralized, distributed and hybrid) for hierarchical power-distribution networks in both urban and rural environments.

[1] http://www.grid4eu.eu/
[2] http://www.smartwatts.de/

- We introduce and model several key cost indicators, such as energy consumption, processing power, storage requirements, communication bandwidth, accuracy and privacy.
- We provide a detailed cost-benefit analysis of the proposed architectures, which SG designers can use to select the architecture that best fits their requirements.

8.1. RELATED WORK

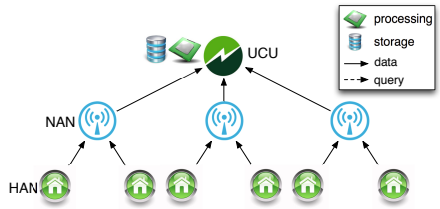
The majority of architectures proposed hitherto focus only on a specific architectural aspect, like communication, storage or processing. In recent SG deployments, smart meters collect data at an interval of 5 to 15 min compared to traditional systems that only records meter data once a month [10]. Data values obtained as an average over a 15 minute interval may not be detailed enough to support energy services such as advanced distribution automation, asset management, and appliance energy disaggregation [85]. Current, smart meters in the near future may well measure values every 30 s, posing a significant challenge in processing and storage of huge amount of data generated.

A secure decentralized data-centric information infrastructure for the SG is proposed in [60]. Kim et al., describe challenges in low-latency communication protocols, and security mechanisms for SG. Balancing supply and demand is mapped to a constraint-satisfaction problem and evaluated using a decentralized hierarchical architecture in [100]. A cloud-based SG information management model is proposed in [103] and [46], along with a discussion on key challenges. The focus on cloud-computing approaches is to provide adequate resources for the SG. *In contrast with the above works, we not only propose and analyze several architectures, but we also model important key cost indicators such as energy, communication, storage, processing, accuracy and privacy based on the physical topology of the grid.* The cloud-based Demand Response (CDR) architecture using a distributed information infrastructure is proposed in [67]. Scalability aspects of data storage and processing of monthly bills in SG is investigated in [21]. Several data-storage mechanisms like centralized relational databases, distributed relational databases and file systems are compared and evaluated. Similarly, scalability aspects of data communication for smart metering in SG is investigated in [124]. They also study how communication cost scales with the number of smart meters and sampling frequency.

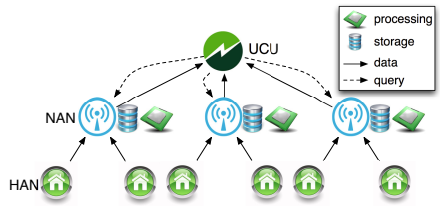
A comparison of centralized and distributed monitoring architectures for billing and demand-response applications is proposed in [79]. Martinez et al. explore the potential benefits of having distributed architectures compared with centralized ones. In [79], the authors evaluate the proposed architectures by considering a fixed number of houses. In comparison, our work improves on the existing literature to provide a comprehensive analysis of various data-processing architectures with realistic environments *viz,* urban and rural environments. We consider accuracy and privacy cost indicators to provide a detailed cost-benefit analysis along with other cost indicators like energy, communication, storage, and processing.

8.2. DATA PROCESSING ARCHITECTURES

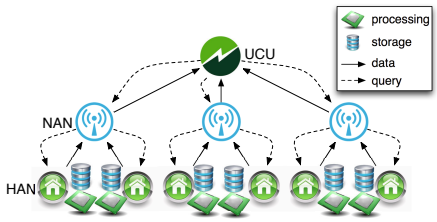
The current topology of the power distribution network is arranged according to the voltage [20]. The distribution network is organized into multiple subgrids and consequently forming a hierarchical topology. Our architectural model adopts hierarchical topology of



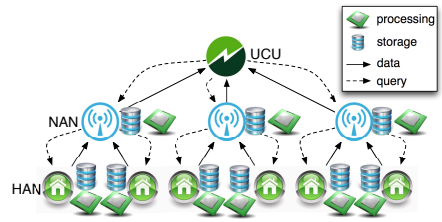
(a) Centralized Architecture



(b) Decentralized Architecture



(c) Distributed Architecture



(d) Hybrid Architecture

Figure 8.1: Data processing architectures for the smart grids.

the power distribution networks. The key elements of our architectural model are the following.

Home Area Nodes (HANs)- are devices interconnected with the smart meter at the consumer premises. HAN receives energy-consumption information from all appliances in the household and can employ mechanisms or receive information to match supply and demand at the household level.

Neighborhood Area Nodes (NANs)- act as an intermediate node between consumers and utility providers, and it serves a small geographical area, i.e., a neighborhood consisting of several houses. NAN receives information from the households within the neighborhood. Multiple NANs are deployed to cover utility's territory.

Utility Control Unit (UCU)- is the central control entity of utility providers. This node is responsible for billing, maintaining data, determining electricity price and carrying out demand response. UCU acts as the root node in our architectural model.

Design Choices.

In our architectural model, the HANs at each household periodically sense energy consumption and transmit to the respective NAN. NANs act as the intermediate nodes between HANs and UCU. Communication between HANs and NANs is based on $sub-1GHz$ transceivers which are best suited for both indoor and outdoor environments [22]. The interconnection between NANs and UCU is based on IEEE 802.16 (WiMAX) which supports a maximum data rate up to 1 $Gbps$. The communication choice is supported by some of the recent works [124], [20].

Data aggregation at the nodes can minimize the overall data communicated and also help in preserving sensitive information of customers. Three major data aggregation mechanisms considered in literature are:

a) *Time-wise aggregation*: where consecutive time-stamped energy consumption readings are aggregated to reduce the granularity of the data collected.

b) *Value-wise aggregation*: where similar energy consumption readings are bucketed to ob-

tain discrete energy readings and thus reducing number of readings.
c) *Consumer-wise aggregation*: where the energy consumption values of several individual customers are aggregated into one time series to obfuscate the consumption of each individual customer.

We consider only time-wise data aggregation with different granularity. UCU can acquire information from the households by initiating a query and nodes can respond to the query depending on their roles. Based on the data aggregation, processing and storage capabilities of HANs, NANs and UCU, different architectures are proposed. By default, all nodes can send and receive a message, which is the minimum capability assumed at each node. The storage and processing icons in the Fig. 8.1, show the additional capability available at each node depending on the architecture.

Centralized Architecture.

In centralized architectures as illustrated in Fig. 8.1a, only the UCU has data processing and storage capability. HANs periodically sense and transmit the energy consumption values to the respective NANs. NANs act as relays and forward it to the UCU. UCU has all the information and is responsible for processing and storage of the data. Thus, the information flow is uni-directional from HANs to UCU via NANs. No data aggregation is applied in centralized architecture.

Decentralized Architecture.

In this architecture, only NANs have data processing and storage capabilities as shown in Fig. 8.1b. HANs transmit data periodically to the respective NAN similar to the centralized architecture. Instead of forwarding the data, NAN stores and processes this data locally. In decentralized architectures, since all data is available at the NAN, fine grained data aggregation is possible. NANs can aggregate hourly energy consumption and report to UCU. UCU generates queries to retrieve information from the NANs only when required. Thus, NANs act as central entities in this architecture.

Distributed Architecture.

In distributed architectures, all HANs have data processing and storage capabilities. HANs periodically sense and store the energy consumption values locally. UCU initiates a query to fetch data, which is forwarded to the NAN and in turn to the HANs. HANs process the query and send the reply to UCU via NAN. Thus, making the architecture completely distributed as illustrated in Fig. 8.1c. HANs are assumed to have sufficient data storage and processing capability and communicate only upon reception of a query.

Hybrid Architecture.

In hybrid architectures, HANs and NANs both have data processing and storage capabilities as shown in Fig. 8.1d. Hybrid architectures are extension of distributed architectures, where HANs not only sense and store but also transmit aggregated energy values to NANs. For instance, HANs can sense and store energy values periodically and at the end of the day send an aggregate energy consumption reading to the NAN. The data aggregation granularity may vary depending upon the energy services considered.

8.3. KEY COST INDICATORS

Architectures proposed in the previous section are characterized based on the availability of data storage and processing capabilities at the node, hence monetary costs of a node needs to be modeled in architecture evaluation. Data aggregation is employed by architectures to reduce the amount of data communicated as well as increasing the privacy of the

customer. However, data aggregation results in decreased accuracy, since the UCU might need to disaggregate energy readings that have been aggregated over time by the HAN or the NAN. The trade-off between accuracy and privacy cost as a function of data aggregation granularity provides a key insight in the design of architectures. Our cost-benefit analysis hence considers monetary cost, accuracy and privacy as the key cost indicators to evaluate the performance of proposed architectures.

Monetary cost: Monetary cost C_M is the cost (in \$) to deploy and operate the nodes in the architecture (HANs, NANs, UCU).

$$C_M = C_D + C_O. \tag{8.1}$$

The deployment cost C_D is a one-time cost that accounts for the deployment of storage, processing and communication capacity.

$$C_D = C_S + C_P + C_T, \tag{8.2}$$

where C_S is the cost of storage, C_P is the cost of the processing units, and C_T is the cost of the transceivers.

The operational cost C_O is the cost incurred for the operation of the entire network for one month period.

$$C_O = E_{\text{total}} \cdot f_E, \tag{8.3}$$

where E_{total} is the average energy (in Joules) required by all nodes to be operational for a period of one month, and f_E is the price of energy (in \$/Joule).

Apart from these factors, the deployment and operational cost may include other factors such as cooling, sensors, peripherals and maintenance, which are not considered in our cost modeling. The components of C_D and C_O are described in detail in the following sections.

Energy consumption: The energy required for the operation of the entire network (expressed in Joules, J) includes the various activities the nodes can perform, such as reading from and writing into the storage, communicating, processing, etc. Energy consumption is calculated for the duration of one month. The energy consumed by a HAN is given by,

$$E_{\text{HAN}} = E_t^{H\to N} + E_{r/w}^H + E_p^H, \tag{8.4}$$

where $E_t^{H\to N}$ is the energy consumed for communication, $E_{r/w}^H$ is the energy consumed for reading from and writing into the storage, and E_p^H is the energy consumed for processing.

The energy consumption for communication is,

$$E_t^{H\to N} = e_{tx}^{H\to N}\ell_{tx}^{H\to N} + e_{rx}^{H\to N}\ell_{rx}^{H\to N}, \tag{8.5}$$

where $e_{tx}^{H\to N}$ ($e_{rx}^{H\to N}$) is the energy required for transmission (reception) of one byte of information between HAN and NAN, and $\ell_{tx}^{H\to N}$ ($\ell_{rx}^{H\to N}$) is the length in bytes of the messages that have been transmitted (received) by the HAN in one month. The energy consumption due to storage is defined as,

$$E_{r/w}^H = e_r^H \ell_r^H + e_w^H \ell_w^H, \tag{8.6}$$

where e_r^H (e_w^H) is the energy required to read (write) one byte of information, and ℓ_r^H (ℓ_w^H) is the length in bytes of the messages that have been read from (written into) the storage in one month. Finally, the energy consumption of processing is defined as,

$$E_p^H = e_p^H n_p^H, \tag{8.7}$$

where e_p^H represents the energy required for processing a byte of information at HAN, and n_p^H is the number of processed bytes.

Similarly, energy consumption for a NAN is given by,

$$E_{\text{NAN}} = E_t^{H\to N} + E_t^{N\to U} + E_{r/w}^N + E_p^N, \tag{8.8}$$

where $E_t^{H\to N}$ is the energy consumed for communication between HANs and NANs, $E_t^{N\to U}$ is the energy consumed for communication between NANs and UCU, $E_{r/w}^N$ is the energy consumed for reading from and writing into the storage, and E_p^N is the energy consumed for processing. These terms are defined as,

$$\begin{aligned} E_t^{H\to N} &= e_{tx}^{H\to N}\ell_{tx}^{H\to N} + e_{rx}^{H\to N}\ell_{rx}^{H\to N} \\ E_t^{N\to U} &= e_{tx}^{N\to U}\ell_{tx}^{N\to U} + e_{rx}^{N\to U}\ell_{rx}^{N\to U} \\ E_{r/w}^N &= e_r^N\ell_r^N + e_w^N\ell_w^N \\ E_p^N &= e_p^N n_p^N. \end{aligned} \tag{8.9}$$

Finally, for the UCU we have,

$$E_{\text{UCU}} = E_t^{N\to U} + E_{r/w}^U + E_p^U. \tag{8.10}$$

The terms $E_t^{N\to U}$ (energy consumption for communication), $E_{r/w}^U$ (energy consumption for storage reading/writing) and E_p^U (energy consumption for processing) are defined as,

$$\begin{aligned} E_t^{N\to U} &= e_{tx}^{N\to U}\ell_{tx}^{N\to U} + e_{rx}^{N\to U}\ell_{rx}^{N\to U} \\ E_{r/w}^U &= e_r^U\ell_r^U + e_w^U\ell_w^U \\ E_p^U &= e_p^U n_p^U. \end{aligned} \tag{8.11}$$

Thus, the total energy consumption for the entire network in a month is,

$$E_{\text{total}} = E_{\text{UCU}} + \sum_{i\in\mathcal{N}} E_{\text{NAN}}(i) + \sum_{j\in\mathcal{M}} E_{\text{HAN}}(j), \tag{8.12}$$

where $\mathcal{N}$ is the set of NANs and $\mathcal{M}$ is the set of HANs in the network.

Communication: The communication cost accounts for the data rate (expressed in bits per second, bps) needed to transmit data from a HAN to the UCU through a NAN. Data rate for a HAN is expressed as,

$$T_{\text{HAN}} = \frac{8\,\ell_m^{H\to N}}{t^{H\to N}}, \tag{8.13}$$

where $\ell_m^{H\to N}$ is the length of the message that has to be transmitted from the HAN to the NAN, and $t^{H\to N}$ is the time period within which a HAN needs to transmit its information to the NAN. Given T_{HAN}, the resulting monetary cost for communication at HAN j is,

$$C_T(j) = T_{\text{HAN}} \cdot f_T(T_{\text{HAN}}), \tag{8.14}$$

where $f_T(\cdot)$ is a non-linear function that models the cost of bandwidth (expressed in \$/bps). Similarly, data rate for a NAN is expressed as,

$$T_{\text{NAN}} = \frac{8\,\ell_m^{N\to U}}{t^{N\to U}}. \tag{8.15}$$

The resulting monetary cost for communication at NAN i is,

$$C_T(i) = T_{\text{NAN}} \cdot f_T(T_{\text{NAN}}). \tag{8.16}$$

Therefore, the total communication cost required for transmission between HANs to NAN and NANs to UCU is expressed as,

$$C_T = \sum_{i \in \mathcal{N}} C_T(i) + \sum_{j \in \mathcal{M}} C_T(j). \tag{8.17}$$

Storage: The storage cost accounts for the total amount of storage capacity (expressed in bytes) required by the node. The storage cost depends on the sampling interval τ and the time duration ΔT for which storage is needed. Thus, the storage requirement for a node k is expressed as,

$$S_k = \Delta T \frac{\ell_m}{\tau}, \tag{8.18}$$

where ℓ_m indicates the length of a message. Depending on the architecture selected and application requirement, ℓ_m and τ varies for each HANs, NANs and UCU. Given S_k, the resulting monetary cost for storage at node k is,

$$C_S(k) = S_k \cdot f_S(S_k), \tag{8.19}$$

where $f_S(\cdot)$ is a non-linear function that models the cost of storage (expressed in \$/byte).

Thus total storage cost of the network for one month is given as,

$$C_S = C_S(\text{UCU}) + \sum_{i \in \mathcal{N}} C_S(i) + \sum_{j \in \mathcal{M}} C_S(j). \tag{8.20}$$

Processing: The processing cost accounts for the number of operations (ops) required to respond to a query received at the node. The in-node operations to respond to a query includes mainly arithmetic and relational operations. Processing cost depends on the number of messages to be processed and number of operations to be performed based on the query. The processing cost at node k calculated for one month is expressed as,

$$P_k = \sum_{q \in \mathcal{Q}} n_m \cdot n_q, \tag{8.21}$$

where $\mathcal{Q}$ is the set of queries generated in the network, which depends on the supported energy services, n_m is the number of messages to be processed and n_q represents the number of operations to be performed for query q. These values depend on the architecture selected, the types of query and the node. Given P_k, the resulting monetary cost for processing at node k is,

$$C_P(k) = P_k \cdot f_P(P_k), \tag{8.22}$$

where $f_P(\cdot)$ is a non-linear function that models the cost of processing (expressed in \$/ops).

Thus total processing cost of the network for one month is given as,

$$C_P = C_P(\text{UCU}) + \sum_{i \in \mathcal{N}} C_P(i) + \sum_{j \in \mathcal{M}} C_P(j). \tag{8.23}$$

Accuracy cost: As mentioned before, in order to reduce storage and communication, a possible strategy is to do a time-wise aggregation of consecutive energy readings. In this way,

an original time-series X of n readings is reduced to a smaller time-series Y of length m, (where $m < n$) by aggregating each k consecutive values in X into a single value y in Y. However, certain services may need to restore the original time series X from Y, using a disaggregation algorithm. The restored time-series $\widehat{X}$ may differ from the original one, X.

Accuracy is therefore an important measure of how accurately the original data can be retrieved from aggregated data. We utilize Normalized Root Mean Square Error (NRMSE) as our accuracy cost. Let $x_t \in X$ be the real energy consumption value of a HAN j at time t, and $\widehat{x}_t \in \widehat{X}$ the energy consumption value that has been inferred through the disaggregation algorithm. The NRMSE is expressed as,

$$\text{NRMSE}(j) = \frac{\text{RMSE}(j)}{x_{\max} - x_{\min}}, \tag{8.24}$$

where $x_{\max}$ and $x_{\min}$ are the maximum and minimum real energy consumption values of X, and

$$\text{RMSE}(j) = \sqrt{\frac{\sum_{t=1}^{n} (x_t - \widehat{x}_t)^2}{n}}. \tag{8.25}$$

The accuracy cost C_A is therefore defined as the average NRMSE among all the HANs. Formally,

$$C_A = \frac{1}{|\mathcal{M}|} \sum_{j \in \mathcal{M}} \text{NRMSE}(j). \tag{8.26}$$

Privacy cost: Although the compression of the original time-series X into Y through time-wise aggregation reduces the accuracy of the restored time-series $\widehat{X}$, it also enhances the customer privacy. In fact, Y can be considered as an obfuscated version of X. To quantify the privacy of the aggregation of k consecutive values into a single aggregated value y, we use Shannon entropy [107] associated with the disaggregation of y into k values. In general, higher the entropy, the higher is the customer privacy. The entropy of a system with $\mathcal{S}$ states is expressed as,

$$H(y) = \sum_{s \in \mathcal{S}} -p(s) \cdot log(p(s)), \tag{8.27}$$

where $p(s)$ is the probability that the system is in state s. In our case, $\mathcal{S}$ is the set of all possible disaggregations, i.e., all the possible ways a value y can be split into k values such that the sum of the k values equal y. The number of possible disaggregations (i.e., the state space size $|\mathcal{S}|$) is called weak integer composition of y into k parts, and it is computed as,

$$|\mathcal{S}| = \binom{y+k-1}{k-1} = \frac{(y+k-1)!}{(k-1)! \cdot y!}.$$

Assuming that each disaggregation of y into k values has the same probability, we can rewrite Eq. (8.27) as,

$$H(y) = log\left(|\mathcal{S}|\right). \tag{8.28}$$

The average entropy of the aggregated time-series Y of HAN j is,

$$H(j) = \frac{\sum_{y \in Y} H(y)}{|Y|}. \tag{8.29}$$

Table 8.1: Queries generated (E=energy consumption)

Queries generated	BAP	DR	DF	EM
E / 10 seconds		X	X	
E / 30 seconds		X		X
E / 15 minutes		X	X	X
E / hour	X	X		
E / day	X	X	X	X
E / month	X		X	X
(Min, Max, Avg)E / day		X	X	
(Min, Max, Avg)E / month	X			X

Thus, the privacy cost C_H of a data processing architecture is defined as,

$$C_H = -\frac{1}{|\mathcal{M}|} \sum_{j \in \mathcal{M}} H(j). \tag{8.30}$$

8.4. EVALUATION

Services in SG have different data requirements, which imply different data acquisition queries generated by the UCU. In our evaluation we consider Billing and adaptive pricing (BAP), Demand Response (DR), Demand Forecast (DF) and Emergency Management (EM). The requirements of the services considered for our cost-benefit analysis are described in Table 8.1.

Billing and adaptive pricing (BAP). In the future, utility providers will be able to bill consumers based on the real-time demand-supply balance. Consumers will also get real-time pricing information in order to alter their energy demand. Thus, queries related to minimum, maximum and average energy consumption, as well as hourly and monthly energy consumption are generated by UCU for this application.

Demand-response (DR). DR strategies are designed to reduce or shift energy consumption from peak periods to off-peak periods. Thus, energy consumption readings at high frequency during peak periods and low frequency consumption readings at off-peak periods are required to envisage DR.

Demand forecast (DF). Demand forecast algorithms can assist utility providers towards efficient distribution of electricity and better planning of resources. Aggregate energy consumption readings and high frequency readings at peak periods can assist in accurate demand forecast.

Emergency Management (EM). Cascading failures and robustness of the grid are some of the challenges that are handled using emergency management strategies. To detect abnormal energy consumption patterns, readings at high frequency are required.

Environment: In our cost-benefit analysis we consider two environments *viz*, *urban* and *rural*. We define the number of HANs and NANs in an urban and rural setup based on Electric Power Research Institute (EPRI) [11] survey about the NAN, population density and number of households in the USA.

Average total population in an urban environment is around $4.8M$, with a maximum population density of roughly $33.7K$ persons per km^2 and land area of 121 km^2. Thus,

Table 8.2: Energy consumption for different operations. [95]

Operations	Energy consumption
Transmission @sub-1GHz	0.164 mJ/byte
Reception @sub-1GHz	0.08 mJ/byte
Transmission @IEEE 802.16	0.324 mJ/byte
Reception @IEEE 802.16	0.100 mJ/byte
Read from flash	0.09 μJ/byte
Write to flash	0.8 μJ/byte
Processing	0.14 μJ/byte

an urban environment is composed of $1.6M$ households. To provide adequate coverage to the collection of energy data from the households, 73 NANs operating at sub-1GHz are required [11].

Rural environments with different terrain and population density are considered to have total population of $1.4M$ and land area of 215000 km^2. Thus, rural environment consists of around $476K$ households, with 76 NANs operating at sub-1GHz to provide coverage [11].

Simulation parameters: A standard wireless sensor node (WSN) is considered as HAN and its configuration depends on the architecture. Each HAN samples data by default every 5 minutes, which can be programmed based on the requirement or upon reception of the query. Each HAN is associated with a sub-1GHz transceiver to communicate with the NAN. Similarly, NANs are equipped with both sub-1GHz and WiMAX transceivers to communicate with HANs and UCU respectively. Table 8.2 shows the energy consumption for different operations performed by the HAN.

Data message contains HAN number, time stamp and energy consumption values. The *Query* message includes the HAN number and query number. Similarly, the *Query-reply* message carries the energy consumption value, HAN number and query number. Finally, the *Aggregated data* includes HAN number, aggregation granularity and aggregated energy consumption value. Message size of data, query, query-reply, aggregated messages are considered to be 50, 5, 10 and 10 bytes respectively.

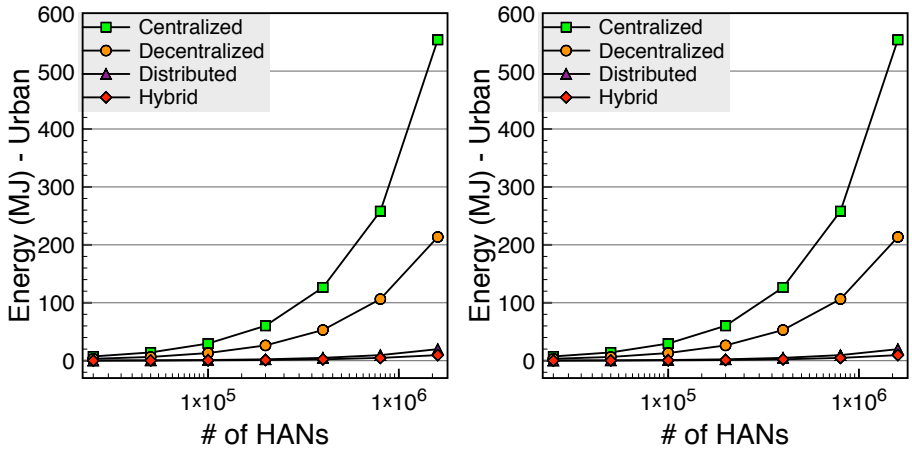
8.5. RESULTS

This section describes the performance of each architecture based on the key cost indicators for urban and rural environments. To calculate the key cost indicators, we used the data over a duration of one month in our simulations.

Energy consumption.

Energy consumption cost per architecture for both urban and rural environments[3] is illustrated in Fig. 8.2a and Fig. 8.2b. In *urban environments*, it is evident that centralized architecture consumes significant amount of energy compared to other architectures. In centralized architecture, complete data needs to be relayed to the UCU, thus increasing the number of transmissions and the energy required. Distributed and hybrid architectures

[3] In our experimental evaluation we considered two cases: (i) each NAN has the same number of HANs, and (ii) each NAN has a uniformly distributed random number of HANs. We found that there is not much difference in energy consumption between the two cases. Thus, for simplicity we consider equal number of HANs being allocated to each NAN.



(a) Urban (b) Rural

Figure 8.2: Energy consumption across architectures.

Table 8.3: Energy consumption distribution for urban environments with 1.6M HANs.

Architectures	Storage	Proc.	Comm.	Total Energy
Centralized	4%	12.6%	83.4%	554.0 MJ
Decentralized	2.3%	16.2%	81.5%	213.5 MJ
Distributed	15.9%	20.4%	63.7%	19.8 MJ
Hybrid	7.1%	3%	89.9%	9.6 MJ

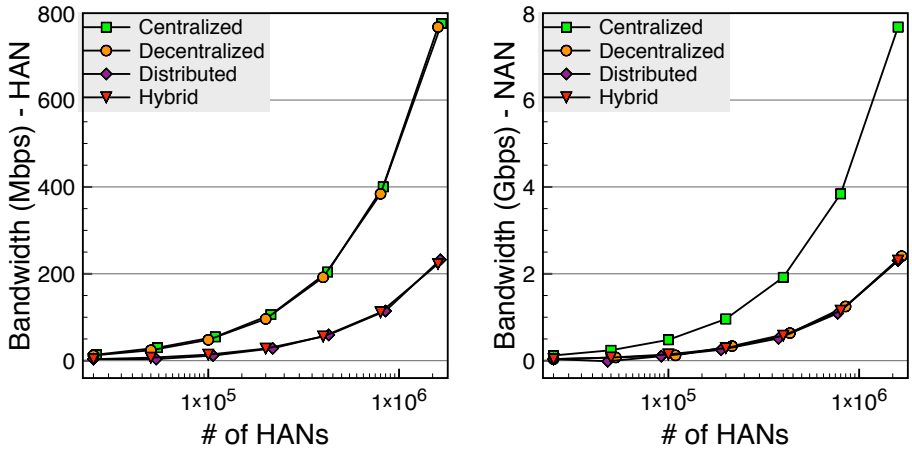
consume much lower energy compared to centralized and decentralized architectures. The significant energy saving in distributed approaches is due to the reduced number of transmissions. Energy consumption of the hybrid architecture is the lowest compared to all other architectures. This energy saving is achieved by sending aggregated data to NANs as compared to storing data only at HANs, as in distributed architecture.

In general, the total energy consumption increases rapidly as the number of houses increases for centralized and decentralized architectures. In case of distributed and hybrid architectures, the increase in energy consumption is very gradual, thus increasing their scalability. Similar trends can be seen for rural environments, as shown in Fig. 8.2b.

Table 8.3 shows energy consumption distribution for different architectures (number of HANs = 1.6M). We remark that the energy consumption considers only communication, storage and memory operation, although other factors could be considered, such as cooling, lights, etc. It is evident that the most significant energy factor in all the architectures is communication. Hence, reducing communication needs can in turn reduce overall energy consumption, as can be seen in distributed and hybrid architectures.

Communication.

The communication cost as described in Section 8.3 is the average data rate required to support the SG services considered. The time of reference $t^{H\rightarrow N}$ and $t^{N\rightarrow U}$ in Eq. (8.13) and Eq. (8.15) are considered to be 1 s. Table 8.4 shows the average bandwidth requirement at each HAN and NAN for both urban and rural environments with 1.6M HANs and 476K HANs respectively. Data rate requirement for a HAN is same, irrespective of the environment, as each HAN transmits the same data based on the architecture selected. However,



(a) Bandwidth-HAN (b) Bandwidth-NAN

Figure 8.3: Bandwidth required in urban environment.

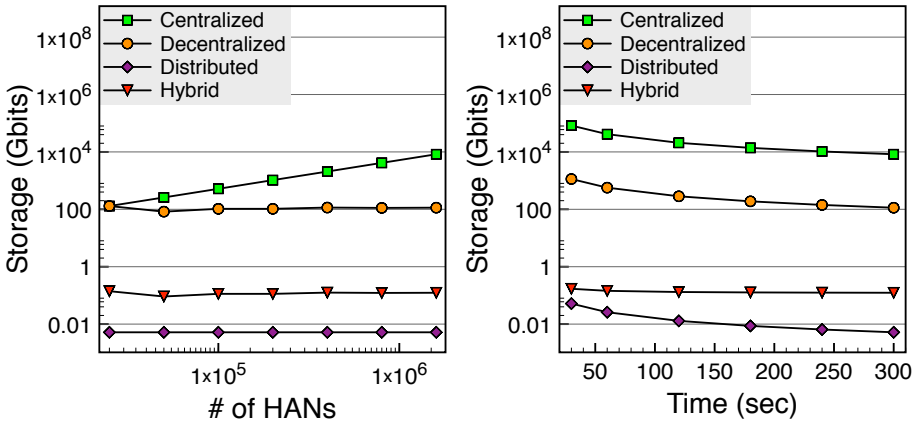
Table 8.4: Bandwidth required for various architectures.

	Urban		Rural	
Architectures	HAN	NAN	HAN	NAN
Centralized	480 bps	11 Mbps	480 bps	3 Mbps
Decentralized	480 bps	32 Mbps	480 bps	1 Mbps
Distributed	144 bps	3 Mbps	144 bps	1 Mbps
Hybrid	138 bps	2 Mbps	138 bps	0.5 Mbps

the data rate required at NANs in urban environment is higher than rural environment, since more HANs are associated with each NAN in an urban environment. The bandwidth requirements from HANs to NAN and NANs to UCU in an urban environment are shown in Fig. 8.3a. The needed bandwidth between HANs and NAN is higher for centralized and decentralized architectures.However, since distributed storage and processing is adopted in distributed and hybrid architectures, the number of transmissions performed at the HAN is reduced. Thus, the bandwidth requirement is significantly reduced in these architectures. Similarly, the average bandwidth requirement between NANs and UCU is shown in Fig. 8.3b. In general, the bandwidth increases with the number of houses as seen in the Fig. 8.3. Similar trends with scaled-down bandwidth requirements are observed for rural environments.

Storage.

Storage required by each node for different architectures in urban environment, for duration $\Delta T = 1$ year is shown in Fig. 8.4a. The default sampling interval of 5 minutes is considered to determine the storage cost as described in Eq. (8.18). As with other costs, storage cost for centralized architecture is the highest compared to other architectures, as all data is stored at one place i.e. the UCU. The storage required by other architectures is much lower than the centralized architecture, with distributed architecture having the lowest. As the number of HANs increases, the required storage increases linearly in centralized architectures. On the other hand, since equal number of HANs are allocated to each NAN, for all the other architectures storage is constant, as seen in the Fig. 8.4a.



(a) # of HANs (b) Sampling Interval

Figure 8.4: Required storage in urban environment.

Table 8.5: Processing operations in urban environment.

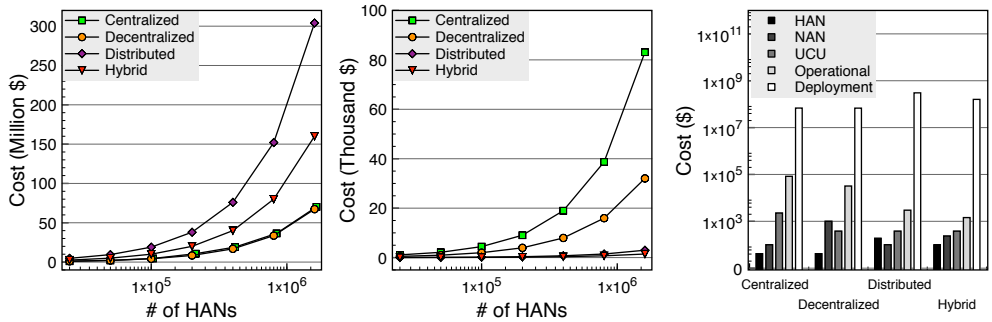
Architectures	Number of operations
Centralized	9000 M-ops
Decentralized	9000 M-ops
Distributed	0.36 M-ops
Hybrid - NAN	15.75 M-ops
Hybrid - HAN	0.008 M-ops

Finally, the required storage as a function of sampling interval is shown in Fig. 8.4b. Higher sampling intervals indicate less frequent sensing of energy values. Storage cost in general decreases with increase in sampling interval, regardless of the architecture.

Processing.

Processing accounts for the number of operations performed to respond to a query as described in Section 8.3. Processing requirements depend on the number of messages the node has to process before replying. For each query, all messages until the reception of query are processed and each query is independent of other queries. Thus the processing requirement depends on when the query is received (in turn number of messages to be processed) and the operations performed. The processing requirement for one month duration in an urban environment with 1.6M HANs is shown in Table 8.5.

In centralized architectures, since UCU performs all processing, the processing requirements increase with the number of houses. Since number of HANs per NAN is constant, processing at each NAN in decentralized architectures is merely a constant with increase in number of houses. In a distributed architecture, processing is done in a distributed manner at each HAN, thus reducing the number of operations at each HAN by order of four compared to decentralized architecture. In hybrid architectures, processing effort is distributed at both HANs and NANs and has the least processing cost at each HAN. Similar trends are also observed for rural environments.



(a) Deployment cost (b) Operational cost (c) Cost distribution

Figure 8.5: Monetary cost details for deployment, operational and each node in urban environment.

Cost-Benefit Analysis.

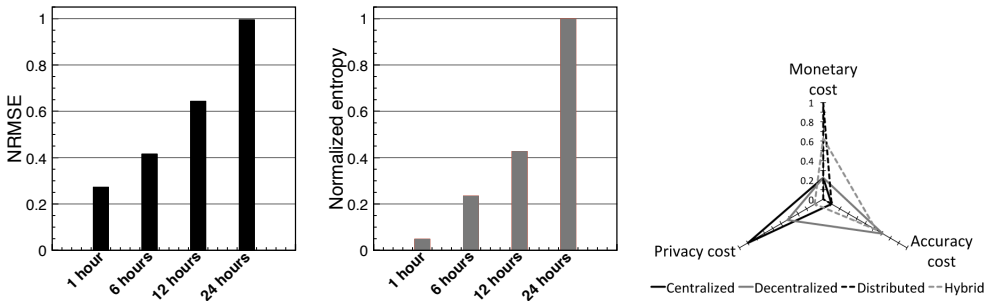
Monetary cost: The monetary cost as described in Section 8.3 accounts for deployment and operational costs. Deployment cost is the cost for installing nodes (HANs, NANs, UCU) and varies based on the capabilities provided to each node with respect to processing, storage and communication. Operational cost is calculated based on the energy required to operate all nodes for one month. The cost of electricity (f_E) in Eq. (8.3) is considered to be 0.194 $/KWh, based on Pacific Gas and Electric Company[4] yearly average electricity costs. Based on the storage, processing and communication requirements obtained in previous Section, appropriate modules and price details are considered from digikey[5]. In general, $f_S(\cdot)$ in Eq. (8.19) model the cost of storage per byte and is considered to vary from 2$ to 60$ for 256KB to 500GB of storage. Similarly $f_T(\cdot)$ in Eq. (8.14) models the cost of transceivers and is considered to be 5$ and 10$ for $sub-1GHz$ and WiMAX transceivers respectively. $f_P(\cdot)$ in Eq. (8.22) models the cost of processing and varies from 5$ (MSP43016xx processor) to 60$ (ARM Cortex-M3 processor).

8

The deployment cost for each architecture as a function of number of houses is shown in Fig. 8.5a. Since processing and storage are performed by only UCU in centralized architecture, the total deployment cost is the lowest compared to all other architectures. In decentralized architectures, all the NANs have storage and processing capabilities. The distribution of processing and storage capabilities to the NANs overcomes single point failure but follows the same trend in monetary cost as compared to centralized architectures. In distributed architecture, each HAN is equipped with processing and storage capabilities, thus increasing the monetary cost of each node in the network. Due to sheer number of HANs, the total deployment cost of the distributed architecture increases rapidly with number of HANs. The deployment cost in hybrid architectures is lower compared to a distributed architecture, since processing and storage are distributed at both HANs and NANs. The operational cost for various architectures in urban environment is shown in Fig. 8.5b.

Fig. 8.5c shows the break-up of monetary cost incurred for each architecture in urban environment with 1.6M HANs. The figure shows the cost in $ for each HAN, NAN, UCU

[4] http://www.pge.com/.
[5] http://www.digikey.com/



(a) NRMSE (b) Normalized entropy (c) Cost-Benefit analysis

Figure 8.6: Accuracy, privacy and cost-benefit analysis.

(with storage, processing, communication), as well as the operational and deployment cost across various architectures. As seen in the figure, adding storage and processing features to each HAN increases the deployment cost of distributed and hybrid architectures. However, distributed and hybrid architectures have the least operational cost compared to centralized and decentralized architectures and hence more energy efficient.

Accuracy cost: Centralized and distributed architectures do not employ data aggregation, as data is stored either in UCU or HAN. Since complete data is available at each NAN in a decentralized architecture, low data aggregation granularity of 1 hour is employed. In hybrid architecture, NAN receives data aggregated with granularity every 12 hours from each HAN. NRMSE is the metric used to determine the accuracy based on data aggregation granularity. To calculate the accuracy cost, we used two weeks of data collected by the REDD [72]. Data aggregation granularity of 1, 6, 12, 24 hours are considered, while the step-size of the time-series to be restored is assumed to be 5 minutes. The focus here is not on accurate disaggregation algorithms. For this reason, we employ a rather simple algorithm that equally split an aggregated reading into the 5 minute buckets of the time-series to be restored. Fig. 8.6a shows the NRMSE values for various data aggregation granularities. It is evident that, higher the data aggregation granularity, the higher is the NRMSE and thus lower is the accuracy.

Privacy cost: This cost factor also depends on the data aggregation granularity. Intuitively, higher the data aggregation granularity, the higher is the entropy and thus lower is the privacy cost. In centralized architectures, all sensitive data of customers is available at UCU, thus making centralized architecture the least privacy-preserving architecture. On the other hand, distributed architectures with distributed storage ensure that customer's sensitive data is stored locally at HAN, making them completely privacy-preserving.

However, privacy cost varies in decentralized and hybrid architectures. As before, we used two weeks of data collected by the REDD initiative [72] to calculate the privacy cost. Fig. 8.6b shows the average entropy for different data aggregation granularities. The figure shows that the higher the data aggregation granularity, the higher the entropy, therefore the lower the privacy cost (see Eq. (8.30)).

The radar plot in Fig. 8.6c shows the performance of each architecture according to the monetary, accuracy and privacy costs. All cost indicators have been normalized to the [0,1] interval. Clearly, the lower the value of monetary, accuracy and privacy costs, the more desirable is the architecture for the smart grids.

Centralized architectures has low monetary cost, high privacy cost and low accuracy cost, thus making them less privacy-preserving but economically cheaper, since all data is stored and processed at the UCU. Thus, centralized architectures are less scalable and suffer from single point of failure but with low deployment cost. *Decentralized architectures* on the other hand have low monetary cost, moderate privacy cost and moderate accuracy cost. The privacy cost reduction is achieved with data aggregation at NANs, which also increases the accuracy cost. Decentralized architecture distributes the processing and storage efforts to the NANs thus achieving moderate privacy and low monetary cost. *Distributed architectures* have the highest monetary cost, the lowest privacy and accuracy costs. Distributed architectures are clearly the most privacy-preserving system, since data is stored and processed locally, however this increases the monetary cost of the architecture. *Hybrid architectures* have lower monetary cost compared to distributed architectures, with low privacy cost and moderate accuracy cost. Distributing storage and processing at both HANs and NANs reduces the deployment cost, with a loss in accuracy due to data aggregation at NANs. Also, hybrid architectures are more energy efficient but with high deployment cost.

8.6. CONCLUSIONS

In this chapter, we presented different data-processing architectures for SG. To gain insights, the proposed architectures were evaluated with respect to energy consumption, processing power, storage requirements, communication bandwidth, accuracy of data collection and privacy. From our cost-benefit analysis, we conclude that even though centralized architectures perform well in terms of accuracy and deployment cost, they are less scalable and the least privacy preserving. On the other hand, distributed architectures overcome privacy issues with local storage and processing, but with additional deployment cost. Decentralized architectures perform well in terms of accuracy and monetary cost, while hybrid architectures increase the privacy by increasing the deployment cost. Thus, the choice of the architecture could be to have a more energy-efficient architecture or highly-scalable distributed architecture with high deployment cost or simple less-scalable architecture and depends upon the objective of the implementation. The proposed comprehensive analysis on various data-processing architectures shows how to cope with the overwhelming data generated from smart meters.

9

CONCLUSIONS

"Two roads diverged in a wood, and
I took the one less traveled by, and that has made all the difference."

Robert Frost

Sustainable development while reducing the energy footprint has become an important field of study. This calls for many innovative solutions in the generation, transmission, and consumption of energy. In this thesis we majorly focused on the consumption-side of energy systems.

We explored how to develop effective personalized energy services towards designing sustainable smart energy systems. We argue that the design and development of energy services should follow a personalized approach giving paramount importance to consumer preferences. We followed a bottom-up approach where energy services are developed from *nano* scale targeting individual households, to *micro* scale targeting smart buildings, to *macro* scale targeting neighborhoods and cities. To this end, we presented an approach – physical analytics for sustainable and smart energy systems – that combines IoT data, physical modeling and data analytics to develop intelligent, personalized energy services. Our results are grouped in three parts as follows.

In Part I, we discussed energy services that raise awareness among occupants of a household by providing fine-grained energy consumption information in real-time. Specifically, we developed a simple, effective energy-disaggregation algorithm to derive appliance-level information. Further, we designed a personalized energy-disaggregation system moving from appliance-level towards user-level information. This energy-disaggregation system provides per-occupant energy-usage information that is used to create better energy management techniques apart from raising awareness.

In Part II, we proposed energy services to reduce/shift demands by considering the preferences of individual occupants. We presented a demand-scheduling algorithm that minimizes the user discomfort and electricity cost by shifting usage of appliances. Several metrics to analyze user preferences and appliance-usage patterns from historic energy-consumption data were described. Further, consumer preferences with respect to indoor temperature and lighting levels were modeled to develop a demand-reduction algorithm.

The proposed iLTC system improves user comfort and efficient usage of HVAC and lighting systems in smart buildings.

In Part III, we explored how to develop energy services that can be applied across various segments of consumers. To this end we presented a temporal demand-regulation scheme that can identify target consumers in a neighborhood for various DR programs. After identifying consumers, utilities can target specific groups of consumers to adjust their demand, such as reduction in average and peak demand. Further, to develop effective energy services at the city level, a comprehensive analysis on where to store and process smart-meter data was required. We presented four data-processing architectures and analyzed which architecture is best suited to efficiently store and process the data for various energy services.

9.1. RECAPITULATION

As mentioned above, the essence of this thesis is to raise awareness, reduce energy consumption and at the same time improve user comfort in smart – homes, buildings, and neighborhoods. We now enlist the three main achievements of the work presented in this thesis:

- We demonstrated that, with minimal intrusion and computational complexity, fine-grained energy-consumption information can be derived accurately in real time to raise awareness (Part I).
- We have achieved a significant reduction in energy consumption and electricity cost by modeling consumer preferences and energy-usage patterns (Part II).
- We presented novel schemes toward developing consumer-centric energy services that analyze and classify households based on the energy-consumption data and consumer preferences to provide targeted recommendations (Part III).

Through the above accomplishments, this thesis advocates developing personalized energy services towards the design of sustainable, smart energy systems. The proposed energy services follow a distributed approach with low-complexity algorithms that are scalable, from one household to tens and thousands of households. Let us now recapitulate our contributions through the prism of personalization.

Data collection. Fine-grained data collection is the cornerstone of developing personalized energy services. While smart-meter deployments have paved the way to collect energy-usage data at the house level, fine-grained data at appliance-level and user-level is still not readily available. We presented algorithms that utilize least amount of data from sensors already deployed in the households to derive fine-grained consumption data at appliance and user levels. State-of-the-art algorithms work well only on the specific household from where training data was collected and involve high-computational complexity. The algorithms presented in this thesis circumvent these limitations and perform significantly better than the state-of-the-art algorithms with low complexity. The derived fine-grained data is used by personalized energy services to raise awareness and provide information on faulty appliances, per-appliance and per-occupant energy consumption. Furthermore, energy services can use this information to model per-occupant energy-usage behavior, find anomalies in energy usage and predict energy consumption. Through the proposed inference algorithms we demonstrated that effective personalized energy services can be developed without deployment of additional sensors, thus avoiding maintenance and deployment cost.

User preferences. To tailor energy services to individual consumers, apart from data collection, accurate modeling of consumer characteristics is crucial. We presented two approaches to derive user preferences keeping in mind, (i) the applicability across households and (ii) the ability to run in real time. Our first approach uses data-driven techniques to infer user preferences from historic energy-consumption information. With the inferred user preferences, the proposed personalized energy services such as demand shifting and TDR outperform traditional energy management services. The second approach employed a crowd-sourced technique to collect user preferences with the help of a smartphone App. We demonstrated that with a few user-provided inputs one can develop comprehensive models of user comfort requirements. Hitherto, energy services were generic and hence ineffective. Energy services based on consumer preferences enable active participation and promote sustainable energy usage. Even though our data-driven approaches provide the required information to tailor energy services quite effectively, there is still a need to develop generic algorithms to model large numbers of consumers.

Scalability. While we tailor energy services to cater to individual consumer preferences, it is also important to ensure applicability of personalized energy services at various scales – households, buildings, neighborhoods, and cities. We address the scalability of energy services at two levels (i) at occupant level – the ability to provide information to multiple occupants in a household based on their preferences and (ii) at household level – the ability to provide information to multiple households in a neighborhood/city. To support scalability at occupant-level, we developed low-complexity online algorithms that can run on consumer devices such as smartphones and send only events to the local server (a Raspberry PI), thus eliminating the need to transfer all the data to a server for processing. This allows the local server to process event streams from multiple occupants in real time. Similarly, at household-level we proposed algorithms that utilize historic energy consumption of a household to characterize energy-usage behavior and can run locally at each household. They were evaluated on a dataset with more than 4,000 households to characterize their energy-usage behavior and provide targeted recommendations towards energy reduction. This validates the personalized energy services that are developed for large numbers of consumers. Furthermore, it is also crucial to design efficient end-to-end energy systems addressing the issues related to data storage and processing for these services. Our proposed data-processing architectures provide insights such as where to store and process large-scale smart-meter data from millions of households for various energy services.

Real-time feedback. To raise awareness and enable active consumer participation, personalized energy services need to provide real-time feedback to the consumers. The real-time requirements vary depending on the energy service. Services such as identification of faulty appliances and current energy usage of an occupant need information relaying in real time to alert consumers, whereas services such as day-ahead demand shifting and monthly electricity-bill estimation may tolerate higher latency. The former can be implemented with a distributed architecture to minimize latency; the latter can follow a hybrid architecture where some processing can be offloaded to aggregation nodes. The designers can choose the best architecture depending on the latency requirements. Furthermore, we developed a smartphone App to provide real-time recommendations to consumers on appliance-level consumption, per-appliance energy usage and cost per month, and estimated monthly energy cost based on usage patterns. Further analysis on how consumers react (or adapt) to these recommendations is needed to develop effective feedback mechanisms.

In summary, we demonstrated that personalized energy services can be implemented at large scales to promote sustainable energy usage. We showed that using less data and with the help of personalized feedback and recommendations in real time we can enhance consumers –awareness, –participation and –engagement. We followed a bottom-up approach wherein energy services were proposed at various scales. Unlike prior works focusing on synthetic data, we advocated data-driven techniques to model energy usage behavior of consumers and energy consumption patterns. The proposed personalized energy services were implemented and evaluated in a real-world setting and across multiple datasets. As such this thesis provides a promising direction towards the development of real time, scalable, personalized energy services consequently leading to designing sustainable, smart energy systems at scale.

9.2. FUTURE WORK

In this section, we offer some ideas for future work into the topics covered in this thesis. While we tackled some of the key challenges involved in developing personalized energy services, there still exist several open challenges for the large-scale development of personalized energy services. We enlist some of them below.

Data-driven techniques. The adoption of energy services at large scale is still limited due to the lack of consideration of consumer preferences. Hitherto, energy services were evaluated either using numerical analysis or simulations resulting in unrealistic assumptions about acceptance and usage of these services. In this thesis, energy services were developed by analyzing real-world data. While we demonstrated the advantages of our data-driven energy services at households and buildings, there is a need to evaluate these services at large scale. This requires collecting information about energy consumption of occupants at the city level. Currently there are a few pilot projects like the Smart Grid, Smart City project [8] and Pecan street [59] that collect this information. Future energy services should be evaluated across large-scale using real-world datasets from various cities for its wide adoption.

Towards generic algorithms. The energy disaggregation and apportioning algorithms proposed in this thesis require information about the occupants (location and activity) and their appliances (type and state). Collecting this information in large-scale deployments across households is not feasible. Deriving appliance-level information in a household is challenging due to two reasons *viz.*, multiple appliances with similar energy-consumption profile and appliances with multiple states. LocED addresses this by considering occupants' location to constrain the number of appliances during disaggregation. However, when there are occupants in each room it fails to constrain the set of appliances. Developing generic algorithms to predict the number/type of appliances, user location and activities is non-trivial. Recently, deep neural networks based energy-disaggregation algorithms [65, 80] were proposed, where neural nets are trained on few appliances in a household. This trained model is later used to infer appliance-usage from a different household. They have been found to provide very good results. Similarly, deep neural-nets are also employed for human activity recognition [96]. While these are promising, they require large amounts of labeled data. It is also difficult to analyze this data for inferring an event on an embedded device. Some open questions are: how to develop an effective model with little/no training data? Can the trained model run locally on an embedded device?

Integrating energy services on the fly. In this thesis, we developed energy services to raise awareness, provide actionable feedback and promote sustainable energy usage. While

these services target specific scenarios, there is a need to support additional services without much overhead. Currently, with large-scale penetration of renewable sources such as solar and wind, consumers also produce energy at various capacities. Energy services targeting these new actors should be explored. Supporting additional services on the fly is non-trivial as the system needs to learn new relevant information about energy production and usage. A generic framework needs to be developed that tackles challenges apart from the ones described in this thesis such as privacy-preserving mechanisms and user-behavior analysis, and can support additional services on the fly.

Applicability of personalized services to other domains. Finally, the applicability of the proposed mechanisms across various domains, especially to other resources such as water and gas needs to be investigated. We believe personalized services such as disaggregation, apportioning and demand reduction can be directly applied to water consumption in smart houses and buildings.

BIBLIOGRAPHY

[1] "Building the power grid of the future." [Online]. Available: http://www.news.gatech.edu/features/building-power-grid-future

[2] "Dutch residential energy dataset." [Online]. Available: http://www.st.ewi.tudelft.nl/~akshay/dred/

[3] "Energy outlook 2010. Energy Information Administration." [Online]. Available: http://www.eia.doe.gov/oiaf/ieo/index.html

[4] "International data corporation (idc)." [Online]. Available: http://www.idc.com/getdoc.jsp?containerId=prUS25903815

[5] "Measurement and instrumentation data center." [Online]. Available: http://www.nrel.gov/midc/

[6] "Meteorologisk institutt." [Online]. Available: http://www.yr.no/

[7] "Physical analytics." [Online]. Available: http://researcher.watson.ibm.com/researcher/view_group.php?id=6566

[8] "Smart grid and smart city project and." [Online]. Available: https://www.energyaustralia.com.au/sgsc

[9] "Smart grids strategic research agenda towards 2035." [Online]. Available: http://www.smartgrids.eu/documents/sra2035.pdf

[10] "U. S. dept of energy, advanced metering infrastructure," 2008. [Online]. Available: https://www.smartgrid.gov/sites/default/files/pdfs/advanced_metering_infrastructure_02-2008.pdf

[11] "Electric Power Research Institute, Wireless Field Area Network Spectrum Assessment," 2010.

[12] "The commission for energy regulation (cer), "Electricity customer behaviour trial",," 2012.

[13] "Council of european energy regulators, 2020 vision for europe's energy customers. a discussion paper," 2012, ref. C12-SC-02-04.

[14] "Electricity customer behaviour trial, the commission for energy regulation (cer)," 2012.

[15] "Who urban population growth," 2014. [Online]. Available: http://www.who.int/gho/urban_health/situation_trends/urban_population_growth_text/

[16] "Price waterhouse coopers (pwc)," 2016. [Online]. Available: http://www.pwc.co.uk/

[17] "U. N. world cities report," 2016. [Online]. Available: http://wcr.unhabitat.org/

[18] W. Abrahamse and L. Steg, "How do socio-demographic and psychological factors relate to households' direct and indirect energy use and savings?" *Journal of Economic Psychology*, vol. 30, October 2009.

[19] U. S. E. I. Administration. [Online]. Available: http://www.c2es.org/technology/overview/buildings

[20] A. Aggarwal, S. Kunta, and P. Verma, "A proposed communications infrastructure for the smart grid," in *Innovative Smart Grid Technologies (ISGT)*, 2010.

[21] M. Arenas-Martinez, S. Herrero-Lopez, A. Sanchez, J. Williams, P. Roth, P. Hofmann, and A. Zeier, "A comparative study of data storage and processing architectures for the smart grid," in IEEE SmartGridComm, 2010.

[22] S. Aust, R. Prasad, and I. Niemegeers, "Performance evaluation of sub 1 ghz wireless sensor networks for the smart grid," in IEEE Local Computer Networks, 2012.

[23] M. Azizyan, I. Constandache, and R. Choudhury, "Surroundsense: mobile phone localization via ambience fingerprinting." In Proceedings of the 15th annual international conference on Mobile computing and networking (MobiCom), 2009.

[24] Z. Baharlouei and M. Hashemi, "Efficiency-fairness trade-off in privacy-preserving autonomous demand side management," *IEEE Transactions on Smart Grid*, vol. 5, no. 2, pp. 799–808, 2014.

[25] E. Bakshy, I. Rosenn, C. Marlow, and L. Adamic, "The role of social networks in information diffusion," in *Proceedings of the 21st international conference on World Wide Web (WWW '12), ACM*, 2012.

[26] M. Baranski and J. Voss, "Genetic algorithm for pattern detection in nialm systems," in *In Proc. SMCS. IEEE*, 2004.

[27] C. Basu, J. J. Caubel, K. Kim, E. Cheng, A. Dhinakaran, A. M. Agogino, and R. Martin, "Sensor-based predictive modeling for smart lighting in grid-integrated buildings," *Sensors Journal, IEEE*, vol. 14, no. 12, pp. 4216–4229, 2014.

[28] N. Batra, H. Dutta, and A. Singh, "Indic: Improved non-intrusive load monitoring using load division and calibration," in *In Proc. of ICMLA*, 2013.

[29] N. Batra, J. Kelly, O. Parson, H. Dutta, W. Knottenbelt, A. Rogers, A. Singh, and M. Srivastava, "Nilmtk: An open source toolkit for non-intrusive load monitoring," in *In Proc. e-Energy. ACM*, 2014.

[30] P. Berkhin, "A survey of clustering data mining techniques," in Grouping multidimensional data. , *Heidelberg*. Berlin: Springer, 2006, pp. 25–71.

[31] D. Bonino, F. Corno, and L. D. Russis, "Home energy consumption feedback: A user survey," *Energy and Buildings*, vol. 47, pp. 383–393, 2012.

[32] D. Brunelli and L. Tamburini, "Residential load scheduling for energy cost minimization," *Energy Conference (ENERGYCON), 2014 IEEE International*, 2014.

[33] R. F. Chang and C. N. Lu, "Feeder reconfiguration for load factor improvement," *Power Engineering Society Winter Meeting, IEEE*, vol. 2, no. 7, pp. 980–984, 2002.

[34] C. Chen, J. Wang, and S. Kishore, "A distributed direct load control approach for large-scale residential demand response," *in IEEE Transactions on Power Systems*, vol. 29, no. 5, pp. 2219–2228, 2014.

[35] S. Chen, N. B. Shroff, and P. Sinha, "Heterogeneous delay tolerant task scheduling and energy management in the smart grid with renewable energy," *in IEEE J. Sel. Areas Communication*, 2013.

[36] X. Chen, T. Wei, and S. Hu, "Uncertainty-aware household appliance scheduling considering dynamic electricity pricing in smart home," *in IEEE Transactions on Smart Grid*, vol. 4, no. 2, pp. 932–941, 2013.

[37] Y. Cheng, K. Chen, B. Zhang, C. j. M. Liang, X. Jiang, and F. Zhao, "Accurate real-time occupant energy-footprinting in commercial buildings," in In Proceedings of ACM BuildSys, 2012.

[38] G. Chicco, R. Napoli, and F. Piglione, "Comparisons among clustering techniques for electricity customer classification," *IEEE Transactions on Power Systems*, vol. 21, no. 2, pp. 933–940, 2006.

[39] B. D. Chuong and B. Serafim, "What is the expectation maximization algorithm?" *Nature Biotechnology*, vol. 26, no. 8, pp. 897–899, 2008.

[40] S. Darby, "The effectiveness of feedback on energy consumption, billing and direct displays," *A review for DEFRA of the literature on metering*, 2006.

[41] R. D. Dear and G. S. Brager, "Developing an adaptive model of thermal comfort and preference," *Center for the Built Environment*, 1998.

[42] D. Easley and J. Kleinberg, "Networks, crowds, and markets: Reasoning about a highly connected worldnew york,," 2010.

[43] V. L. Erickson, S. Achleitner, and A. E. Cerpa, "Poem: power-efficient occupancy-based energy management system," in *In Proceedings of international conference on Information processing in sensor networks (IPSN)*, 2013.

[44] V. L. Erickson and A. E. Cerpa, "Thermovote: participatory sensing for efficient building hvac conditioning," in *Proceedings of the 4th ACM Workshop on Embedded Systems for Energy-Efficient Buildings*. ACM, 2012, pp. 9–16.

[45] X. Fang, S. Misra, G. Xue, and D. Yang, "Smart grid - the new and improved power grid: A survey," *Communications Surveys & Tutorials, IEEE*, vol. 14, no. 4, pp. 944–980, 2012.

[46] X. Fang, D. Yang, and G. Xue, "Evolving smart grid information management cloudward: A cloud optimization perspective," *Smart Grid, IEEE Transactions on*, vol. 4, no. 1, pp. 111–119, 2013.

[47] C. Fischer, "Feedback on household electricity consumption: A tool for saving energy?" 2008.

[48] C. Frale and A. E. Raftery, "How many clusters? which clustering method? answers via model-based cluster analysis," *Computer Journal*, vol. 41, pp. 578–588, 1998.

[49] M. Fürsch, S. Hagspiel, C. Jägemann, S. Nagl, D. Lindenberger, and E. Tröster, "The role of grid extensions in a cost-efficient transformation of the european electricity system until 2050," *Applied Energy*, vol. 104, 2013.

[50] A. D. Galasiu and J. A. Veitch, "Occupant preferences and satisfaction with the luminous environment and control systems in daylit offices: a literature review," *Energy and Buildings*, vol. 38, no. 7, pp. 728–742, 2006.

[51] A. Ghahramani, F. Jazizadeh, and B. Becerik-Gerber, "A knowledge based approach for selecting energy-aware and comfort-driven hvac temperature set points," *Energy and Buildings*, vol. 85, pp. 536–548, 2014.

[52] F. Giannotti, D. Pedreschi, A. Pentland, P. Lukowicz, D. Kossmann, J. Crowley, and D. Helbing, "A planetary nervous system for social mining and collective awareness"," *The European Physical Journal Special Topics*, no. 1, pp. 49–75.

[53] U. Gneezy, E. Haruvy, and H. Yafe, "The inefficiency of splitting the bill," *The Economic Journal*, vol. 114, no. 495, pp. 265–280, 2004.

[54] Y. Guo, M. Jones, B. Cowan, and R. Beale, "Take it personally: personal accountability and energy consumption in domestic households," 2013.

[55] S. Gupta, M. S. Reynolds, and S. N. Patel, "Electrisense: Single-point sensing using emi for electrical event detection and classification in the home," in *In Proc. UbiComp. ACM*, 2010.

[56] G. W. Hart, "Nonintrusive appliance load monitoring," in *Proc. of the IEEE*, 1992.

[57] S. Hay and A. C. Rice, "The case for apportionment," 2009.

[58] L. Hernandez, C. Baladron, J. M. Aguiar, B. Carro, A. J. Sanchez-Esguevillas, J. Lloret, and J. Massana, "A survey on electric power demand forecasting: future trends in smart grids, microgrids and smart buildings," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 3, pp. 1460–1495, 2014.

[59] C. Holcomb, "Pecan street inc.: A test-bed for nilm," in *In Proc. NILM workshop*, 2012.

[60] Y. j. Kim, M. Thottan, V. Kolesnikov, and W. Lee, "A secure decentralized data-centric information infrastructure for smart grid," *Communications Magazine, IEEE*, 2010.

[61] Y. j. Wen and A. M. Agogino, "Wireless networked lighting systems for optimizing energy savings and user satisfaction," in *Wireless Hive Networks Conference*. IEEE, 2008, pp. 1–7.

[62] C. Joe-Wong, S. Sen, S. Ha, and M. Chiang, "Optimized day-ahead pricing for smart grids with device-specific scheduling flexibility," *IEEE J. Sel. Areas Communication*, vol. 30, no. 6, pp. 1075–1085, 2012.

[63] J.Z.Kolter and T.Jaakkola, "Approximate inference in additive factorial hmms with application to energy disaggregation," *In Proc. AIS-TATS*, 2012.

[64] S. Karjalainen, "Consumer preferences for feedback on household electricity consumption," *Energy and Buildings*, vol. 43, pp. 458–467, 2011.

[65] J. Kelly and W. Knottenbelt, "Neural nilm: Deep neural networks applied to energy disaggregation," in *In Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*, 2015.

[66] A. Kelman, Y. Ma, and F. Borrelli, "Analysis of local optima in predictive control for energy efficient buildings," *Journal of Building Performance Simulation*, vol. 6, no. 3, pp. 236–255, 2013.

[67] H. Kim, Y. j. Kim, K. Yang, and M. Thottan, "Cloud-based demand response for smart grid: Architecture and distributed algorithms," in IEEE SmartGridComm, 2011.

[68] H. Kim, M. Marwah, M. Arlitt, G. Lyon, and J. Han, "Unsupervised disaggregation of low frequency power measurements," in *In Proc. SDM. SIAM*, 2010.

[69] Y. Kim, T. Schmid, Z. Charbiwala, and M. Srivastava, "Viridiscope: Design and implementation of a fine grained power monitoring system for homes." In UbiComp, 2009.

[70] L. Klein, J. Y. Kwak, G. Kavulya, F. Jazizadeh, B. Becerik-Gerber, P. Varakantham, and M. Tambe, "Coordinating occupant behavior for building energy and comfort management using multi-agent systems," *Automation in Construction*, vol. 22, pp. 525–536, 2012.

[71] J. Kleinberg, "The convergence of social and technological networks," *Communications of the ACM*, vol. 51, no. 11, pp. 66–72, 2008.

[72] J. Z. Kolter and M. J. Johnson, "Redd: A public data set for energy disaggregation research." In SustKDD, 2011.

[73] A. Kusiak, M. Li, and F. Tang, "Modeling and optimization of hvac energy consumption," *Applied Energy*, vol. 87, no. 10, pp. 3092–3102, 2010.

[74] A. H. Y. Lam, Y. Yuan, and D. Wang, "An occupant-participatory approach for thermal comfort enhancement and energy conservation in buildings." in *Proceedings of the 5th international conference on Future energy systems*. ACM, 2014, pp. 133–143.

[75] S. Lee, D. Ahn, S. Lee, R. Ha, and H. Cha, "Personalized energy auditor: Estimating personal electricity usage," in *In Proc. of the International Conference on Pervasive Computing and Communications*, 2014.

[76] T. Logenthiran, D. Srinivasan, and T. Z. Shun, "Demand side management in smart grid using heuristic optimization," *in IEEE Transactions on Smart Grid*, vol. 3, no. 3, pp. 1244–1252, 2012.

[77] J. Lu, T. Sookoor, V. Srinivasan, G. Gao, B. Holben, J. Stankovic, E. Field, and K. Whitehouse, "The smart thermostat: Using occupancy sensors to save energy in homes," in *In Proc. SenSys*, November 2010.

[78] L.Zhang, A. Lam, and D. Wang, "Strategy-proof thermal comfort voting in buildings." in *Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings*. ACM, 2014, pp. 160–163.

[79] R. Martinez, F. Ramos, and S. e. a. A. Gormus, "Comparison of centralized and distributed monitoring architectures in the smart grid," *In Systems Journal, IEEE*, December 2013.

[80] L. Mauch and B. Yang, "A new approach for supervised power disaggregation by using a deep recurrent lstm network," in IEEE Global Conference on Signal and Information Processing (GlobalSIP), 2015.

[81] M. Minou, G. hanos, M. Vasirani, T. Ganu, M. Jain, and A. Gylling, "Evaluating demand response programs: Getting the key performance indicators right," 2014.

[82] S. J. Moss, "Market segmentation and energy efficiency program design," 2008.

[83] S. N. A. U. Nambi, A. R. Lua, and R. Prasad, "Loced: Location-aware energy disaggregation framework," in *In Proceedings of ACM BuildSys*, 2015.

[84] S. N. A. U. Nambi, E. Pournaras, and R. V. Prasad, "Clustering and temporal evaluation of energy demand," 2016, arXiv preprint.

[85] S. N. A. U. Nambi, T. G. Papaioannou, D. Chakraborty, and K. Aberer, "Sustainable energy consumption monitoring in residential settings," in *In Proc. IEEE INFOCOM*, 2013.

[86] D. Pan, A. Lam, and D. Wang, "Carrying my environment with me in iot-enhanced smart buildings," in *Proceeding of the 11th annual international conference on Mobile systems, applications, and services*. ACM, 2013, pp. 521–522.

[87] N. L. Panwar, S. C. Kaushik, and S. Kothari, "Role of renewable energy sources in environmental protection: A review," *Renewable and Sustainable Energy Reviews*, vol. 15, pp. 1513–1524, 2011.

[88] O. Parson, S. Ghosh, M. Weal, and A. Rogers, "Non-intrusive load monitoring using prior models of general appliance types," in *In Proc. AAAI*, 2012.

[89] R. Paulson, C. Basu, A. M. Agogino, and S. Poll, "Inverse modeling using a wireless sensor network (wsn) for personalized daylight harvesting," *In SENSORNETS*, pp. 213–221, 2013.

[90] L. Peretto, "The role of measurements in the smart grid era," *in IEEE Instrumentation & Measurement Magazine*, vol. 13, pp. 22–25, 2010.

[91] M. A. Piette, G. Ghatikar, S. Kiliccote, E. Koch, D. Hennage, P. Palensky, and C. McParland, "Open automated demand response communications specification," 2009.

[92] D. Pisharoty, R. Yang, M. W. Newman, and K. Whitehouse, "Thermocoach: Reducing home energy consumption with personalized thermostat recommendations," in *Proceedings of the 2nd ACM Conference on Embedded Systems for Energy-Efficient Buildings*. ACM, 2015, pp. 201–210.

[93] E. Pournaras, M. Vasirani, R. E. Kooij, and K. Aberer, "Decentralized planning of energy demand for the management of robustness and discomfort," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 4, pp. 2280–2289, 2014.

[94] D. M. W. Powers, "Evaluation: From precision, recall and f-factor to roc, informedness, markedness & correlation," *In Journal of Machine Learning Technologies*, pp. 37–63, 2011.

[95] T. Prabhakar, S. N. Akshay, R. V. Prasad, S. Shilpa, K. Prakruthi, and I. Niemegeers, "A distributed smart application for solar powered wsns," *In LNCS NETWORKING*, 2012.

[96] V. Radu, N. D. Lane, S. Bhattacharya, C. Mascolo, M. K. Marina, and F. Kawsar, "Towards multimodal deep learning for activity recognition on mobile devices," in In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct, 2016.

[97] A. Rai, K. Chintalapudi, V. Padmanabhan, and R. Sen, "Zee: zero-effort crowdsourcing for indoor localization," in In Proceedings of ACM Mobicom, 2012.

[98] J. Ranjan, E. Griffiths, and K. Whitehouse, "Discerning electrical and water usage by individuals in homes," in *In Proceedings of ACM BuildSys*, 2014.

[99] J. Ranjan, Y. Yao, and K. Whitehouse, "An rf doormat for tracking people's room locations," in *In Proceedings of the ACM International Joint conference on Pervasive and Ubiquitous Computing*, 2013.

[100] D. Rech and A. Harth, "Towards a decentralised hierarchical architecture for smart grids," in *Proceedings of the Joint EDBT/ICDT Workshops*, U. Acm., Ed., 2012.

[101] E. Rendon, I. Abundez, A. Arizmendi, and E. M. Quiroz, "Internal versus external cluster validation indexes," *International Journal on Computer Communication*, vol. 5, no. 1, pp. 27–34, 2011.

[102] A. Rowe, M. Berges, and R. Rajkumar, "Contactless sensing of appliance state transitions through variations in electromagnetic fields," in *In Proc. BuildSys. ACM*, 2010.

[103] S. Rusitschka, K. Eger, and C. Gerdes, "Smart grid data cloud: A model for utilizing cloud computing in the smart grid domain," in IEEE SmartGridComm, 2010.

[104] M. Saha, S. Thakur, A. Singh, and Y. Agarwal, "Energylens:combining smartphones with electricity meter for accurate activity detection and user annotation," in *In Proceedings of 5th International Conference on Future Energy Systems*, 2014.

[105] T. F. Sanquist, H. Orr, B. Shui, and C. Bittner, "Lifestyle factors in U.s. residential electricity consumption," *Energy Policy*, vol. 42, 2012.

[106] C. Sarkar, S. N. A. U. Nambi, and R. V. Prasad, "iltc: Achieving individual comfort in shared spaces," in in International Conference on Embedded Wireless Systems and Networks (EWSN), 2016.

[107] C. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, no. 379, pp. 623–656, July 1948.

[108] M. Shargal and D. Houseman, "The big picture of your coming smart grid," vol. 2009. [Online]. Available: http://www.smartgridnews.com/artman/publish/commentary/The_Big_Picture_of_Your_Coming_Smart_Grid-529.html

[109] W. Shen, V. Babushkin, Z. Aung, and W. L. Woon, "An ensemble model for day-ahead electricity demand time series forecasting," in *In Proceedings of the fourth international conference on Future energy systems, ACM e-Energy*, 2013.

[110] R. E. Smalley, "Top ten problems of humanity for next 50 years," 2003.

[111] J. Stephenson, B. Barton, G. Carrington, D. Gnoth, R. Lawson, and P. Thorsnes, "Energy cultures: A framework for understanding energy behaviours," *Energy Policy*, vol. 38, 2010.

[112] Y. Strengers, "Designing eco-feedback systems for everyday life," in *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*. ACM, 2011.

[113] P. Tan, M. Steinbach, and V. Kumar, "Introduction to data mining."

[114] Y. Tung and K. Shin, "Echotag: Accurate infrastructure-free indoor location tagging with smartphones," in In Proceedings of ACM Mobicom, 2015.

[115] J. S. Vardakas, N. Zorba, and C. V. Verikoukis, "A survey on demand response programs in smart grids: Pricing methods and optimization algorithms," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, pp. 152–178, 2015.

[116] C. Wang, Y. Zhou, J. Wu, J. Wang, Y. Zhang, and D. Wang, "Robust-index method for household load scheduling considering uncertainties of customer behavior," *in IEEE Transactions on Smart Grid*, vol. 6, no. 4, pp. 1806–1818, 2015.

[117] M. Weiser, "The computer for the 21st century," *Scientific american*, pp. 94–104, 1991.

[118] M. Weiss, A. Helfenstein, F. Mattern, and T. Staake, "Leveraging smart meter data to recognize home appliances," in *In Proc. PerCom. IEEE*, 2012.

[119] T. K. Wijaya, M. Vasirani, and K. Aberer, "When bias matters: An economic assessment of demand response baselines for residential customers," 2014.

[120] Y. Yuan, D. Pan, D. Wang, X. Xu, Y. Peng, X. Peng, and P. j. Wan, "A study towards applying thermal inertia for energy conservation in rooms," *ACM Transactions on Sensor Networks (TOSN)*, vol. 10, p. 1, 2013.

[121] M. Zafar, I. Valera, M. Rodriguez, and K. P. Gummadi, "Fairness constraints: A mechanism for fair classification," in *Proceedings of ICML*, 2015.

[122] M. Zeifman and K. Roth, "Nonintrusive appliance load monitoring:review and outlook," 2011.

[123] Z. Zhao, W. C. Lee, Y. Shin, and K. B. Song, "An optimal power scheduling method for demand response in home energy management system," *in IEEE Transactions on Smart Grid*, 2013.

[124] J. Zhou, R. Hu, and Y. Qian, "Scalable distributed communication architectures to support advanced metering infrastructure in smart grid," 2012.

[125] Z. Zhu, J. Tang, S. Lambotharan, W. H. Chin, and Z. Fan, "An integer linear programming based optimization for home demand-side management in smart grid," in *in Proc. IEEE PES ISGT*, 2012.

ACKNOWLEDGMENTS

Doing a PhD was always my ambition and the last four years have been quite remarkable. This thesis is the result of few amazing collaborations within and outside TU Delft. While my work was solitary at times (always?), I had a good bunch of friends, colleagues and family members who supported and pushed me to achieve my ambition. I would like to sincerely thank all the people who have inspired, advised, criticized and helped me along the way.

First, I would like to thank my promotor, Koen Langendoen (Koen) and daily supervisor, Venkatesha Prasad (VP) for their unwavering support along my PhD journey. I would like to explicitly thank Koen's extra mile efforts to hire me as a PhD student without a master's degree. When doing a PhD, it is very important to work on a topic, which you are passionate about. Koen gave me a free hand to pursue a topic of my interest. Even though we did not have weekly meetings and write papers together, he has always shown interest and kept tabs on me. I have thoroughly enjoyed our coffee/tea discussions and fun-filled technical meetings. Over the years, he has played a crucial role in shaping my thinking, improving my presentation skills (koenification) and expressing my ideas/problems/solutions in a structured way. Through your Dutch pragmatic thinking and punctilious attitude, you have taught me how to become a thorough researcher. I was also lucky enough to enjoy numerous bike rides with you exploring the flat land.

I would also like to thank my daily supervisor, VP. I was very fortunate to have a daily supervisor who supported everything I ventured into. VP gave me complete freedom to pursue any problem I am interested in, collaborate with others, and buy necessary equipment. Generally a daily supervisor's role is to push his PhD students, however, in my case it was the other way around. I have pushed VP to several all-nighters to finish up the papers, interact and manage my master students, support conference trips and much more. I am truly grateful for the support and trust bestowed upon me.

Apart from my supervisors, I also had privilege to interact with several colleagues and students from EWI. I would like to thank all my colleagues from embedded software group – Andreas, Andrei, Coen, Ioannis, Marco (x2), Przemek, Qing, Qingzhi, Sinan, Stef, and Yunus – who have been great company. I want to especially thank Chayan – my partner in crime, be it to write a paper for a conference in a fancy location or to vanish for a week in greek islands; Kishor – my well-wisher, he has constantly helped me throughout my stay in Delft; Vijay – my adviser (and sounding board), he was always there if I had to share something and without his pampering and kindness I would not have completed my PhD in time. I would also like to thank Marco Zuniga; I have had the privilege to be part of his course on smartphone sensing, co-guide master students, and write amazing papers. I have enjoyed our numerous in-depth technical discussions and his attitude towards deep analysis of a problem has helped me improve the quality of my publications.

During my PhD I also interacted and worked with MSc students – Antonio, Bontor, Duncan, Johnny, Luis, and Nikos; I want to thank them for choosing me as their supervisor and for their efforts in various projects. I would also like to thank our group staff members – Janneke, Munire, Paula, Paulo, Rina, and Stephen for their support on logistics and systems

related queries (often asked last minute). This thesis would not have been possible without the funding from the iCore EU FP7 project. It was an amazing experience to be part of this project, attend project meetings at various locations in Europe and I would like to thank all the project partners of iCore. I would also like to thank Prof. Ignas Niemegeers and Fernando Kuipers who were part of my Go/No-go committee and trusted my PhD plan. I am grateful to all the committee members for agreeing to be part of my PhD committee.

I have been very fortunate to work with a bunch of wonderful external collaborators. My passion towards systems research grew when I started working in ZENLab in DESE, IISc under Prof. H.S. Jamadagni and T.V. Prabhakar. I have always admired their passion towards systems challenges and it has helped me to identify fundamental societal problems that exist just around us. I would also like to thank all the research assistants who worked with me in DESE, IISc for their constant support and motivating me to pursue a PhD. I would like to thank Prof. Karl Aberer and LSIR group members in EPFL for hosting me as an intern. My special thanks to Matteo Vasirani, he is the best post-doc researcher I have worked with till date. I would also like to thank Evangelos Pournaras; I really enjoyed our brainstorming sessions and it has significantly contributed to chapter 6 in this thesis. I am very grateful to my mentor Dr. Venkat Padmanabhan who hosted me in Microsoft Research India in the summer of 2016. I thoroughly enjoyed my time in MSR India and I would like to thank Manohar, Saikat and all my colleagues at MSR India for their support.

During my stay in Delft I was lucky to have an amazing bunch of friends – Anuradha, Geetha, Kruthika, Mihir, Nimish, Sakshi, Shruti and Wendy who have fed me and cheered me up on numerous occasions. I have greatly enjoyed the time we have spent together on various vacations and day-outs. I would also like to thank the open-minded, English speaking people of The Netherlands who welcomed me and made my stay very pleasant. I greatly admire the Dutch culture – Dutch treat, being direct and straight forward, bargain on stuff, and bicycle always (be it rainy or windy) – and hope to follow them.

Finally, I would like to thank my mother, brothers and other relatives. You always believed in me and trusted that I would do the right thing in my career. In particular, I want to thank my mother (Veda) who has selflessly cared for me. I remember the days/weeks' prior to paper deadlines where I used to go home only to get some food and sleep. I will certainly try to ensure to spend more quality time with you. My brothers (Ranga and Arjun) have been constantly supporting me in all my endeavors and have taken the responsibility to manage the family back in India. Without your love and kindness, I wouldn't have come this far!

Akshay Srirangam Narashiman
Delft, May 2017

LIST OF PUBLICATIONS

Journals:

1. **S. N. A. U. Nambi**, A. R. Lua and R. V. Prasad, Decentralized Energy Demand Regulation in Smart Homes, in *IEEE Transactions on Green Communications and Networking,* 2017.

2. **S. N. A. U. Nambi**, Evangelos Pournaras and R. V. Prasad, Temporal Self-regulation of Energy Demand, in *IEEE Transactions on Industrial Informatics, Vol. 12, Issue. 3, pp. 1196-1205,* 2016.

3. **S. N. A. U. Nambi**, and R. V. Prasad, Toward the Development of a Techno-Social Smart Grid, in *IEEE Communications Magazine, Vol. 54, Issue 11, pp. 202-209,* 2016.

4. C. Sarkar, **S. N. A. U. Nambi**, R. V. Prasad, A. Rahim, R. Neisse, G. Baldini, DIAT: A Scalable Distributed Architecture for IoT, in *IEEE Internet of Things Journal, Vol. 2, Issue. 3, pp. 230-239,* 2014.

5. V. S. Rao, **S. N. A. U. Nambi**, R. V. Prasad, I. Niemegeers, On systems generating context triggers through energy harvesting, in *IEEE Communications Magazine, Vol. 52, Issue 6, pp. 70-77,* 2014.

Conferences:

1. **S. N. A. U. Nambi**, L. A. Gonzalez, R. V. Prasad, CoachMe: Activity Recognition using Wearable Devices for Human Augmentation, in *ACM International Conference on Embedded Wireless Systems and Networks, EWSN,* 2017.

2. **S. N. A. U. Nambi**, A. R. Lua and R. V. Prasad, Decentralized Energy Demand Regulation in Smart Homes, in *59th Annual IEEE Global Communications Conference (Globecom),* 2016.

3. C. Sarkar, **S. N. A. U. Nambi**, and R. V. Prasad, iLTC: Achieving Individual Comfort in Shared Spaces, in *ACM International Conference on Embedded Wireless Systems and Networks (EWSN),* 2016.

4. **S. N. A. U. Nambi**, A. R. Lua and R. V. Prasad, LocED: Location-aware energy Disaggregation Framework, in *2nd ACM International Conference on Embedded Systems For Energy-Efficient Built Environments, BuildSys,* 2015.

5. **S. N. A. U. Nambi**, M.Vasirani, R. V. Prasad, K. Aberer, Performance Analysis of Data Processing Architectures for the Smart Grid, in *5th IEEE PES Innovative Smart Grid Technologies (ISGT) European 2014 Conference,* 2014.

6. **S. N. A. U. Nambi**, M.Vasirani, R. V. Prasad, K. Aberer, A Cost-Benefit Analysis of Data Processing Architectures for the Smart Grid, in *Wireless and Mobile Technologies for Smart Cities (WiMobCity) with ACM MobiHoc,* 2014.

7. **S. N. A. U. Nambi**, C. Sarkar, R. V. Prasad, A. Rahim, A unified semantic knowledge base for IoT. In *IEEE World Forum on Internet of Things (WF-IoT),* 2014.

8. C. Sarkar, **S. N. A. U. Nambi**, R. V. Prasad, A. Rahim, A scalable distributed architecture towards unifying IoT applications. in *IEEE World Forum on Internet of Things (WF-IoT),* 2014.