Solving Human Centric Challenges in Ambient Intelligence Environments to Meet Societal Needs

---

A Dissertation

Presented to

the Faculty of the School of Engineering and Applied Science

University of Virginia

In Partial Fulfillment

of the requirements for the Degree

Doctor of Philosophy (Computer Science)

by

Erin Griffiths

December 2019

Approval Sheet

This dissertation is submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy (Computer Science)

Erin Griffiths

This dissertation has been read and approved by the Examining Committee:

Kamin Whitehouse, Adviser

Jack Stankovic, Committee Chair

Mary Lou Soffa

A.J. Brush

John Lach

Accepted for the School of Engineering and Applied Science:

Dean, School of Engineering and Applied Science

December 2019

To everyone who has helped me along the way.

Abstract

In the world today there exists a large number of problems that are of great societal concern, but suffer from a problem called the tragedy of the commons where there isn't enough individual incentive for people to change their behavior to benefit the whole. One of the biggest examples of this is in energy consumption where research has shown that we can reduce 20-50% of the energy used in buildings if people would consistently modify their behavior. However, consistent behavior modification to meet societal goals that are often low priority on a personal level is often prohibitively difficult in the long term. Even systems design to assist in meeting these needs may be unused or disabled if they require too much effort, infringe on privacy, or are frustratingly inaccurate. We need a way to automatically meet societal needs while reducing or eliminating the burden on users.

The vision of Ambient Intelligent Environments (AmI) has been dedicated to meeting this and other human supportive goals since first being introduced in the late 1990s. Ambient intelligence environments are computational systems embedded in the physical environment that sense, reason about, and act for the benefit of the people in that environment and their objectives. The key factor, and challenge, of an ambient intelligence environment is that it serves its objectives invisibly and transparently, with little to no requirements on the user's behavior or cognitive load, but still incorporates the user in all its actions. It is not by accessing a computer terminal that people interact with the system, but by normal interactions with their environment and the objects within. This seems ideal for meeting societal concerns where users do not want to put any additional effort into the system.

This dissertation aims to explore some of the practical uses and challenges of sensing and reasoning in ambient intelligence environments. First, I show how AmI environments are needed to personalize solutions to societal needs, such as shifting the energy peak created by daily energy use. This personalization can improve widespread adoption by allowing a system to benefit every user and not just the average one. Then, I explore two novel solutions to the two main human-centric challenges in ambient intelligence environments that may prevent the adoption of solutions for meeting societal needs: *privacy vs. data collection* and *having the human in the loop vs. automation*. Privacy is challenging in AmI because personal data is needed to create

personalization, but many systems collect more private data than they need for operation. We present a novel solution to one of the worst examples of this, cameras, by using hardware limitations to preserve privacy. Keeping humans in the loop in AmI is challenging since we cannot ask for their preferences and instead need to learn through their natural interactions with the environment. In physical AmI, the interactions used to learn such preferences are often few and far between, resulting in a "small data" problem for learning. This problem is further exacerbated as preferences change over time and the system must quickly adapt. We present a novel solution that leverages the similarity between users in reinforcement learning to quickly learn in the face of these challenges. We anticipate that this work and these solutions to the two challenges of AmI will increase the accuracy, privacy, and ultimate adoption of AmI technologies to meet societal needs.

Contents

List of Tables

List of Figures

Chapter 1

Motivation: Meeting Societal Needs Through Sensors, Learning, and Control

1.1 Overview and Motivation

In the world today there exists a large number of problems that are of great societal concern, but suffer from a problem called the tragedy of the commons where there isn't enough individual incentive for people to change their behavior to benefit the whole. One of the biggest examples of this is in energy consumption where research has shown that we can reduce 20-50% of the energy used in buildings if people would consistently modify their behavior [2, 3]. But despite an increased understanding of the problems and an increase in available solutions, we still haven't achieved these savings. The main reason? It's often prohibitively difficult to get people to change their behavior consistently and over the long term to meet a societal need [4]. People's willingness to commit to behavioral changes is often based on the value they place on the outcome of that change, and whether or not that outcome is a priority in their life at any given time. Issues with school, work, and personal and family life can all outweigh a commitment to a change at any time. Additionally, while asking people to turn off lights and program their thermostats might seem small and reasonable changes, the behavioral modifications required for more complex societal needs are often not. For example, in the energy system used today, the peak in electricity demand that occurs as people consume higher amounts of electricity during the day is a significant societal problem. This peak governs the number of generators that need to

be built - an ever increasing number each year as our overall demand for electricity rises. These peaking plants are generally inactive and underused at night when power demand is lower and are less efficient/more expensive plants that are built to quickly spin up in peak situations. The only solution that currently exists to this problem, besides extensive and expensive battery storage, is asking people to change their behavior when it comes to energy usage - like taking showers, washing clothes, and doing dishes at 3 am to flatten out the usage peak. The complexity of the timing and extent of these behavioral changes will only increase as we begin to rely more heavily on dynamic and variable renewable energy sources. Given the tragedy of the commons problem and constantly shifting individual priorities, manual behavioral change will likely never happen on a large enough scale to make a difference for this issue. Hence, we need a way to automatically meet societal needs without requiring constant participation from users - reducing or eliminating the burden on users and increasing the likelihood of meeting societal needs.

The vision of Ambient Intelligent Environments (AmI) has been dedicated to meeting this and other human supportive goals since first being introduced in the late 1990s [5]. Ambient intelligence environments are computational systems embedded in the physical environment that sense, reason about, and act for the benefit of the people in that environment and their objectives. The key factor, and challenge, of an ambient intelligence environment is that it serves its objectives invisibly and transparently, with little to no requirements on the user's behavior or cognitive load, but still incorporates the user in all its actions. It is not by accessing a computer terminal that people interact with the system, but by normal interactions with their environment and the objects within. A user does not have to remember to use the system or manually provide information about themselves during operation. Because of this, both highly personal, individual services (i.e. those a person will participate for) as well as less personal societal benefits (i.e. where people may forget to participate) can be equally supported with this technology. Such systems can help to support the projected 83.7 million U.S. elderly in 2050 [6] and work to reduce the 40% of the world's energy consumed in buildings. Additionally, they can sense, learn, and act to meet preferences that feed into health, happiness and better productivity [7]. Much of the work thus far in AmI has been researching and building the infrastructure necessary to connect a person's environment to the computational resources necessary to sense, learn about, and control that environment. It is here that AmI overlaps with the growing fields of ubiquitous and pervasive computing, the internet of things (IoT), cyber physical systems, and more to create this infrastructure in a way that's fast, cost-effective, accurate, and low power. These fields also overlap in creating systems for sensing and learning about human behaviors in tasks like activity recognition and person tracking. Additionally, researchers have begun incorporating machine learning into AmI and its infrastructure to sense and learn about the people of society on a larger scale than ever before. If we can effectively leverage this sensing and control, we can begin to have a larger impact on societal needs by

seamlessly integrating computational support into people's lives.

Researchers in AmI have recently identified two major challenges to this vision above and beyond the infrastructure challenges being researched in related fields [8]. These challenges look at the adoptability of AmI in the wider human populace and focus on how to integrate sensing and control infrastructure with common human concerns and behaviors. These two challenges concern the human-centric issues of privacy and human control. Specifically, the first challenge deals with the privacy concerns of people participating in AmI environments. In general, systems that work with or around human preferences and behavior must learn about the people in that environment in order to be effective. This learning can fall on a wide spectrum and many systems in place today operate on the average behavior of people (e.g. the common thermostat temperature in public buildings is often based on the average male metabolism [9]). As technology and companies are now pushing towards more personalized services, learning the behaviors of individuals, over the average, is becoming the norm. With this increase, more private information is being recorded in the push towards obtaining more actionable data for smart services. The more individual data that is recorded, both in sample rate and sensing mode, the more personal services can become. However, much of this data must come from the private domain to fully personalize services. Hence, there is a trade-off between the need for a smart system to collect more data and the need for users to have a minimum amount of data collected to minimize privacy violations. This *privacy vs. data collection* problem is infrequently addressed in current smart systems. Much of the privacy issue is handed off to security research, where the prevention of hacking or authorization protocols are presented as the solution to privacy concerns. However, even with perfect security, users may not trust the smart system itself to preserve their privacy and many want only the minimum data possible collected for the service they're consenting to. While some data must be collected for the smart system to operate, many systems inadvertently collect more data than is necessary for their stated objective. One of the best examples of this is using visual sensors to collect data. Here, all data within the visual field can be collected and recorded, not just that which fulfills the smart objective. For example, a camera designed to collect data about who is in a space automatically also collects data about what they are doing, what they are wearing, the layout of the space, what objects are in the space, and more. Minimizing this dissonance between collected data and needed data for smart systems needs to be addressed in AmI to mitigate privacy concerns and increase adoptability.

The second challenge in AmI deals with the trade-off between human participation and automation. Like with privacy concerns, a person's willingness to participate in a system will depend on how much they value the outcome. Hence, even manual operation of a system might be acceptable if the value is high enough. Many IoT systems have the connectivity of smart systems, but require manual human operation to turn them on and off. This means humans are always in control and the system is in operation only when it is valued

and the actions are approved. On the other hand, fully automated systems often provide too little control for the user. Actions that are not personalized or that incorrectly infer preferences can frustrate or actively work against the user's objectives and preferences. If frustrations get too high, users may disable the system entirely. Since issues like societal concerns are often not consistently high value enough for manual operation, systems must use some form of automation to provide long term operation. The challenge to AmI systems is to balance this *human in the loop vs. automation* trade-off so minimal onus is placed on the user for operation, but the system still operates to meet objectives while taking the preferences and behaviors of users into account. One option is to have a user pre-program individual preferences at the start of operation to account for human input (such as having the users set the time periods of temperature on their thermostat). However, this relies on both the ability of users to articulate all their preferences in full and their ability to select the best options to meet other objectives (e.g. energy minimization). Since users often do not have the ability, knowledge, or time to determine this a priori, AmI grew to adopt machine learning techniques to learn and incorporate user preferences. This allows the system to collect samples of behavior in various situations or states to then be incorporated into decision making. The challenge with machine learning in AmI is that machine learning requires many samples of user preferences and behavior in different environmental states in order to successfully learn and predict the behavior. In the physical environments of AmI these samples may be few and far between. Systems can increase the samples by performing specific training periods (a necessary step for supervised learning) but this also increases the effort on the part of the users as a kind of "action based" preference survey and may need to be repeated frequently as users' preferences change over time. Ideally, an AmI system could learn user preferences through interactions with the environment alone. Hence, a balance between learning preferences through only natural interactions and learning them quickly and accurately enough to serve the user must be found in AmI to ensure users have influence over the system and its control but are not burdened in the process.

This dissertation aims to increase the use of AmI to support solutions to societal problems by presenting novel ways to address these challenges. We do so in three parts. First, we look at the importance of including individual human behaviors and preferences in solutions for societal needs. Specifically, we show that incorporating individual behaviors in systems that aim to deal with the energy peak can both significantly decrease the peak energy use and increase the monetary savings for the individual. These monetary savings are only possible for an individual with a personalized system, motivating the need to include users in the loop of both sensing and control to meet societal needs. With this inclusion of users and their personal information in systems, solving the two human-centric challenges of AmI becomes necessary for adoption. Hence, the second part of this dissertation addresses the privacy vs. data collection challenge with a novel approach to preserve privacy through the deliberate creation of hardware limitations for visual sensors. We focus on

limiting data collected to only that necessary for the application by purposely limiting the hardware of the camera such that, even if the camera were hacked, only the consented application data could be extracted. No other data available to the sensor (e.g. human clothing or activities) could be obtained by either the smart system or the hacker. The third part of this dissertation addresses the human in the loop vs. automation challenge with a novel approach to reinforcement learning to learn preferences. Reinforcement learning learns human preferences and the best responses to their behaviors by continuously interfacing with users and their environments naturally, making it an ideal AmI machine learning approach. Our novel addition to this technique leverages the similarity between users to speed learning in an AmI environment accounting for users' preference changes over time and the general infrequency of interactions with the system in a physical environment. We anticipate that this work and these solutions to the two challenges of AmI will increase the accuracy, privacy, and ultimate adoption of AmI technologies to meet societal needs.

1.2 Thesis Statement

Preserving image privacy through hardware limitations and learning preferences in the face of natural and infrequent system interactions in reinforcement learning can increase the adoptability and accuracy of ambient intelligence environment solutions that can meet societal needs more effectively by including personalization in their design.

1.3 Contributions

The contributions of this dissertation are the following:

- An ambient intelligence system for load shifting using thermal energy storage and water heaters called ThermalThrift useful to utilities (for load shifting) and homeowners (for monetary savings).
- An analysis of the importance of personalized behavioral information on the potential adoptability of an AmI system to shift peak electrical load.
- The collection and analysis of in-situ water use data for 10 people over 78 total days for an evaluation of the AmI peak shifting system.
- A camera based identity sensor useful for tracking residents in homes without requiring active user participation called Lethe.
- An analysis of the privacy-preserving capabilities of hardware limitations for the use of cameras in people identification.

- The collection and analysis of 1500 camera events from 21 different participants to evaluate the accuracy of a privacy-preserving person tracking camera.
- A general collaborative reinforcement learning algorithm that uses similarity information between participants to collaborate for faster learning called KindredRL.
- The creation of three simulations of blinds, thermostat, and water heating control to evaluate the learning capabilities of a reinforcement learning algorithm leveraging behavioral similarities.

1.4 Thesis Outline

The rest of the dissertation is organized as follows:

- Chapter 2 describes the state of the art technologies related to ambient intelligence environments.
- Chapter 3 describes our peak shifting water heater project that shows the importance of personalization in AmI Environments designed to meet societal needs.
- Chapter 4 describes our privacy based approach to identifying users with cameras in AmI Environments.
- Chapter 5 describes our user-similarity based approach to learning with reinforcement in AmI Environments.
- Chapter 6 concludes the dissertation by summarizing the contributions, discussing limitations, and describing future work.

Chapter 2

Background and Related Work: Ambient Intelligence Environments

2.1 Ambient Intelligence Environments

The idea of AmI was first publicly presented at the Digital Living Room Conference in 1999 by Roel Pieper. Its first mention in a publication was by Aarts and Appelo in 1999 where they noted that work grew out of ubiquitous computing, a term coined by Mark Weiser in 1988 at PARC [5, 10]. Like ubiquitous computing, AmI relies on a physically distributed, wide variety of interconnected computing devices embedded in the environment. While the terms are sometimes used interchangeably, along with pervasive computing and sometimes the internet of things, AmI emphasizes environments that are sensitive and responsive to people. AmI may leverage the technologies and interconnected computing devices of fields such as ubiquitous computing and the Internet of Things, but its focus will always be using these devices to support people in their everyday life activities. In general, AmI is characterized by systems and technologies that have [8]:

- implicit interaction: networked devices in the environment can be interacted with as they normally would in everyday life
- context awareness: sensors throughout the environment can recognize people and their situation context
- personalization: the system can be tailored to meet human needs
- adaptivity: the system can change in response to people
- anticipatory behavior: the system can anticipate human desires without manual intervention

These characteristics are achieved through three main components: sensing, learning, and control. Here, sensing provides both the implicit interaction and context awareness necessary to an AmI environment. Sensing in this manner must cover two main areas: people and non-human objects. An AmI system must understand when and how a person interacts with an object (such as opening the fridge, turning the stove on, etc. to make dinner) and how the people are oriented in the space (in fact two different people are interacting with the fridge and stove to make dinner together). Sensing embedded within objects is quickly becoming commercially available to AmI systems with the Internet of Things rapidly bringing internet connected objects that can report when and how they are used to the market. Examples here include, but are not limited to, door locks, lighting and HVAC systems, smart toasters, cars, alarm clocks, vending machines and more. This sensing can monitor the devices themselves, such as those present in manufacturing facilities, or act as a medium to monitor people and provide context to their actions and environment. In AmI, the current research and state of the art sensing comes into play when people are the ones being sensed. In general, humans are difficult to sense because of the varied and creative ways they use and interact with their space and the objects within it. Carried devices, and wearables in particular like smartwatches and RFID tags, are one of the currently growing forms of state of the art human sensing, with the ability to directly sense the people they are monitoring. However, currently, such devices do not fit into the AmI vision since remembering to carry a device requires an additional cognitive load on the user. In the future, if wearables ever becoming fully integrated with daily life such that wearing them becomes natural they may provide another avenue of human sensing in AmI. For now, AmI must rely on using the environment as either a medium or staging place for human sensing.

Human sensing often falls into three main categories: identification, tracking, and activity recognition. These three categories correspond to the who, where, and what questions necessary for contextual awareness. State of the art systems in these categories are numerous. Device-free identification systems use biometrics such as height [11], body shape [12], gait analysis [13, 14], and facial recognition [15] to identify individuals. Tracking systems often pair with some form of identity sensor for room-level tracking [11, 12] or use some other sensor, such as newer Wi-Fi tracking, to detect where a person is within a room [16, 14]. Activity recognition is generally the most complex, spanning a wide variety of activities from walking/standing/sitting to cooking, watching TV, interacting with family and sleeping. Systems in activity recognition are widely varied, such as a fall detection system for the elderly using smart tiles [17], computer vision approaches that detect walking, stretching, and interacting with objects and other people [18], and Wi-Fi based systems that detect hand gestures [19]. In general, the more complex an activity (i.e. the more an activity is composed of many smaller activities, such as cooking) the harder it is to identify as the same person can perform that activity differently every time. Some systems also use a form of infrastructure mediate human sensing,

where the people are assigned activities through only what the infrastructure senses of their actions, such as movement by differential air pressure in the HVAC [20] or bathroom use through sensing of the shower and bathroom lights [21]. All of these sensors provide the context needed to learn which actions the AmI system should take to meet user needs.

The latter two components, learning and control, are often combined in AmI. Both components aim to fulfill the personalization, adaptivity, and anticipatory behavior characteristics of an AmI system. The challenge here is to incorporate both environment (e.g. weather) and human (e.g. identity) context information to learn the desired actions to meet both human preferences and system objectives. Hence, learning often implicitly creates a control policy. This control policy can either provide direct control of the environment or control the issuing of feedback or alerts. Alerts and feedback are common in systems where the user is outside of the AmI environment, such as caretaking and ambient assisted living systems [22, 23, 24]. Here, learning focuses on understanding typical behavioral patterns and detecting any anomalies with techniques such as information fusion, feature extraction, pattern analysis and recognition, and anomaly detection. A large variety of approaches are used for anomaly detection including, but not limited to, Hidden Markov Models [22] and unsupervised classification [25]. Anomaly detection itself is tightly coupled with the accuracy of the AmI system's context detection as what could be an anomaly in the eyes of the system could also just be an normal event given some unobserved context. Overall, anomaly detection is still an active area of research [26]. In feedback based systems, one an anomalous evens has been detected an alert is sent to a caretaker.

AmI systems outside of caretaker based systems focus on better learning and predicting typical behavior and largely ignore anomalies. This systems often focus on direct control, where the system modifies the environment without a human intermediary. Here, AmI leverages other areas, such as the Internet of Things, to gain mechanical control of the environment such as connected HVAC, lighting controls, remote ovens and various other connected devices. With mechanical control, AmI then applies the behavioral predictions gained through learning and/or sensing to perform actions that meet the user's needs. These systems often combine learning and control into a computational agent that can learn, plan, make decisions, and act as one technological piece. Rule-based systems, such as fuzzy-logic controllers [27, 28], are often used to turn context and behaviors into actions with expert supplied or learned rules such as "turn off the lights if no one is in the room". Other systems use model predictive control to make guesses about the future environment to improve the long term analysis of actions and others use supervised machine learning to turn context directly into action based on context action pairs. Some systems are purely reactive to the current environment, such as lighting systems that immediately turn of lights when a person leaves. Here the challenge is in the sensing system, where missed detection of occupants can cause errors. More closely aligned with AmI, other systems

must focus on not just learning current context, but predicting future context to perform optimal actions, such as HVAC systems that take multiple minutes or hours to warm up a space in preparation for occupants. Overall, all these components work together to fulfill a particular AmI application.

2.2 Applications in AmI

As the availability of sensing and control infrastructure grows AmI and similar smart technologies have been applied to a wide variety of applications [29]. State of the art systems in AmI include applications to assist the elderly with systems such as monitoring for and alerting emergency services of falls [30], systems to prevent wandering in dementia patients [31], and alerting caretakers of changes in behavior [22, 24]. Other systems provide personalized product services, such as [32] which uses context sensors to improve media recommendations (such work music vs. fun music and detecting when a family friendly movie is needed with kids) and control the environment to set mood lighting. Numerous building control systems have also been created that use personal preferences to daylight harvesting with LEDs [27] and control temperature [33, 34, 35].

The challenge in selecting applications for AmI is that the ease at which the system can become human-centered depends on the personal value a user places in the system. Application related to human health are often highly valued, which is part of why Ambient Assisted Living (AAL) applications that focus on health and elderly AmI system have one of the most extensive bodies of research work in AmI. An AmI review in 2015 found 132 individual practical implementations of AAL with 1,048 related papers and that body of work has only grown in the following years [36]. These systems often target high privacy areas, such as homes, while lower personal value system, such as energy use systems, are only widely accepted in commercial buildings. To make the situation even more complex, different users often have different values in the application space. Some people may place so much value on monitoring their pets at home that they litter the space with Dropcams, but others may not be willing to have cameras in their private spaces even to monitor for emergency medical situations. Some users may be willing to constantly update an HVAC on their schedule to save energy, while others may not prioritize using a system to check their vital signs on a daily or weekly basis.

This tradeoff between application value and privacy or participation concerns means that lower personal value applications (e.g. those that fulfill a societal rather than personal concern) have more of a need to decrease privacy and participation concerns then systems for things like health applications in AmI. Privacy concerns tend to mean that the majority of work in AmI energy systems are in public spaces, such as HVAC and lighting control in office buildings and on campuses [37, 27, 38]. Alternately, such systems in homes are

built on minimally invasive, and typically less accurate in the face of a real environment, sensors [11, 19, 17]. Low levels of interest in participation on a consistent basis also cause problems here, particularly in the public space, as many users don't learn preferences on the fly and the benefits of the AmI system are directed to the building and not the occupants. Many state of the art systems require the user to manually input preferences for comfort such as setting thermal, visual, air quality, and humidity ranges that are acceptable [27, 35, 34]. As these values may change over time, or be hard to articulate in one sitting, consistent participation in the long term may be required to maintain personalization in these applications. Hence, many practical energy savings systems do not attempt finer grained personalization, instead relying on user averages in a very peripherally AmI way. However, as we will show in Chapter 4, personalization can greatly benefit these lower value applications such as in our example of peak load shifting. Hence, increasing privacy preservation and decreasing required participation can help bring lower individual value application into the AmI application space and further work to meet societal needs.

2.3 Privacy in AmI

Privacy in AmI is a complex and multifaceted problem. While there is no universal definition, privacy is often described as the ability of the user to choose who data should be disclosed to. Based on Westin's definition from 1967: privacy is "the claim of individuals, groups or institutions to determine for themselves when, how, and to what extent information about them is communicated to others" [39]. For any AmI system to operate, some data about an individual must be collected. For example, a system that efficiently turns lights on and off in a house must know where people are located in the light zones. Depending on the data and specific individual, this data might be considered private – meaning it should not be disclosed to others unless permission is given by the user. An individual may decide that the collection of room level data is acceptable when used only for the lighting system, but shouldn't be released to friends and family that could use that data to tell when the individual is home or away. This disclosure might also be situation dependent, the individual might want the room level data released to the fire department if their house is on fire. Many AmI systems have used the ability to select the granularity of disclosure to provide privacy, such as ambient assisted living systems that allow patients to determine the information given to care takers [22].

The private information that an individual might disclose to an AmI system involves a complex function including parameters such as how private they consider the information, their value of the system using the information, the risk given public exposure of the information, and their trust in the security of the system. The value of the system can greatly affect this decision, with emergency services or health assistive technologies for elderly or impaired uses having some of the highest acceptability rates [40]. This may in

part explain why there are many fewer works in ambient energy use assistance than ambient assisted living – the personal value is just not high enough for the privacy risks. Recent work has even begun to explore monetizing private data that users can sell to corporations or other parties, providing another possible avenue of value [41]. The privacy level, value, and exposure risk are often highly application and individual dependent. There is no general way to increase the value of a system (besides data monetization) that can affect all applications, nor is there a general way to decrease privacy and exposure risk. Hence, given that studies have found that often the largest concern for users is the misuse of the collected data, most work in AmI privacy focuses on increasing the security and therefore trust of the system [40, 42].

Current work in security includes, but is not limited to, authentication and authorization frameworks for smart objects [43], end-to-end security protocols [44, 45], and data encryption at sensor nodes [46]. This security must span the whole system, from the collection of sensor data on individual nodes, to the processing of such data in the cloud. Many systems attempt to limit security risks by processing data on-site or on-node, so that only aggregated data is passed to a central control system or the cloud [47]. Challenges here include limited processing power and storage space available to security measures, requiring security to be more lightweight compared to traditional internet solutions [48]. While security is a necessary area in AmI, and still being actively researched, it doesn't cover all privacy concerns. Especially concerns that, no matter the security measures, a system can be hacked.

If we entertain the possibility that any AmI system can be hacked (or private data may be incorrectly released by the system itself), then limiting the privacy invasion when that event occurs becomes a primary focus of privacy research. It is here that the challenge of privacy vs. smartness of AmI systems becomes most visible [49]. Given the assumption that we cannot ensure the protection of private data through security, then the best privacy protection an AmI system can offer is by not collecting private information. However, many AmI systems collect more data than is necessary to the given application and don't build this idea of limiting data collection for privacy into their design. Systems that do limit privacy concerns through the selection of data collection tend to do so through their choice of sensor. Mainly, this means deliberately creating systems that do not use audio or video sensing [22, 11]. Other systems limit data collection by aggregating data as close to the sources as possible, such as the perceptual hashing of images for facial recognition in [47] so that the image itself is not stored. But this strategy still presents the opportunity for hacking at the source to expose the image. Our work in privacy aims to allow the use of cameras, a highly information-rich but also privacy invasive sensor, by limiting the hardware capability of the sensor to account only for the AmI system's required data and thus eliminating the collection and storage of unnecessary image data.

2.4 Human Participation in AmI

The level of human participation in any system we use varies widely across applications and technology. On one hand, some systems are fully manual–like many of the connected objects in IoT that can provide users with capabilities and access but do not operate without human control. When these systems are simple they can provide the most accurate and timely control (e.g. the house lights will turn off exactly when you want them to). However, if the objectives of the system or the consequences of controls are complex (e.g. such as deciding when to turn on the HVAC to make sure the house is warm when you get home, given heating time and potential traffic) manual control becomes onerous and error-prone. Alternately, fully autonomous systems tend not to require, or even allow, human participation. While human preference might be considered at the start of operation, such as setting a "comfortable" temperature or setting a general occupancy schedule, there is often no mechanism for a user to correct or interact with the system without reprogramming it [50]. This can make autonomous systems error-prone and frustrating to the users – potentially to the point that they disable the system.

All AmI systems sit somewhere between fully manual and fully autonomous and attempt to gain the benefits of both extremes. While the goal of AmI is to not require extra human participation in the operation of the system, many systems currently use human participation to either speed learning, provide context, correct system errors, or perform recommended actions. AmI generally aims to use a hybrid approach where a user can interact with the system if needed or wanted but does not have to for the system to continue operation. This can come in the form of pre-programing rules for operation, presetting preferences, having override capabilities and/or providing feedback to the system on its operation. Even sensing in AmI can require human participation, such as devices worn to track people and their context that users may lose or forget to wear [51]. Many AmI systems solicit preferences from the users at the start of operation and allow adjustments by the users during operation [27, 37]. Others have users validate or filter the rules of the system, even those generated automatically, to ensure correct operation [22]. Often these interventions are not through a user's natural interactions with the environment, but through accessing the system through a terminal and manually setting comfort ranges or action rules. These human interventions help the system to perform well at a given task but apply a cognitive and/or physical load on the users that may not be sustainable over the long term. Like with privacy, this effort on the part of the user can detrimentally effect participation in systems that have low personal value – such as those that meet societal concerns.

Hence, creating a system that has a natural interaction mechanism that can still keep human preferences in the system loop is critical to AmI adoption for societal based systems. The AmI system iDorm uses some natural interactions, such as turning the light off at night, to determine the action to take given current

context [52]. Many other systems have used natural interactions with lights, heating, appliances, and even a persons gaze to infer preferences [53, 54, 55]. The main drawback to learning in this manner is that the system must wait for natural interactions in order to learn specific preferences. Learning these preferences is further complicated by context, as a specific interaction and context combination might occur on only a daily, monthly, or yearly basis. For example, knowing what a user prefers when it snows is not something that can be learned over the summer and, depending on the geographical location, might be something experienced only a few days a year. This has a benefit over survey based learning in that context that is not experienced by the user is ignored whereas a survey of preferences wouldn't know that it didn't need to obtain those preferences. However, these interactions might be so infrequent in the natural learning case that the system may not be able to learn preferences before either the context or the preferences themselves change. Current systems that attempt to learn preferences naturally tend to use reinforcement learning, or other learning systems like fuzzy logic with a negative reinforcement rule deletion process, to learn preferences from these interactions [52, 56]. While these systems can learn individual preferences, they tend to require many sample interactions over a long period of time and may not learn necessary preferences before those preferences change. Our work in participation in AmI aims to increase learning speed and therefore control accuracy by leveraging the similarity between users of AmI systems to effectively increase the sampling of interactions for an individual in reinforcement learning.

Chapter 3

Using AmI to Meet Societal Needs

3.1 Introduction

To evaluate the importance of including personalization in societal applications in AmI we look at the problem of the energy peak and thermal energy storage. The energy peak is a phenomenon that occurs mainly during the day when high energy system are used like HVAC and lighting systems in larger commercial buildings. This large peak in energy usage governs the use of less efficient peaking power plants and costs more energy on the grid than normal usage. Encouraging users to manually shift the time of their usage of usage is often infeasible, as energy needs to be consumed during the peak for normal business operation, or can get inconsistent compliance over the long term. Hence many solutions look at storing energy produced at off-peak hours to decouple the usage peak from the generation peak and level out the required generation. The challenge is that this energy must be stored for later use and traditionally battery storage is highly expensive.

Thermal energy storage (TES) is the practice of storing energy in a thermal mass for later use. Of the four main types of energy storage (thermal, mechanical, electrical, and chemical), TES has some of the lowest capital costs ($60/kWh), even compared to Pb-acid batteries ($400/kWh) [57]. However, converting the stored heat back into electricity results in a loss of 40-70% of the energy [57], and so TES is most commonly used for heating and cooling (HVAC) so that the thermal energy is used directly and conversion to electricity is avoided. To achieve this, TES technologies such as stored ice, chilled water, or heat bricks are used to store or remove heat during off-peak hours to meet HVAC needs during peak hours, thereby reducing a building's energy bill and reducing peak load on the energy grid [58, 59, 60]. Recent work has begun to look beyond HVAC and target water heaters for TES due to their similar thermal end use during peak hours in the form of showers, dishwashing, and washing machine use in the morning and afternoon. However, current approaches

do not address how to make water heater TES cost effective for consumers given the highly personalized and individual nature of hot water usage. Without cost-effective TES, consumers may not be incentivized to adopt the technology and the potential 7-19 GWh of storage in the nearly 100 million homes in Europe with hot water tanks remains unused [61, 62]. In comparison, the combined capacity from thermal, battery, compressed air, and flywheel storage currently available totals about 12 GWh worldwide [63]. If it could be made cost effective for consumers, the existing culture of water heaters has the potential to provide large scale grid storage at very low cost.

The main goal in water heater TES is to store enough energy in the tank to facilitate turning off the heating elements for a period during peak hours and using stored thermal energy to service demand [64, 65, 66, 67, 68]. Current studies focus on two main techniques to increase this off period: increasing the size of the tank (or adding additional tanks) and raising the temperature setpoint of the tank [69, 70]. However, there are two key challenges that prevent these approaches from being cost effective. The first challenge is *usage dynamics*: water usage patterns vary greatly from one household to the next. The second challenge is *standby loss*: the loss of heat through the walls of the tank due to imperfect thermal insulation. All TES systems suffer from some standby loss, typically discharging 0.5-1% of their stored energy per day [57], but standby loss is especially problematic for water heaters, which can discharge 11% or more (even the most expensive tanks discharge at least 3%). To make storage cost effective, standby loss must therefore be carefully managed. However, doing so is challenging because of water usage dynamics. For example, storing too much energy on a day with little peak water usage can actually increase a household's energy bill by creating excess standby losses. Similarly, more energy is required to perform load shifting for a hot shower at the end of peak hours than at the beginning because heat must be stored for a longer period. Existing solutions focus exclusively on load shifting, as opposed to consumer costs, and use a static tank size or static setpoint temperature during off-peak hours without explicitly addressing usage dynamics and standby loss [69, 70]. Their results indicate these approaches reduce peak load and total cost on average across many homes, but they do not achieve optimal performance for individual households and can even increase some households' energy costs. In this work, we analyzing whether water heaters can be cost effective TES devices for the individual consumer in an AmI setting by personalizing the operation of the system and dynamically managing standby loss based on individual water usage profiles and real time-of-use pricing schemes.

We present an AmI system that learns the water usage patterns of each household in order to create cost effective TES. We call the system *ThermalThrift*. It first builds a statistical model of the household's historical water usage over time to make predictions about future hot water demand. Then, it combines these predictions with a time-of-use (TOU) pricing scheme and a thermal model of the water tank to decide whether and how much thermal energy to store in advance of peak hours. If raising the temperature is

expected to increase total cost for the consumer, ThermalThrift does not store energy and effectively reverts to a conventional setpoint water heater. ThermalThrift can create cost effective storage in two ways: a *consumer-facing* variant that minimizes the total energy cost for the consumer, favoring storage TOU savings that exceed standby loss costs and a *utility-facing* variant that minimizes peak load such that the cost of energy for the consumer does not exceed a conventional, non-TES water heater. These two variants represent the two main stakeholders: consumers, who may install such a system to respond to TOU prices and save money, and the utility, who wants consumers to install such a system to reduce peak load.

To evaluate ThermalThrift, we collect water usage data from 6 pairs of participants over periods of 2-3 weeks each, for 78 total days. The collection of this empirical data was essential to our evaluation because the goal of ThermalThrift is to provide customized performance for each household in an AmI fashion, and it therefore cannot be evaluated in simulation using the ASHRAE domestic hot water consumption profile [71] used by other studies [69, 70]. Additionally, we collect 12 different TOU pricing schedules from existing electric utilities. Using these 6 usage datasets and 12 pricing schemes, we analyze ThermalThrift's use of TES for personalized, and therefore cost effective, load shifting. Results indicate that after only 7 days of training data for a given household, ThermalThrift is able to achieve within 5% and 36% of the optimal cost and peak load reduction. The consumer-facing ThermalThrift reduces consumer cost by 25% and peak load by 47% on average, using individualized storage temperatures ranging from 51°C to 93°C. Additionally, the utility-facing ThermalThrift reduces peak load 62% for water heating without increasing energy costs for consumers. For utility companies looking to reduce peak load, ThermalThrift's savings translate to a potential collective 1GWh of peak load reduction for approximately 500,000 homes. Additionally, ThermalThrift never increases costs for the studied households – indicating that by including personalized information and converting this application to an AmI system we are able to meet a societal need in a way that would encourage adoption in individual users.

3.2 Related Work

Work related to cost effective TES in water heaters can be roughly categorized into three areas: cost effective TES approaches in HVAC and other applications, leveraging existing stored energy in water heaters, and current TES approaches to load shifting with water heaters.

The foundation of cost effective TES is the design of efficient storage materials that minimize energy loss regardless of when energy is generated or consumed. Such materials are often analyzed with respect to variable renewable energy generation, such as solar or combined heating, cooling and power plants, and physical environment of the buildings or spaces where the storage operates [72, 73, 74, 75]. When the consumption of

energy can be controlled–a standard occurrence in buildings with HVAC–techniques such as optimization and model predictive control (MPC) have been shown to save energy, shift peak load, and reduce building operating costs by choosing when to store energy and managing thermal leakage [59, 76, 77, 78]. For example, Oldewurtel *et al.* combined model predictive control with weather predictions to improve energy efficiency in climate control by storing thermal energy in advance of weather changes and Ma *et al.* uses MPC to minimize costs of precooling tank water for a water cooled A/C system subject to TOU pricing on a college campus [79, 80]. However, space heating and cooling is quite different than water heating because the latter is more dependent on occupant behaviors, which are different for each water heater. Effectively using MPC to apply TES for HVAC is very dependent on building an accurate model of the building and having weather information. Occupant usage patterns affect heating and cooling demand, but only by 15-30% at most [81, 82] and therefore existing studies do not account for the effects of dynamic occupancy when shifting peak heating or cooling load. Hence, these MPC HVAC system rarely collect the human context found in AmI systems. Applying MPC to water heaters also requires a thermal model, but the models aren't as unique as buildings and their HVAC systems. In contrast, the key to success is modeling the occupants' hot water usage patterns. Unlike weather predictions, which can be found online, occupant usage patterns are different in every building and must be learned. In this work, we evaluate how quickly ThermalThrift can learn personal water usage patterns in a AmI environment to make TES in water heaters cost effective for consumers.

Several studies have explored the use of thermal energy stored during normal operation of a tank (49-60°C) for load shifting. In these approaches heating is simply turned off during a period of peak hours, without first charging the tank with extra heat, and the tank coasts on whatever energy is already in the tank [64, 65, 66, 67, 68]. This approach is currently used by many utility companies today. However, this approach allows the temperatures to drop below user setpoints when coasting through peak periods. Therefore, it can and often does affect user comfort, e.g. when using the shower during peak hours, making it not a human-centric AmI system. Research in this area focuses on minimizing or adjusting the resulting payback period when water heaters turn on at the end of peak hours [83]. Some work uses a fuzzy logic controller to decide when to turn off the heating elements, while Kepplinger et. al. used linear optimization to control water heaters based on stock exchange prices [67, 64, 65]. Since no additional energy is used for TES, cost effective TES is not considered in these approaches.

A recent white paper by EPRI and study by Lacriox were among the first to evaluate water heaters for TES by actively charging the tanks with more energy than is required for normal operation. The tank setpoints were set to arbitrarily high temperatures (92°C and 77°C respectively) during off-peak hours to achieve longer coasting periods [69, 70]. They also studied the use of extra large tanks in order to increase storage capacity [70]. However, these approaches use static tank sizes and temperatures during off-peak hours

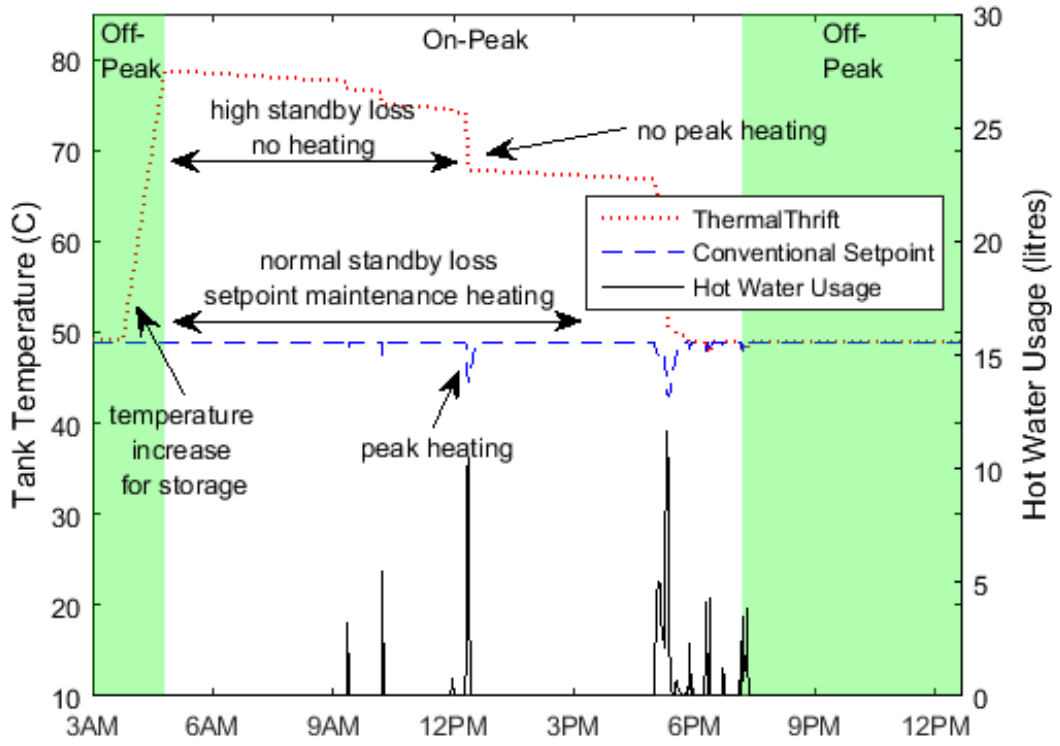


Figure 3.1: ThermalThrift stores thermal energy in off-peak hours (solid, edges) in preparation for usage during peak hours (white, center). ThermalThrift prevents the purchase of energy during peak hours while higher temperature stored energy is available. A conventional water heater purchases energy during the entire peak period.

and do not actively manage standby loss or usage patterns. Because of this, an excess of thermal energy could be stored. This may increase costs for the consumer and render TES not cost effective. This paper is the first to actively create cost effective TES using water heaters by demonstrating that hot water usage patterns can be used to better manage standby loss.

3.3 ThermalThrift: An Ambient Intelligence Water Heater

ThermalThrift achieves cost effective TES with water heaters by leveraging TOU pricing to buy cheaper energy during off-peak hours, store it thermally in the tank, and consume it during more expensive peak hours. ThermalThrift stores thermal energy by increasing the tank temperature above a *conventional setpoint* (49-60˚C) to a *TES temperature* (60-93˚C). Any hot water used during peak hours reduces the tank temperature, eliminating the need to consume additional energy until the temperature drops back to the conventional setpoint. This process is illustrated in Figure 3.1. This approach requires two related parameters

to be defined: the *storage start time* j_{TES} and the *TES temperature* t_{TES}. ThermalThrift must select the values of these parameters to shift peak load and be cost effective for the consumer. If j_{TES} is too early or t_{TES} too high, thermal storage costs due to standby loss will outweigh savings during peak hours and cost more to the consumer than conventional setpoint heating. If j_{TES} is too late or t_{TES} too low, the opportunity to shift peak load will not be utilized to the fullest and the energy peak may not be reduced significantly. Therefore, ThermalThrift has two variants that select these parameters to fulfill one of two objectives: minimizing costs for the consumer (*consumer-facing*) or minimizing peak load for the utility while not increasing costs for the consumer (*utility-facing*).

Both ThermalThrift variants optimize the predicted costs of TES, based on historical hot water usage data and a model of tank operation, to select these parameters in a process similar to model predictive control. In the consumer-facing variant, ThermalThrift implicitly finds j_{TES} by evaluating the predicted costs of TES at each time step moving forward. When the predicted cost of TES falls below that of a conventional setpoint heater, storage begins. In the utility-facing variant, ThermalThrift explicitly selects t_{TES} by predicting the cost of a sweep of TES temperatures and choosing the temperature that shifts the most peak load without exceeding the predicted cost of a conventional setpoint heater. ThermalThrift then calculates the storage start time, j_{TES}, needed to reach t_{TES} when peak hours begin and starts storing thermal energy at that time.

The details of ThermalThrift's consumer-facing and utility-facing variants are described below. While both approaches use different methods to select j_{TES} and t_{TES}, they share common components for predicting heating costs. Hence, the following sections describe the common water heater tank model, derived from the physical properties of water heaters, as well as the novel components specific to the consumer-facing and utility-facing approaches to TES.

3.3.1 Water Heater Tank Model

To predict the cost of TES, ThermalThrift estimates the amount of heating that will be required in the future through a series of modeling constraints. To ensure predicted costs accurately reflect the usage in a household, ThermalThrift estimates the costs based on the historical hot water usage of that household. This usage is stored in a matrix U, where each row is a different day of historical data and each column is a time step over the course of that day. Hence, $U_{i,j}$ is the amount of how water drawn from the tank (litres) for use on day i in time step j. The main task of prediction is to convert this historical data U into an estimated heating requirement matrix H based on a model of the tank and its operation in either the consumer-facing or utility-facing variant. Depending on the variant, the matrix H can reflect a variety of control decisions

Table 3.1: The notation used to model ThermalThrift and predict usage, heating, and costs.

Tank Parameters	
g_{tank}	tank water size (L)
s_{tank}	tank surface area (m^2)
m_{tank}	tank water weight (g)
R_{tank}	tank insulation value (e.g. R-1.8, m^2°C/W)
t_{conv}	comfortable user temperature (°C)
p_{tank}	heating power of the tank (W)
t_a	ambient air temperature (°C)
t_c	cold water temperature (°C)
c_w	specific heat of water (4.186 J/g°C)
$maxheat$	maximum possible change in temperature (°C/sec)
Input Data	
U	historical hot water usage (L)
$\vec{p_{TOU}}$	vector of electricity prices ($ /Wh)

Prediction Notation	
S	temperature of tank at each interval (°C)
H	change in temperature due to heating (°C)
W	hot water drawn from the tank (L)
L	change in temperature due to thermal losses (°C)
T	temperature of tank after losses and mixing (°C)
n	length of prediction horizon
m	interval where peak hours begin, $m < n$
τ	time step interval (seconds)
Control Notation	
c	consumer-facing TES start control ({0,1})
d	consumer-facing TES delay control ({0,1})
t_{TES}	TES temperature (°C)
j_{TES}	TES start time

(e.g. TES at a specific start time j_{TES} or conventional heating) and the control decision that produces the optimal cost for that heating is selected. Each tank model can be individualized to a home though the tank parameters (size, insulation, surface area, heating power, ambient air temp, cold water temp, etc.), listed in Table 3.1. The water tank model, represented in equations 3.1-3.6, is derived from the physical properties of water heaters and applies to any electrically heated tank.

To determine the predicted heating H, ThermalThrift implements six matrices to model the operation of the tank: S, H, W, U, L, and T. The matrices S, W, L, and T represent the predicted tank temperature in each interval, the predicted hot water drawn from the tank, the temperature change due to standby loss, and the tank temperature before heating respectively, with entries corresponding to the matrices U and H. Hence, each matrix holds predicted values for all historical days (rows, i) and each time period (of τ seconds) during those days (columns, j). The tank's state (i.e. its temperature) at the end of each time period is represented by the matrix S. Each entry $S_{i,j}$ is calculated from the previous temperature $S_{i,j-1}$ by calculating the temperature in the tank after losses and usage, $T_{i,j}$, and adding the necessary heating $H_{i,j}$. ThermalThrift calculates $T_{i,j}$ from the temperature change due to standby loss $L_{i,j}$ and the influx of cold water $W_{i,j}$ (of temperature t_c) entering the tank due to hot water usage given the size of the tank g_{tank}. A flow diagram for the interactions of these matrices is shown in Figure 3.2. Formally:

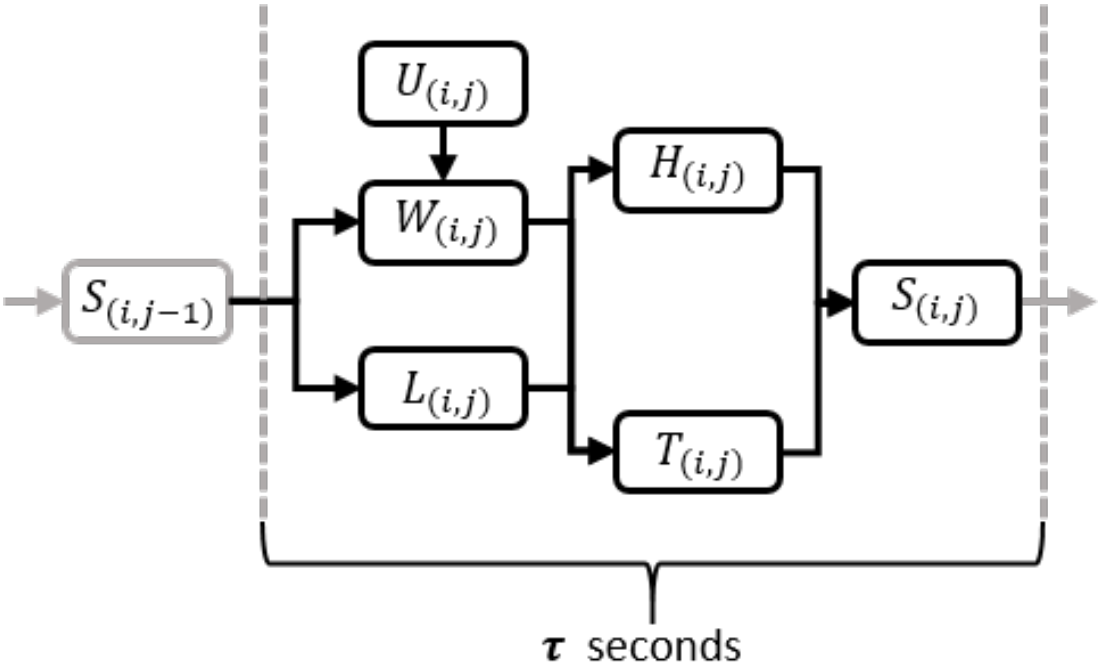


Figure 3.2: ThermalThrift predicts the state of the tank $S_{i,j}$ though a series of modeling constraints. The predicted hot water use $W_{i,j}$ is calculated from historical use $U_{i,j}$ and the previous tank temperature $S_{i,j-1}$. Thermal losses $L_{i,j}$ are also calculated from $S_{i,j-1}$. These losses $T_{i,j}$ are then used to predict the heating requirement $H_{i,j}$ and the final tank temperature $S_{i,j}$ for the τ second interval.

$$T_{i,j} = \frac{(S_{i,j-1} - L_{i,j}) * (g_{tank} - W_{i,j}) + W_{i,j} * t_c}{g_{tank}}, \forall i, j \tag{3.1}$$

$$S_{i,j} = T_{i,j} + H_{i,j}, \forall i, j \tag{3.2}$$

where $S_{i,0}$ is initialized to the current measured temperature of the tank. The upper tank temperature is bounded below t_{limit} (conservatively 93°C) to prevent boiling by:

$$S_{i,j} <= t_{limit}, \forall i, j \tag{3.3}$$

All components of equations 3.1 and 3.2 (H, L, and W) must be calculated to predict the tank state for the next interval.

The component L, the change in temperature due to standby loss, is calculated from the predicted standby loss rate over the time interval of length τ. The loss rate is calculated from the tank surface area (s_{tank}), the insulation rating (R), the ambient air temperature (t_a), the specific heat of water (c_w), and the mass of the

tank water (m_{tank}) as shown in:

$$L_{i,j} = \frac{s_{tank} * (S_{i,j-1} - t_a)}{R * c_w * m_{tank}} * \tau, \forall i, j \tag{3.4}$$

This equation shows that higher tank temperatures cause proportionally more standby loss, hence why tank setpoints are often kept in the lower thermostatic range (49-57°C). Additionally, this increase in loss at higher temperatures highlights the need for standby loss considerations in TES.

The component W represents the amount of water needed from the predicted tank to fulfill the historical hot water use U. If no thermal energy is stored, the predicted usage $W_{i,j}$ equals the historical usage $U_{i,j}$, since any needed hot water is drawn directly from the tank. However, high TES temperatures storing thermal energy (above 60°C) are too hot for direct human use due to potential scalding. To provide comfortable temperatures, thermal storage water is mixed with cold water from the mains before being dispatched to the consumer. Hence, when thermal energy has been stored, W is calculated from the mix of cold water (of temperature t_c) with predicted tank water (of temperature $S_{i,j-1}$) to produce a comfortable t_{conv} temperature for the user. Formally:

$$W_{i,j} = \frac{t_{conv} * U_{i,j} - U_{i,j} * t_c}{S_{i,j-1} - t_c}, \forall i, j \text{ when } S_{i,j-1} > t_{conv} \tag{3.5}$$

While the S equations 3.1, 3.2 and 5.2, the L equation 3.4, and the W equation 3.5 are common to both the consumer-facing and utility-facing variants, the heating component H and overall control mechanisms differ in each variant. Hence, the component H and the final optimization for each variant are described in the following sections. To simplify the equations in both variants, we define a parameter *maxheat* to be the maximum rate of change in temperature (°C per second) possible due to heating. *maxheat* is calculated from the water mass (m_{tank}) and heating power (p_{tank}) of the tank and the specific heat of water (c_w):

$$maxheat = \frac{p_{tank}}{m_{tank} * c_w} \tag{3.6}$$

In both variants, when ThermalThrift stores energy for TES it uses the maximum heating possible (i.e. *maxheat*) to do so, since slower heating provides more time for energy to be lost due to standby. How each variant selects the start time j_{TES} or temperature t_{TES} of thermal storage is describe below.

3.3.2 Consumer-Facing Cost Optimization

The consumer-facing ThermalThrift uses TES to minimize the cost of water heating for the homeowner. Though the goal of this variant is cost reduction, the TOU pricing monetarily rewards reductions in peak

energy. Hence, minimizing costs for the consumer also tends to reduce peak energy consumption. However, if the cost of peak power is only slightly larger than off-peak, the increased standby loss due to higher temperatures may mean TES is predicted as not cost effective. In this situation, the consumer-facing ThermalThrift will neither store energy nor shift peak load, but will operate as a conventional water heater.

The consumer-facing approach implicitly selects the start time j_{TES} of storage by evaluating the cost of TES at each time period (τ seconds) of its operation before peak hours. At each time step, ThermalThrift optimizes the average predicted cost of the heating component H given the TOU price of electricity $p_{\vec{TOU}}$, subject to the tank modeling Equations 3.2, 5.2, 3.4, 3.5 and its specific constraints on the heating component H. The objective function can be formally stated:

$$\min_{c} \; avg((\frac{H * m_{tank} * c_w}{3600}) * p_{\vec{TOU}}) \tag{3.7}$$

Each time ThermalThrift predicts costs, the values of component H can represent three different control decisions: store thermal energy now ($j_{TES} = 1$), store it in the next time period ($j_{TES} = 2$), and do not store thermal energy. If $j_{TES} = 1$ is optimal, storage begins immediately. If either $j_{TES} = 2$ or no thermal storage is optimal, ThermalThrift maintains the conventional setpoint until the next time period, where it again evaluates the costs of TES. These three control decisions are represented in the prediction equations for H using two control parameters: the immediate control $c = \{0, 1\}$ and the delay control $d = \{0, 1\}$. To store thermal energy now, c must be 1. To store energy in the next time period, $c = 0$ and $d = 1$. To never store thermal energy, $c = 0$ and $d = 0$. Only the control decision c is applied to the tank the optimization of the cost of H is complete. While the delay decision d is predicted for the current optimization, the next time step's optimization may make a different decision due to new usage information.

The modeling of H is divided into four equations: heating predictions in the first time step, heating in the second time step, heating from the third time step to the start of peak hours, and heating during peak hours. Storage heating is performed in the first three equations. The fourth equation models the effect of storage on required heating during peak hours. For each time period j in the equations, H must hold either the heating amount required for storage or for maintaining the comfortable setpoint t_{conv}. Storage heating is denoted with the maximum possible change in temperature $maxheat * \tau$. Maintaining a comfortable setpoint is denoted with the temperature change $max(0, min(\ maxheat * \tau, t_{conv} - T_{i,j-1}))$, where the conventional temperature is maintained after temperature loss to the ability of the heating elements. Hence, for the first H equation representing the first interval $j = 1$, either storage heating is performed or the conventional temperature is maintained according to:

$$H_{i,1} = c * maxheat * \tau + (1-c) * max(0, min(maxheat * \tau, t_{conv} - T_{i,1})), \forall i \tag{3.8}$$

In the second equation and prediction interval, heating could either continue storage heating from the previous interval ($c = 1$), have delayed the start of TES heating to this interval ($c = 0$, $d = 1$), or be maintaining a conventional temperature ($c = 0$, $d = 0$). Formally:

$$H_{i,2} = c * maxheat * \tau + (1-c) * (d * maxheat * \tau + (1-d) * max(0, min(maxheat * \tau, t_{conv} - T_{i,2})), \forall i \tag{3.9}$$

The third heating equation either continues increasing storage or maintains a conventional temperature until peak hours being. The beginning of peak hours is denoted by the interval m, when the price of electricity increases over the current price. Even when the current price is considered "peak", if the next price change increases the price then TES will be evaluated. This ensures that even 3-tiered peak systems (off-peak, peak, and super-peak) are evaluated for cost effective TES. Hence, until peak hours are reached at interval m, heating follows the equation:

$$H_{i,j} = max(c,d) * maxheat * \tau + (1 - max(c,d)) * (max(0, min(maxheat * \tau, t_{conv} - T_{i,j})), \forall i, 2 < j < m \tag{3.10}$$

Once peak hours being in interval m, thermal energy storage ceases since any electricity bought during these periods will not reduce peak consumption and will be more expensive. However, evaluation of heating costs extends past the start of peak hours m to the time step where the price of electricity returns to or falls below the current price (i.e. the prediction horizon), denoted as interval n. This ensures that the predicted cost of TES includes the subsequent monetary savings during peak hours. Hence the fourth H equation predicts the cost of only maintaining the conventional setpoint in the period between the start of peak hours m and the end of peak hours n. Formally:

$$H_{i,j} = max(0, min(maxheat * \tau, t_{conv} - T_{i,j}),$$
$$\forall i, m \leq j < n \tag{3.11}$$

We end prediction at interval n, often far short of the end of a day, since there will be no monetary savings from the current TES decision after this point – electricity can be bought for immediate use at off-peak prices without the standby loss due to storage. Each time prediction takes place for the new time period, n and m are recalculated to reflect the new price-change times for that prediction. This allows pricing schemes with two peaks (e.g. one in the morning and one in the evening) to have both peaks evaluated for TES.

In the consumer facing approach, ThermalThrift stores thermal energy when costs are predicted to be lower than a conventional setpoint heater. As a byproduct of the TOU pricing, this also tends to shift peak load. For the utility-facing approach, described below, shifting peak load is the focus of prediction – provided TES is still cost effective.

3.3.3 Utility-Facing Peak Optimization

The utility-facing ThermalThrift takes advantage of TES to reduce peak energy consumption at no more cost to the consumer than their daily, non-TES average. The goal of peak load reduction, and not cost minimization, allows ThermalThrift to store more thermal energy than the consumer-facing approach and cover more usage during peak hours – at the cost of higher standby losses and overall costs. To prevent costs from exceeding a conventional heater, the utility-facing approach incorporates a constraint to keep predicted TES costs below the predicted cost of a non-TES water heater to ensure storage is still cost effective.

The utility-facing variant optimizes for the TES temperature, t_{TES}, for each peak period during a day. The TES temperature designates the amount of thermal energy to have in storage when peak hours being. Unlike the consumer-facing approach, the prediction evaluation for this control decision is performed only once per peak period – rather than on an interval by interval basis. This allows the utility-facing variant to select the TES temperature that optimizes load shifting, while ensuring costs do not exceed the predicted conventional setpoint heater costs for that entire off-peak and peak period. Prediction is performed either when the day starts, or at the start of the off-peak hours before a peak period. Once the TES temperature is chosen, ThermalThrift calculates the storage start time j_{TES} necessary to reach the temperature and beings energy storage at that time.

To predict what t_{TES} temperature will save the most peak energy while remaining cost effect, ThermalThrift optimizes over a sweep of possible TES temperatures. The TES temperature that is predicted to shift the

most energy on average, when evaluated with predictions from k days of historical data, is selected using this objective function:

$$\min_{t_{TES}} \frac{\sum_{i=1}^{k} \sum_{j=m}^{n} H_{i,j} * m_{tank} * c_w}{k} \tag{3.12}$$

To ensure any chosen t_{TES} is cost effective, the average predicted cost of heating the tank must be below the average predicted cost of operating a conventional setpoint heater. For brevity, the predicted cost of operating a normal water heater on each historical day is denoted $cost_{conv}$. Hence, the main constraint that ensures the selected TES temperature is cost effective is:

$$avg((\frac{H * m_{tank} * c_w}{3600}) * \vec{p_{TOU}}) \leq avg(cost_{conv}) \tag{3.13}$$

To ensure comfort for the user, t_{TES} must be at or above the comfortable water temperature (e.g. 49 °C):

$$t_{TES} \geq t_{conv} \tag{3.14}$$

Additionally, the t_{TES} temperature must be reached just before peak hours begin in interval m. Formally:

$$S_{i,m} = t_{TES}, \forall i \tag{3.15}$$

The heating component H can take on values representing each of the t_{TES} temperatures in the temperature sweep subject to the constraints in Equations 3.13, 3.14, and 3.15. The modeling of H is divided into four equations: predicted conventional heating up to the start of storage, heating for storage, ensuring storage exactly reaches t_{TES}, and conventional heating during peak hours. Because each historical day of data might require slightly different start times to reach t_{TES}, due to hot water usage during energy storage, the storage start time of each predicted day is denoted by the vector $\vec{s}$. Before this start time for each day the tank is predicted as heating only to maintain a conventional temperature in the first H equation:

$$H_{i,j} = max(0, min(maxheat * \tau, t_{conv} - T_{i,j})),$$
$$\forall i, 1 \leq j < \vec{s_i} \tag{3.16}$$

The next two H equations model the heating between this $\vec{s_i}$ interval for each predicted day and the start of the peak period m. First, the tank is heated with $maxheat$ until the m-1 interval just before peak hours. Then, if less than $maxheat$ is required in the m-1 interval, only enough heating is used to reach the TES temperature. Formally:

$$H_{i,j} = maxheat * \tau, \forall i, \vec{s_i} \leq j < m - 1 \tag{3.17}$$

$$H_{i,m-1} = t_{TES} - T_{i,m-1}, \forall i \tag{3.18}$$

The final H equation for the utility-facing approach models conventional setpoint heating during peak hours. As in the consumer-facing approach, this prediction only lasts to the end of the peak period, n, since energy bought after this period will have the same cost as the energy being stored. Formally:

$$H_{i,j} = max(0, min(maxheat * \tau, t_{conv} - T_{i,j})),$$
$$\forall i, m < j < n \tag{3.19}$$

Given these constraints on the prediction model, the utility-facing variant chooses the optimal TES temperature for load shifting while not increasing costs. If costs cause $t_{TES} = t_{conv}$, then ThermalThrift acts as a conventional water heater and performs no thermal storage. As a byproduct of TOU pricing, reducing peak load also tends to reduce consumer costs and allows ThermalThrift to shift load cost effectively.

3.4 Experimental Setup

To assess the potential of ThermalThrift to cost effectively shift peak load in an AmI environment given dynamic usage and standby loss in TES, we collected data on real world hot water usage in a test home and TOU pricing from various utility companies across the US. To ensure we had behavioral variation in our water usage data, we collected data from 6 different pairs of participants. We used this historical data and TOU pricing along with our approach to control a water heater modeled by the energy simulation software TRNSYS [84].

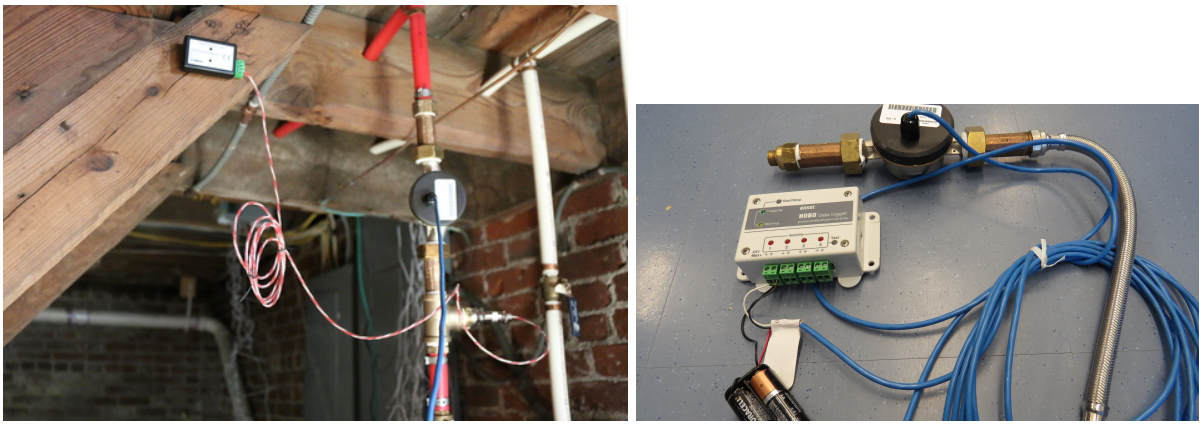


Figure 3.3: An inline water sensor (bottom) with attached data logger collects second granularity data on water drawn from the hot water tank (top) in the test home.

3.4.1 Usage Data

Hot water use traces were collected *in-situ* from 6 pairs of participants (G1-6) living in an instrumented test home (5 pairs for 2 weeks, the 6th for 3). The test home had all the hot water fixtures common to a residential home (kitchen sink, 2 bathroom sinks, dishwasher, washing machine, and shower). To ensure the data had the personal usage patterns of the participants, participants were encouraged to live in and use the home as they normally would use their own. The study had IRB approval and each participant received a $100 incentive. Due to a sensor malfunction the third participant pair, study group G3, has only 8 days of collected data. In total, 78 days of usage data is used in this work.

Usage data was collected externally to the water heater as hot water left the tank. A Seametrics SEA Series Turbine Flow Meter, shown in Figure 3.3, recorded the flow of hot water as it was used by the participants. Flow data was collected using a Hobo Data logger at a sample rate of 1Hz and stored locally. Due to the length of the in-situ studies, and the storage limits of other sensors in the home, data stored on the logger was extracted manually at two week intervals.

The usage data collected from the participants varied widely in the amount of water, type of fixtures, and time of use. For example, bathroom sink usage ranged from 62 to 359 uses across groups and the dishwasher ranged from 0 to 12 uses. Figure 3.4 shows the average hourly usage for each group. Most of the participants' usage follows expected peak patterns, peaking in the morning and evening. However, personal variation is visible in group G4's usage peak in the early afternoon.

3.4.2 TOU Pricing

Pricing schedules for the evaluation were taken from real world TOU pricing schemes from utilities across the United States. The 12 TOU schemes we selected to represent a diverse set of possible pricing schedules

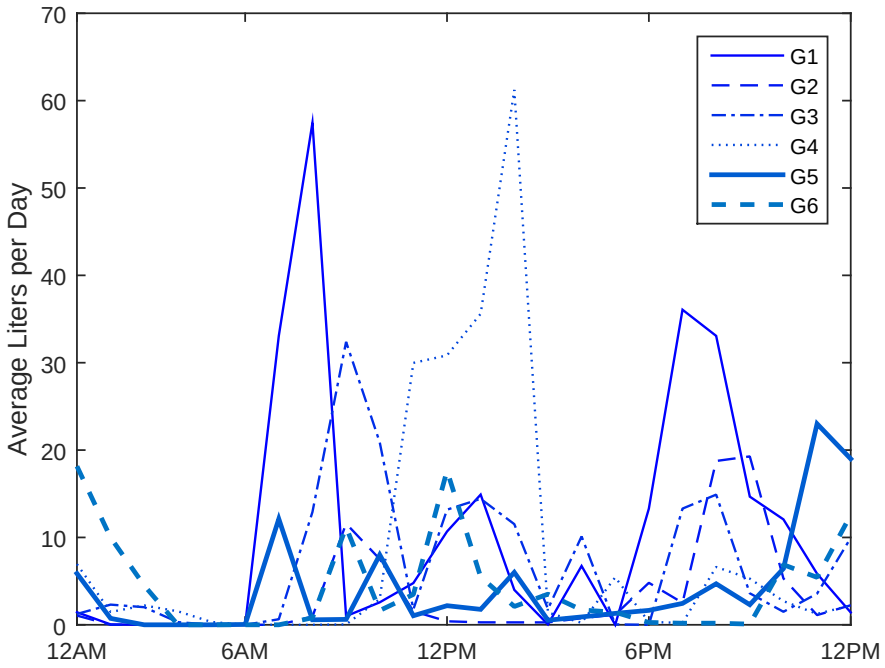
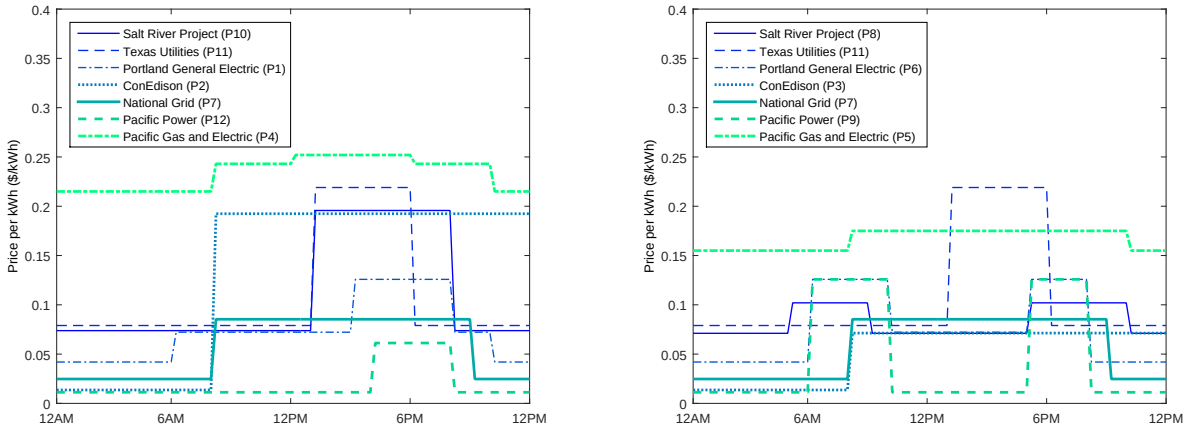


Figure 3.4: Participants' hot water use generally peaks in the morning and evening, though group G4 used the majority of their hot water around noon.

and from different geographic locations. Two power utilities (National Grid and Texas Utilities) have year round schedules, where pricing does not change according to the season. The 5 other utilities have seasonal pricing, where prices and peak power tiers change. Many of the utilities use a 2 tier peak pricing system, with off-peak and on-peak pricing. Others use a 3 tier system with off-peak, mid-peak, and on-peak pricing (sometimes called off-peak, on-peak, and super-peak). The pricing schedules we chose have a variety of each of these properties. For simplicity, the pricing models are hereby referred to as P1-12 in descending order of average conventional water heater peak load used by the study groups (G1-6). In general, the lower the pricing number (P1), the more hot water averaged across the study groups is used during peak hours. The prices themselves can be seen for the summer in Figure 3.5a and winter in Figure 3.5b. The peak price increase ranges from \$0.02 per kWh (P5) to \$0.18 per kWh (P2). The prices and times were obtained from each utility's website.

3.4.3 Baseline and Optimal Algorithms

We compare against two baselines. First, we model a *conventional* water heater that maintains a constant temp of 49°C. The 49°C temperature was chosen as our baseline because it is traditionally recommended as the most energy efficient setpoint for conventional residential water heaters in the U.S. Any lower, and the



(a) Summer. (b) Winter.

Figure 3.5: Many summer TOU pricing schedules have peak hours that span the entire daytime period from 8am-9pm. Some winter TOU pricing schedules have two peaks: one in the morning and one in the evening.

bacteria Legionella would be capable of growth in the tank. Any higher, and standby loss would increase, thereby increasing the amount of energy used to maintain the temperature. In many European countries and Canada, 60 °C is the regulated setpoint temperature with a required mixer for lower temperature delivery, since the bacteria Legionella dies after 30 minutes at this temperature. Despite this, we chose 49°C as our baseline because of the prevalence of hot water tanks in the U.S. and other countries, it's lower consumption of peak energy due to the lower setpoint, and because, with high temperature TES storage, ThermalThrift could be modified to also consider holding a temp of 60°C or higher for 30 minutes to kill the bacteria – allowing a 49°C setpoint temperature to reduce costs while still providing sanitary tank water. Both the consumer-facing and utility-facing approaches are compared to this baseline. If the consumer-facing approach chooses never perform TES, it duplicates the baseline results. If the utility-facing approach chooses a 49°C TES temperature, it also duplicates the baseline results.

Second, we compare against a *usage agnostic* version of ThermalThrift. This baseline represents approaches that statically choose high TES temperatures for off peak hours. However, the agnostic approach only heats to this temperature just before peak hours to minimize easily avoidable off-peak standby loss. Hence, the usage agnostic approach is a stronger baseline than current static approaches, but still does not account for individual household usage. The highest temperature of 93°C is chosen as the pre-peak temperature to represent the main motivation for this type of TES: load shifting. Like ThermalThrift and the conventional baseline, it maintains a minimum comfortable temperature of 49°C.

The results for both approaches are evaluated against their ThermalThrift *optimal* values. We define optimal by using the day being evaluated as the only historical day in U when an optimization is performed.

Essentially, the optimal evaluation has perfect prediction for future hot water usage as it is modeling from oracle data.

3.4.4 Water Heater Simulation

We used the TRNSYS energy simulation software developed by the University of Wisconsin Madison to evaluate ThermalThrift's control on a water heater [84, 85]. We used the Type4a water heater (non-stratified with no hot water inlet) provided with the TRNSYS version 17 library for our simulation. The water heater is modeled after the water heater present in the test home. It uses a 151 litre tank, insulation of $R = 2.6$, $4.5kW$ heating elements, and a conventional or comfortable temperature of $t_{conv} = 49$°C. As the tank was situated in the basement, an ambient air temperature of 10°C is used and cold water entering the tank is set to 10°C. Overall, the tank represents a typical, well insulated residential water heater used in a 1-3 person home.

For the TRNSYS simulation, TES controls are applied to the TRNSYS water heater at each time step according to the two ThermalThrift variants. The electrical use of the TRNSYS water heater and the cost of that usage is recorded. Then, ThermalThrift's control is recalculated by the ThermalThrift variants and applied for the next time step. For our evaluation we chose to use a 2 minute interval since ThermalThrift models the tank as a discrete system. We found no appreciable difference in results at smaller intervals.

3.5 Results

We evaluate cost effectiveness of an AmI/personalization based water heater TES with ThermalThrift by replaying the 2-3 week traces from the participant datasets. Each day is evaluated in order, with the temperature at the end of one day starting the next to ensure any standby loss across multiple days is accounted for. All days in a dataset, except the current day being evaluated, are used as historical data in the matrix U in equation 3.5 for learning. Each dataset has a different number of days, and hence each dataset uses a different number of historical days in the evaluation. The minimum number of days the approaches learn on is 7 in dataset G3. The maximum number is 19 days for G6. ThermalThrift's results using this historical data for both the consumer-facing and utility-facing variants are called *learned*, with the title of each graph or section specifying the variant. Three evaluation metrics are used throughout: cost, peak load, and total load. *Load* is the amount of kWh drawn by the water heater to heat water. Load can be either *peak* energy, when the power is drawn during a pricing schedule defined peak period, or the *total* energy drawn over the course of the day. *Cost* is the cost of this energy over the course of the entire day. We present results

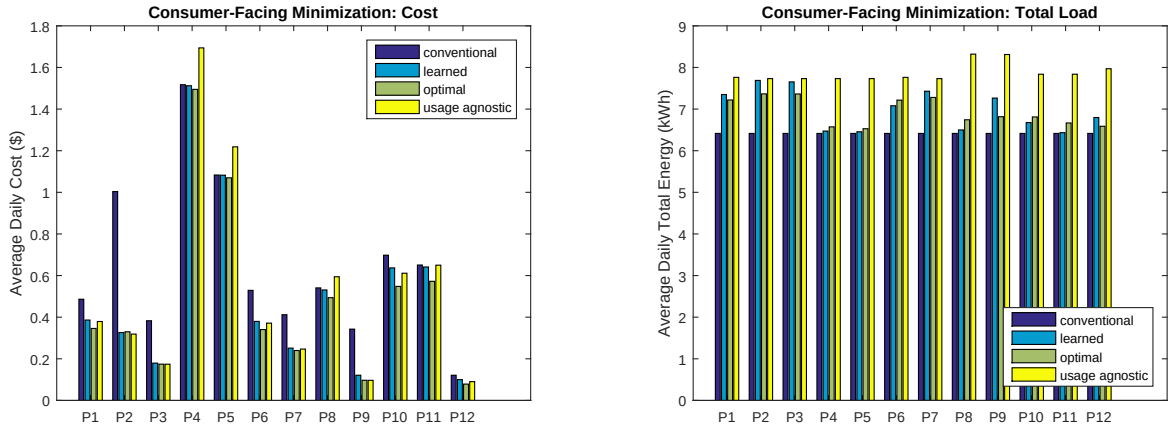


Figure 3.6: Consumer-facing ThermalThrift saves both money and peak energy for every pricing schedule (P1-12). Schedules that have low cost savings tend also to have lower peak energy savings (e.g. P4-5).

as an average of all participant groups for each of the TOU pricing schedules. For all individual households, ThermalThrift never increased their average daily costs over a conventional setpoint heater.

3.5.1 Consumer-Facing Cost Minimization

The consumer-facing approach of ThermalThrift shows that water heater TES can be used to save both cost and peak load for all pricing schedules when performing load shifting, as shown in Figure 3.6 and 3.7. Overall, the consumer-facing approach taken by ThermalThrift uses TES to reduce 47% of the peak electric load and 25% of the cost to consumers over a conventional water heater. For the usage agnostic baseline, cost reductions are often comparable to the learned and optimal ThermalThrift, except for three pricing schedules (P4,5,8) where it costs consumers 11-12% more than a conventional tank to attempt storage without using a learning method to manage standby loss with respect to usage. Additionally, the usage agnostic approach uses more total energy than ThermalThrift's 8% increase, as unused storage is wasted through standby loss as seen in Figure 3.6(b). While the usage agnostic non-AmI approach does save more peak energy, it is unlikely to be adopted by consumers due to these costs and therefore fails to leverage the potential TES of water heaters.

While ThermalThrift reduces costs for all pricing schedules and shows that dynamically managing standby loss is costs effective for load shifting, the total savings vary with TOU pricing schedules. The lowest saving pricing schedules can be categorized into two groups: low relative increase in peak pricing and high relative increase in peak pricing during variable use hours. Pricing models P4, P5, and P8 fit into the first category; they have high peak usage, but low differences between peak and off-peak prices ($0.037, $0.02, and $0.0309).

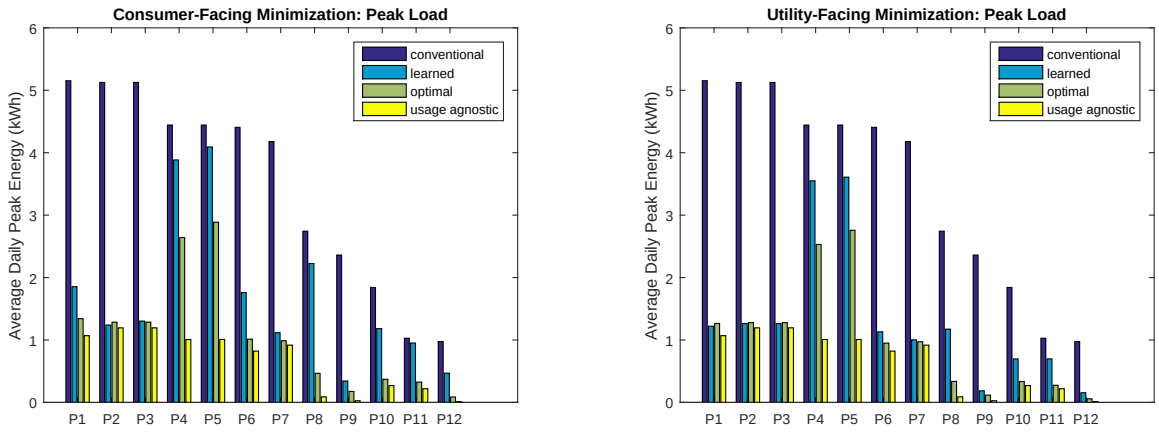


Figure 3.7: Utility-facing ThermalThrift saves more peak energy than the consumer-facing variant.

For these pricing schedules the standby loss cost of incorrectly storing too much TES is too large compared to the possible savings, so ThermalThrift performs little storage. Even in the optimal case, where ThermalThrift knows how much hot water will be used during peak hours, the optimal cost improves little over the learned cost due to the low price differential.

In the second category, pricing schedules P10 and P11 have high price differences between peak and off-peak hours ($0.1219 and $0.14), meaning TES standby loss costs could easily be covered by peak savings. However, most usage within the participant groups occurs outside of these hours, causing the learning component of ThermalThrift to predict low amounts of usage during peak hours and store little thermal energy. However, the occasional use of high flow appliances (shower, dishwasher, washing machine) during these hours is costly and consumes a large amount of peak load. Since ThermalThrift had not learned these events from historical data, it did not store energy in preparation for them. This, and the higher savings exhibited by the optimal ThermalThrift for both cost and peak load, highlight the importance of prediction for water heater TES. The current approach uses a simple algorithm to the average cost across all historical days for prediction. The approach could be more complex, weighting predicted costs based on day of the week or similarity to the current day's usage, to achieve additional savings in P10-11. However, overall, the simple averaging approach already reduces 47% of the water heating peak load. Additionally, the algorithm's simplicity allows ThermalThrift to learn usage and operate effectively with only a short learning period (i.e. a few days).

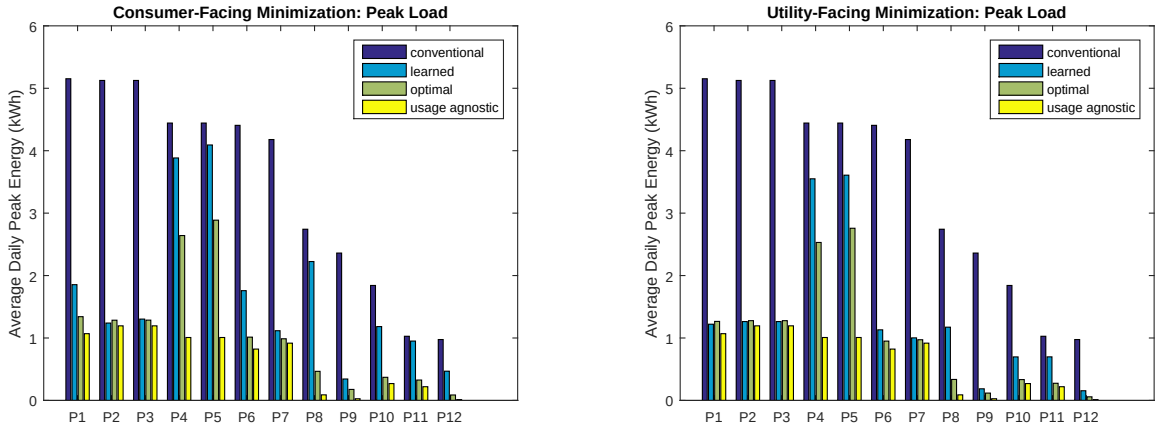


Figure 3.8: The chosen TES temperatures for these savings vary widely across pricing schedules, participant groups, and individual days. The most common TES temperature is the highest possible, 93 °C.

3.5.2 Utility-Facing Peak Optimization

The utility-facing results in Figure 3.7(b) show an even greater reduction in peak energy over the consumer-facing variant. On average, ThermalThrift reduces 62% of peak load while maintaining costs below that of a conventional, non-TES water heater. Since it maintains the conventional cost, ThermalThrift cannot reach the peak shifting achieved by the aggressively peak shifting usage agnostic approach as seen in Figure 3.7(b). This is most visible for schedules P4, 5, and 8 where cost constraints limit the learned ThermalThrift's peak savings. However, by maintaining conventional costs the utility-facing ThermalThrift is more likely to be accepted by consumers, facilitating peak load shifting for utilities using water heater TES. Additionally, due to the design of TOU pricing to monetarily incentivize load shifting, the utility-facing approach does still save consumers money as seen in Figure 3.9.

The chosen TES temperatures for ThermalThrift's savings varied significantly across pricing schedules and participant groups, as shown in Figure 3.8(a). Some schedules (P2,3,7) had the majority of their daily TES temperatures set to 93°C, indicating that TES is highly cost effective for these participant group/pricing schedule combinations. Three low peak load shifting schedules (P4, 5, 8) had TES temperatures clustered around lower temperatures and show the same saving issues as the consumer-facing approach: relative pricing between peak and off-peak is too low to risk TES. Many of the schedules (P1, 6, 9, 10, 11, 12) had TES temperatures that greatly varied between individual participant groups and across days due to the differences in usage patterns during peak hours, indicating that learning these usage patterns for individual households is necessary to manage standby loss for cost effective TES and load shifting. Overall, ThermalThrift shifted water heating peak load for every pricing scheme we evaluated and reduced 15% more peak load on average

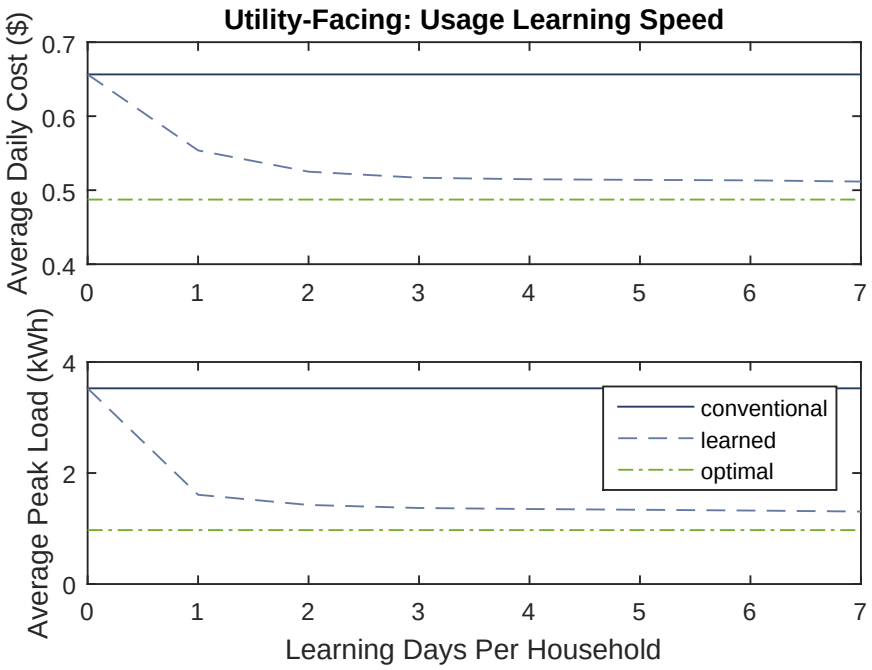


Figure 3.9: ThermalThrift is able to quickly learn usage patterns and successfully use predictions to manage standby loss. After only 7 days of training data for a given household, ThermalThrift is able to achieve within 5% and 36% of the optimal cost and peak load reduction.

than the consumer-facing approach with only a 13% increase in total energy consumption.

3.5.3 Learning Period Analysis

The success of a learning water heater depends in part on how quickly it can learn. Therefore, we also evaluate ThermalThrift with limited learning days in Figure 3.9. The figure shows the average cost and peak load across all participant groups and TOU schedules for all combinations of learning days in 8 days of data using cross-fold validation. This indicates that after only 7 days of training data for a household, ThermalThrift achieves within 5% and 36% of the optimal cost and peak load reduction for the utility-facing approach, with similar results for the consumer-facing approach. Additionally, ThermalThrift halves the difference between the conventional and optimal baselines with only one day of learning data. This indicates that learning usage patterns can quickly make TES cost effective for real consumers.

3.5.4 TES Potential and the Mixing Valve

The addition of a mixing valve on tanks allows the tank to be heated much higher than human tolerable temperatures and therefore provides a larger opportunity for TES. However, ThermalThrift can still provide TES with only human tolerable temperatures (60°C and below), provided a low comfort temperature such as

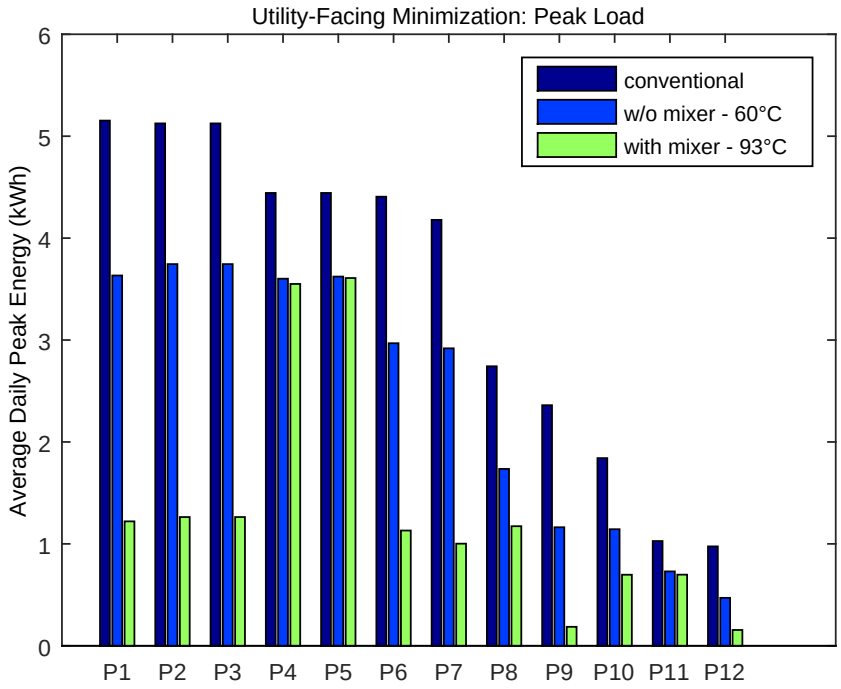


Figure 3.10: With a human tolerable TES limit of 60 °C requiring no mixing valve, ThermalThrift shifts 24% of the peak load. For some pricing schemes no more shifting is possible (e.g. P4-5), but on average a 93 °C TES limit shifts far more of the peak load (62%).

49°C is used. Some pricing schedules P4, P5, P8, and P11 have many chosen TES temperatures at or below 60°C even when higher temperatures are available, as shown in Figure 3.8(b). Thus, for these pricing schemes, the highest potential load shifting is often available cost effectively without exceeding human tolerance limits. Figure 3.10 shows peak energy savings in a TES limited 60°C and 93°C ThermalThrift. It shows that ThermalThrift could save peak load over a conventional tank even without mixing capabilities. Additionally, a lower temperature limit consumes less energy for TES in the pre-peak period than the 93°C limit, reducing total energy consumption. However, pricing schemes that do benefit from the higher TES temperature limit can shift more than twice as much peak load due to the larger storage capacity. ThermalThrift with a 93°C limit can shift 62% of peak load on average across all pricing schemes, while a 60°C limit shifts only 24%.

3.5.5 TES Power Peak

In managing standby loss by implementing storage just before peak hours, there is the potential for a slightly earlier water heating aggregate peak in a neighborhood. However, aggregate energy peaks across the grid consist of more than just water heating energy and include other peak energy use such as HVAC and lighting. A new peak in water heating use may simply help fill the valley still present due to other peak energy use

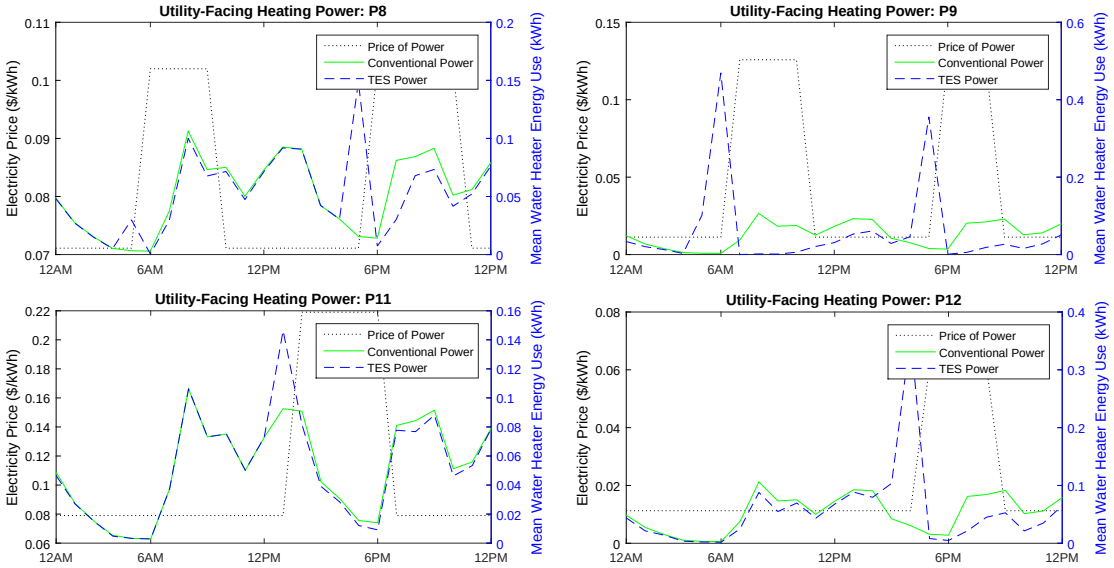


Figure 3.11: Heating TES just before peak hours to minimize standby loss can create a "new peak" in water heating just before peak hours. However, the magnitude and time of this peak depends greatly on the pricing scheme. In widespread use, such prices could be adjusted to suppress this peak or move it to times of excess power generation (e.g. renewables).

and can be used to perform peak leveling. Controlling this leveling may depend on the pricing scheme used. Figure 3.11 shows four pricing schemes and the shift in heating energy caused by the shifting focused utility-facing variant. In each case, shifted energy use creates a water heating peak just before peak hours. The amplitude and time of this spike depends greatly on the pricing scheme itself. While P8 and P9 have peaks that fill in the valleys of the aggregate water heating use of the study groups, scheme P11 heightens a peak in the middle of the day. Adjustments to the price and time of peak hours, in conjunction with information about the historical data and operation of the learning algorithm, could be used on a neighborhood scale to shift water heating to ideal times. Additionally, neighborhood schemes, such as one similar to that presented in [86] to account for the "payback peak" when off water heaters are turned back on after peak hours, could constrain TES charging in a neighborhood to reduce or level the new peak. Hence, while such a peak is a concern when the learning algorithm is widely used, it can be mitigated or shifted on a neighborhood scale.

3.6 Limitations and Future Work

ThermalThrift's results indicate that learning hot water usage patterns in each home using a human-centric AmI approach can produce cost effective TES above a personalization agnostic approach. This approach can reduce costs for the consumer and reduce peak energy for the utility. Future work must explore whether these results extend across a wider variety of households and buildings, including homes with larger families

and a wider variety of household water appliances, or other building types such as laundromats, restaurants, and hotels. Intuitively, buildings with more hot water usage at the beginning of peak hours will experience more savings with ThermalThrift – due to standby losses usage patterns during the ends of peak periods will fare less favorably.

With TES, ThermalThrift maintains a higher temperature during peak hours for a longer period of time than conventional water heaters. Most water activities will be unaffected by more available hot water, such as sink, dishwasher, or washing machine uses, but usage such as showers could potentially be prolonged due to the increased hot water – thereby increasing hot water use during peak hours. Future work must analyze this potential effect, and how ThermalThrift might adapt to these changes in water use.

Future work may also look at the effect of changing temperatures in a tank over time. For water heaters, higher temperatures can increase sediment buildup, covering and/or wearing out the heating elements, shortening the lifespan. However, the effect of limited and dynamic higher temperatures is not well studied (ThermalThrift only uses high temperatures just before and during peak hours) and ThermalThrift shows savings even with water heater recommended temperatures as the upper TES bound (e.g. 60°C). The effect of these temperature changes on lifetime is an avenue of future work. In terms of temperature, future work may also look at adding a Legionella killing temperature constraint to TES operation if 49°C is maintained as the user comfort minimum. Most Canadian homes have hot water thermostats at 60 °C to kill Legionella at the costs of larger standby losses. However, 60 °C need only be maintained for 30 minutes to kill Legionella and such temperatures could be deliberately achieved periodically with TES.

In addition to thermal load shifting, ThermalThrift could also be used for cost effective storage in other demand-response services including storage of renewable energy and performing frequency regulation. Several works, evaluate water heating for ancillary services to aggregately store highly variable renewable energy, but note that heating demand does not necessarily match excess power [87, 88] and Pourmousavi et. al. use load shifting to match with wind generation but do not account for consumer costs [89]. Combining these approaches with ThermalThrift, which accounts for consumer costs and requires excess energy for TES charging, may provide a cost effective way to direct renewable energy to a consumer who will use it to decrease their own costs – something that will vary significantly based on individual usage and TOU pricing. Along with changing TOU pricing, this may also be used to control the potential for a slightly earlier aggregate peak in a neighborhood from ThermalThrift. Evaluating what response ThermalThrift has on changing peak periods, prices, and renewable energy is a direction for future work.

3.7 Conclusions

ThermalThrift shows that incorporating personalization and AmI techniques can improve result in systems geared towards meeting societal needs. Overall, the consumer-facing approach taken by ThermalThrift uses TES to reduce 47% of the peak electric load and 25% of the cost to consumers over a conventional water heater. It does so without ever costing an individual consumer more than a conventional heater – while the usage agnostic approach costs three of the consumers 11-12% more. ThermalThrift achieves this by sensing the context of the user (i.e. their water usage), learning and predicting that usage over time, and controlling their environment (i.e. the water tank) to meet a societal need while still maintaining their comfort. ThermalThrift meets this societal need in a cost-effective way by making a traditionally usage agnostic approach user-centric, thereby showing that AmI has the potential to improve the adoption of systems to meet societal needs. However, ThermalThrift could be further improved by incorporating more information about users. For example, knowing the identity of the person in the home could improve the accuracy of TES by allowing the system to rule out usage predictions for those who are not home. If the person who usually takes a morning shower is out the house, then heat should not be stored for their use. Additionally, ThermalThrift does not currently take into account any changes in user preference over time. A change in jobs, the start of school, a sudden interest in cooking, etc. could all change the usage patterns of water in the household. Incorporating this further human-centric information and adapting to these human changes would mean that ThermalThift would have to deal with the privacy concerns of sensing and the human-in-the-loop problems with learning in AmI.

Chapter 4

Preserving Image Privacy for AmI Sensing

4.1 Introduction

As we have seen in this work so far, meeting societal needs or other human objectives with AmI requires that we sense the people within an environment to understand their preferences and behaviors. One of the best ways to do so is by monitoring the users directly with some form of sensing infrastructure. However, the more information rich and accurate a sensor is and the more directly it monitors people, the more of a privacy concern it presents as it can be capable of collecting information that is unnecessary to the system or unconsented by the user. Due to the convergence of several technology trends, imaging sensors are quickly becoming the most widely used sensors that exemplify this trade-off. Only two decades ago, visible light, thermal, and depth imagers we expensive devices used only for industrial or professional use. In the 1990s lithography developed to the point that CMOS imagers became practical, promising lower power consumption, higher speed, and reduced cost. Soon thereafter, the explosion of demand for mobile devices created a very high volume market, allowing technology investment that further reduced cost and improved image quality, even as pixel sizes shrank and imaging arrays became larger. Today, imaging sensors can be found for pennies per chip and power consumption can be as low 84 pW per frame per pixel [90], and Moore's Law will only enhance these trends. Recently, thermal imagers and depth imagers have been able to leverage these trends due to CMOS-based designs, such as the microbolometer arrays made by FLIR [91] and time-of-flight sensing arrays in the Microsoft Kinect [92]. As a result, thermal and depth imagers are beginning to follow the same cost, performance, and power curves that visible light imagers have seen over the past two decades. As a

result, digital image processing is transforming industries and the human environment, from city-scale face recognition systems to vending machines that personalize choices [93] to vehicles that detect alertness of the driver [94].

Despite the high information fidelity, low cost, low power, and small size of these sensors, image processing is not utilized nearly as much in the home environment. Homes are often filled with cameras for security systems, video games, video conferencing, wearables, and personal electronic devices. However, AmI and automation processes that leverage digital image processing have not had the same transformative effect on the mainstream home environment that they have had on the public environment. For example, since we spend 68% of our time in homes and utilize 20% of our energy footprint there, having AmI systems in the home is necessary to meet energy based societal needs [95, 96]. Several studies indicate that privacy concerns are at least one main cause for this discrepancy [97, 98, 99, 100, 101].

Privacy, as it relates to imaging sensors, can be roughly divided into two categories: *image privacy* is the protection of our image and the potentially unconsented information it holds, while *information privacy* is the protection of knowledge derived from those images for the smart system with given consent. Information privacy is a concern regardless of the sensing equipment used and is currently an active area of research [102, 103, 104]. In our work, we instead focus on protecting the privacy of an image and the extraneous information it collects. We performed a preliminary study in which we surveyed 200 people to assess whether image privacy could be addressed by processing images onboard the camera and never reporting them to the cloud. In essence, this represents the security based solution to image privacy. The results indicate that 21% of participants would still reject such a system due to hacking concerns: even if the system was designed to not report images they could still be collected if the camera system was hacked. This number would likely increase were home camera hacking to become more prevalent than it already is [105].

In this work, we demonstrate a novel hardware-based approach for *privacy-preserving image processing*: the ability to automatically extract information from imaging sensors without the risk of compromising image privacy, even if the system is hacked. The basic idea is to limit both the memory available onboard the camera and the data rate of camera communication: if memory is limited to M bytes and communication is limited to R bytes per time period Δt, a compromised camera could collect at most $M + R$ bytes per time period Δt. Hence, a hacker could extract $M + R$ bytes from one or multiple images during Δt. Within the context of this architecture, we define an algorithm to be *image privacy-preserving* if it can extract useful information using M bytes of memory such that $M + R$ bytes are too small to hold a compromising image. Ideally, $M + R$ is much smaller than the image size. As a proof of concept, we present a system called *Lethe* that uses a binocular thermal camera to perform human identification and tracking, two of the main sensing components of AmI. We do so using only 33 bytes of memory (or 0.69% of the image size) and a data rate of

Figure 4.1: Similar to RGB cameras, thermal cameras collect details of activity, identity, and clothing at high resolutions.

30 bytes per second.

Thermal cameras have been used to track and identify people by facial recognition in combination with visible light cameras and as standalone devices [106, 107, 108, 109]. However, as they are used today, thermal cameras still pose a privacy concern. While low-resolution thermal cameras (e.g., 8x8 pixels) are argued to be privacy preserving [110, 111], the high-resolution cameras needed for identity detection are not (e.g., Figure 4.1). Image obscuring techniques, such as blur, silhouette, pixelization, and masking, can reduce privacy concerns but do not protect against the removal of the full image from the camera itself due to leakage or hacking [112, 113]. To bypass the need for a full image technique like current facial recognition, Lethe only estimates the height and direction of each person. Prior work has shown that this information is often sufficient for room-level tracking in homes, which typically have only a handful of residents [114]. To be an *image privacy-preserving* system, Lethe must perform all functions with limited memory. Therefore, motion based human detection is infeasible since it requires memory-intensive background subtraction algorithms. Additionally, common edge detection algorithms, such as search-based and zero-crossing based edge-detection, require large system memories to compute the required image gradients. In contrast, Lethe evaluates one pixel at a time as they stream in over the communication bus from the camera chip – throwing out each pixel before the next one is received – and extracts only the minimum amount of information required to perform tracking (33 bytes). To do so, Lethe leverages one key feature of thermal cameras: humans are often warmer than background temperatures. As Lethe monitors a space, state information is aggregated over multiple frames to determine the presence, direction, and the relative height of people as they cross the field of view. The state from two thermal cameras stacked vertically is fused through a binocular vision algorithm to estimate a person's height regardless of their distance from the cameras. Hence, Lethe aims to accurately detect people as they pass through the doorway with a memory of M bytes such that $M + R$ bytes per Δt is much smaller than the image size and is not a risk to image privacy.

Our results show that even with this very limited memory footprint, Lethe can identify when a person crosses the field of view with 96.9% accuracy and can determine walking direction with 99.7% accuracy. In a best case of our collected data, it can track 3 individuals with only a 6cm height differential with 94.3% accuracy. On average, it can differentiate 2 individuals 96.0% of the time with a 5cm (˜2in) or greater

difference in walking height and 92.9% with a 2.5cm (˜1in) or greater difference. These results were produced using a thermal camera with only 8 frames per second and a 60×80 pixel image. Our analysis shows that these results will increase dramatically, approaching 100%, as manufacturers increase resolution and frame rate. Overall, an image privacy-preserving tracking system would be an important addition to the suite of AmI sensors. Additionally, we anticipate this hardware-based approach to privacy-preserving image processing can generalize to other visual tasks, helping to advance the acceptability in home AmI systems of one of the most important sensing technologies on the market today.

4.2 Related Work

Much work has done in the recent decades to obscure privacy invasive information in images and video. Approaches include: blur, silhouette, pixelization, masking, and oval and box overlays on recorded video [1, 112, 113]. Peng et. al. extract and encrypt facial regions on CCTV recorded video to protect identities [115]. Edgcomb et. al. focus on in-home fall detection and find that oval overlay obscuring was sufficient for 88% of their participants [112], but blur and silhouette were not [1]. While many of these approaches have been shown to decrease privacy concerns for video images, all focus on obscuring the video after it has been recorded through software applications and many are premised on the fact that the video will be seen by others. In contrast, our work focuses on preventing the recording, storage, and viewing of recorded video through hardware design to provide hard guarantees on risks to image privacy and completely preventing the collection of unconsented data.

Several tracking and identification systems use thermal cameras because they are not confused by lighting levels and can operate in complete darkness. They have long been used to augment RGB cameras for identifying humans [107, 108, 109] and have been used in indoor and outdoor locations for pedestrian tracking, people counting, and fall detection [116, 117, 118]. Thermal cameras have also been shown to record enough detail to successfully identify people through facial recognition [118, 106] and identifying gait features [119, 120]. Additionally, thermal cameras have recently been used to perform gesture recognition for touch surface interaction [121, 122] and to identify air written numbers and letters in combination with color cameras [123]. However, these algorithms do not provide any assurances about image privacy. These algorithms all operate on the full thermal image and therefore have enough memory to leak the images if hacked.

Visible spectrum imaging falls along the same lines: because of privacy concerns, most visual systems are used only in public or office environments despite their potential for identification and tracking in the home environment [124, 125, 126, 127, 128]. Identity detection with visual cameras commonly uses facial,

iris, stride, or gait analysis [129, 130, 131, 132]. However, such systems commonly suffer from occlusion where a person looks away from the camera or is not positioned correctly for gait analysis, often requiring other sensors to compensate [133, 134, 133]. Additionally, image processing in the visible spectrum requires background subtraction, segmentation, and classification algorithms, which require a full frame if not a full video [135, 125, 136, 137], which creates risk for image privacy.

Some thermal camera systems do attempt to preserve image privacy by relying on very low-resolution sensors. For example, some systems monitor the occupancy of a room or track room location using an 8x8 thermal camera and PIR sensor where the entire body of a human is represented by only a few pixels [111, 110]. However, these systems cannot detect the identity of people due to the low resolution. In contrast, Lethe is able to process very high resolution images in order to extract identity while still providing image privacy assurances. To our knowledge, Lethe is the first thermal camera system to do both.

In addition to Lethe, several other AmI systems have attempted to track people in homes with passively collected biometrics such as height and weight. However, because cameras have not traditionally been considered an acceptable sensor in the home due to privacy concerns, such systems are restricted to using sensors that have lower accuracy or higher cost. For example, smart floor systems identify individuals by weight [138, 139], but require extensive and expensive deployment. WiFi based systems that detect identity require extensive personalized training and a very controlled space, such as narrow monitored hallway [140, 141]. Other systems, such as Doorjamb, use ultrasonic based height and width detection for identity [114, 142]. However, these systems often suffer from large numbers of false positives as active sensing signals sent out into the environment can mistakenly detect doors and nearby people. Lethe builds on the same principles as these systems but, by making hardware-based guarantees about image privacy, is able to do so with a sensor that will continue to become cheaper, lower power, and more accurate over time.

4.3 Lethe: A Privacy Preserving Camera for Multi-User Environments

To provide privacy-preserving image processing while tracking and identifying people in AmI systems, Lethe must be designed with consideration for three main components: *crossing detection*, *identity detection*, and *limited video and algorithm memory*. Crossing detection determines when some person has crossed the sensor and in which direction they are traveling. Identity detection determines the identity of the person crossing the field of view. Both components must operate with limited video and algorithm memory to become image privacy preserving. For Lethe, each component operates on a single pixel of the image at a time and

accumulates algorithmic information over the course of multiple pixels and frames. The overall architecture of the system can be seen in Figure 4.2. On the camera, each process uses only a few pixels to store accumulated information and processing (less than 40 bytes). The camera processing then sends a tuple of start time t_s, end time t_e, direction d, and detected pixel h_{pixel} of each event to the cloud or in-home server for final processing. Here, Lethe uses the stereo configuration of the two thermal cameras, where the cameras are stacked on top of each other on one side of the doorframe both facing the opposite side, to biangulate a person's location and height and map that height to their identity h. Finally, the identifying tuple (t_s,t_e,d,h) can be made available for use in higher level AmI applications. The details of how Lethe implements the detection algorithms with limited algorithm memory are described below.

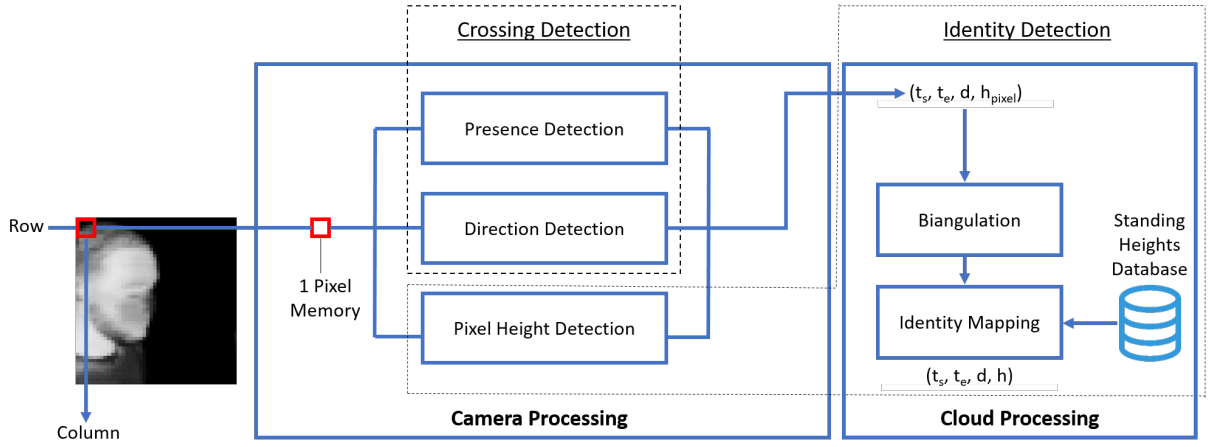


Figure 4.2: When processing a thermal image each pixel is taken in one-by-one and processed by the Presence Detection, Height Detection and Pixel Height Detection algorithms. Note that while the thermal images are presented as full pictures in each of the figures of this paper, only a single pixel at a time is read by the thermal camera for processing. The full image is never stored in memory. Once all the frames in a crossing have been processed, a tuple with the value t_s, t_e, d and h_{pixel} is sent to the cloud or in-home server. There the output from two cameras in stereo are processed to biangulate a height and map that estimated height to an individual.

4.3.1 Crossing Detection

Crossing detection determines when a person crosses the field of view of the sensor. Its main goal is to determine the start time, t_s, end time, t_e, and direction, d, of a person as they cross. Its secondary, but equally important, goal is to determine when a person has *not* crossed the threshold. This secondary goal is vital in real-world deployments where people interact with doorways in ways beyond just crossing the threshold. They might brush up against the doorway as they walk by, reach through to turn on a light but not cross, pause in the doorway to chat or grab their keys, or turn around halfway through the crossing and return the way they came. Hence, crossing detection must detect real crossings with times and a direction,

Table 4.1: The notation used for a thermal camera, biangulation, and identity mapping. Thermal Camera: Values in Memory shows the number of bytes required for each item in memory including four values that we not given notation in this writeup; 33 bytes in total are required.

Thermal Camera: Values in Memory	
t_s	starting timestamp of an event (4 bytes)
t_e	ending timestamp of an event (4 bytes)
t_{last}	timestamp of last frame with a presence (4 bytes)
r_back	background reference temperature (1 byte)
loc_l^i	horizontal location of leftmost detected presence in frame i (x4, 4 bytes)
loc_r^i	horizontal location of rightmost detected presence in frame i (x4, 4 byes)
d_{ind}	the directional indicator derived from the leftmost and rightmost values (x2, 2 bytes)
d_{sum}	the current summation of the direction indicators (x2, 2 bytes)
h_{pixel}	the vertical height of the detected pixel height (1 byte)
	the current time stamp (4 bytes)
	the current pixel temperature (1 bytes)
	the column and row values of the current pixel (2 bytes)

Thermal Camera: Notation	
r_{thresh}	the statically stored temperature threshold between background and human presence
d	the final direction value of an event
Biangulation: Notation	
h_u	h_{pixel} from upper camera
h_l	h_{pixel} from lower camera
p_w	angel width of a pixel (degrees)
d_u	height of the upper camera (cm)
d_c	distance between the two cameras (cm)
d_t	calculated height of a person taller than both cameras (cm)
d_m	calculated height of a person between both cameras (cm)
d_s	calculated height of a person shorter than both cameras (cm)
Identity Mapping: Notation	
h	calculated height from biangulation (cm)
h_m	mean height after clustering (cm)
h_r	measured standing height of a participant (cm)

(t_s,t_e,d), and differentiate these from other doorway interactions. Lethe does this in two steps, detecting the presence of a person in the field of view with *presence detection* and determining the direction they travel in with *direction detection*. To identify and remove non-crossing sensor interactions, direction detection doesn't just determine if they moved from room 1 to room 2, but also if they entered the doorway from room 1 and exited the doorway back into room 1. Hence, the combination of presence and direction detection become Lethe's crossing detection.

Presence Detection

We detect the presence of a person in the field of view based on their temperature. Indoor environments are typically air-conditioned to a 20°C to 22°C range and most objects in the environment conform to this temperature. Human skin temperature ranges between 32°C and 34°C [143]. Additionally, hair and clothing tend to absorb the skin's heat and retain a temperature lower than skin, but higher than the surrounding environment (e.g., 25°C to 30°C). Hence, a person can be detected in the view of a thermal camera by identifying pixels that are warmer than the background temperature. In Lethe, we determine the background temperature r_{back} as the average temperature of the last frame absent human presence. This value is updated

with each new human-less frame to adjust the background temperature to changes in the environment temperature.

We detect a human in a pixel whenever that pixel has a value r_{thresh} degrees higher than the background reference r_{back}. If at any point in the streaming evaluation of a frame a pixel passes this threshold, the frame is declared as having a presence. If this is the beginning of an event, t_s is updated. Once a person is detected, we consider all frames 0.366 seconds after the last frame with a presence (t_{last}) to be part of the event. This aims to cover any gap caused by lost frames in an event, without merging two events in succession. This lag time value is conservatively based on the average walking speed of a person (1.39 m/s) and the average shoulder span of an adult male (0.508 m). Hence, if a person is slightly over a shoulder width away from leaving the frame when data is lost and we catch a last measurement of their shoulder as they leave when video data is regained, we combine this last measurement with the previous event measurements. This value is a heuristic and can be modified for a particular setup. If there is no chance of data loss, the period ($1/fps$) of the video recording can be used for the lag time. A long lag time may merge two events together if people walk in quick succession. In our setup, 0.366 seconds must pass before a person can follow someone into the video range in order for a second event to be formed. If the last human presence frame occurred earlier than this lag, t_e is set and the event is ended.

The exact values of r_{back} and r_{thresh} will depend on the thermal camera used. The FLIR hardware implementation we used for this work (described in detail in Section 4.4.1) does not provide a mapping from the measured value to a temperature but instead returns a value in the range of 0-9000. To turn these values into temperatures, we estimated the measurement to temperature mapping based on readings from an infrared thermometer and found a change of 0.027°C per value. We then truncated the lower readings to 20°C to capture only the typical background and human temperatures. This meant that cold objects in the environment could not significantly shift the background average r_{back}, while still allowing the background temperature to shift in the normal indoor range. Thermal images throughout this paper are shown with the truncated range of 7500-8700 (19°C to 41 °C) to map the values into the 0-255 range needed for images. We then define r_{thresh} to be 3°C. This threshold was chosen based on our general knowledge of the difference between the air temperature in a typical indoor room (20-22°C) and a small sampling of human hair and clothing temperatures from people in our office (25-30°C). When the background is cold, this allows us to detect a person's clothing and longer hair as their presence. If the background is warmer, then this contracts their detected presence to thin clothing, shorter hair, and skin.

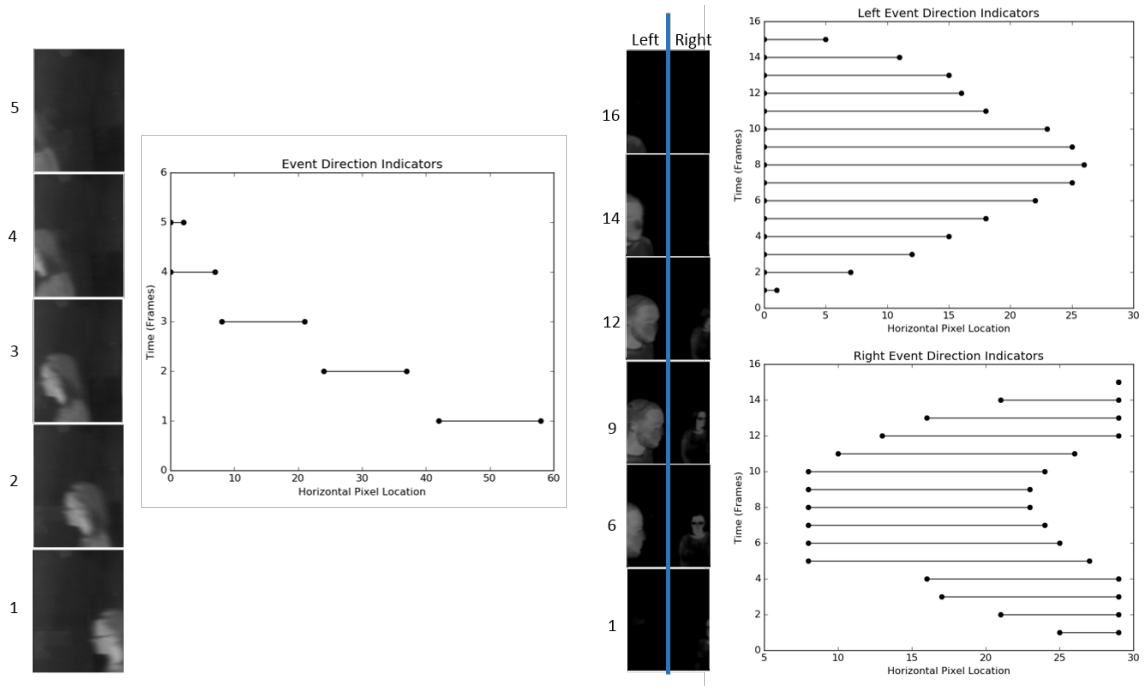


Figure 4.3: A canonical example of a person crossing through the field of view (left) and an examples of a complex non-crossing event (right). Direction detection captures the leftmost and rightmost position of the person in the left and right half of each frame to determine direction over time or filter out non-crossing events based on these indicators (above, left and right). Additionally, note that while the full thermal images are presented here for clarity, only a single pixel is read into memory at a time by the thermal camera for processing and only two pixel locations are retained from the previous frame for direction detection.

Direction Detection

Once presence detection has ended an event with the creation of the (t_s,t_e) tuple, direction detection either assigns a direction d to that event or removes that event as a non-crossing interaction. Direction detection collects the information needed to make this determination in parallel to presence detection. Each streaming pixel evaluated by presence detection is also used by direction. Direction detection uses the same r_{back} and r_{thresh} as presence to determine if someone is present in a pixel. However, direction detection goes beyond presence to use this information to determine a person's location in the frame and how that location changes over time. When a person crosses through the field of view they first appear on one side of the frame and then progress in intervals to the other side before leaving the field of view. A canonical example of this behavior can be seen in Figure 4.3. Lethe captures the progress of the person over time through the field of view by leftmost, loc_l^i, and rightmost, loc_r^i, location of the person in frame i. If the pixels of a frame are processed from left to right, this means Lethe sets loc_l^i to the column of the first pixel with a human presence and loc_r^i to the column of the last pixel with a human presence. In this way, Lethe identifies the region a person is located in for each frame.

Once loc_l^i and loc_r^i for a human region have been identified the challenge becomes turning this information into a direction using limited memory. While we could retain each region for every frame and process the information once the event is over – this would mean storing values double the number of frames in an event in memory. Since there is no limit on the number of frames in an event (a person could stand in the doorway for multiple minutes before walking through), this would not allow Lethe to maintain memory limitations. Instead, Lethe accumulates information about direction by comparing only two regions at any given time. Each frame's region is compared to the stored region of the previous frame and assigned a direction indicator value: 1 if the region has shifted to the right, -1 if it has shifted to the left. Lethe retains only the sum of these direction indicators, d_{sum}, over multiple frames and assigns a direction, either left or right, based on the negative or positive value of d_{sum} respectively at the end of the event.

To detect and remove non-crossing events, Lethe identifies $d_{sum} = 0$ as an event where a person entered and left from the same side of the field of view. However, the basic direction algorithm described above will fail in two scenarios: variable walking speeds and non-crossing events with a person on either side of the door. In the case of variable speed events, a person might walk slowly into the door and then quickly out the same side. Here, d_{sum} might equal $1+1+1+1+1-1-1-1=2$ as the first five frame pairs indicated right and only the last three indicated left. Hence, the event would be mislabeled as a rightward crossing through the frame. To prevent this, only direction indicators that are different from the last indicator, d_{ind}, are added to d_{sum}. In our example, $1+1+1+1+1-1-1-1=2$ would become $1-1=0$ and the event would be correctly identified as a non-crossing. This makes the algorithm agnostic to walking speed. Non-crossing events with a person on either side of the doorway are more complex. An example of such an event, where two people walk into the field of view to chat across the doorway, can be seen in Figure 4.3. To identify these non-crossing events, we split the direction detection algorithm into two and determine a direction for each side of the doorway independently. Each side detects human regions and accumulates a direction sum (i.e., d_{sum1} or d_{sum2}). If d_{sum1} and d_{sum2} both indicate left or both indicate right, the event is given a direction. If both direction sums are 0 or different directions, then Lethe dismisses the event as a non-crossing. Hence, Lethe now retains 12 pieces of information for any pair of frames: loc_{l1}^i, loc_{r1}^i, loc_{l1}^{i-1}, loc_{r1}^{i-1}, d_{ind1}, d_{sum1}, loc_{l2}^i, loc_{r2}^i, loc_{l2}^{i-1}, loc_{r2}^{i-1}, d_{ind2}, d_{sum2}. With this information, Lethe determines the direction of crossing events and detects when a person has passed through the field of view but has not crossed the threshold.

4.3.2 Identity Detection

Lethe detects the identity of an individual by using their height as a proxy. While heights are not unique to individuals, most homes have only a few occupants each of whom can be distinguished by their height. For

Lethe, height was chosen specifically because a full image of the person is not needed to extract height – only the location of the top of their head is required. Hence, a full frame is not needed to identify an individual and the identity proxy of height can be extracted in a memory limited fashion. However, video represents only a 2D image of a 3D space – we have no idea how close the person is to the camera. Hence, the pixel height found in the video may not map to the physical height of the person. For example, a person taller than the camera will appear taller closer to the camera and shorter further away. For a person shorter than the camera the distortion is reversed. To deal with this, Lethe uses two cameras to detect two-pixel heights of a person. Then, it biangulates the person's distance from the camera and physical height given these two measurements. Finally, Lethe maps these estimated heights to individuals. Hence, Lethe performs identity detection in three steps: *pixel height detection*, *biangulation*, and *identity mapping*.

While pixel height detection is performed on the thermal camera, biangulation and identity mapping take place off the camera in either an attached processor that stores all the crossing events over time for a doorway or a centralized location that stores the crossing events for all doorways. Hence, the thermal camera creates and transmits the tuple (t_s,t_e,d,h_{pixel}) where (t_s,t_e,d) is determined by crossing detection and h_{pixel} is determined by pixel height detection. Biangulation and identity mapping turn h_{pixel} into the person's walking height and identity h on a different device with more memory and processing power. This final tuple (t_s,t_e,d,h) can then be used on that applications device for higher level processing AmI systems or other smart applications.

Pixel Height Detection

To determine the height of an individual, each camera must first determine the person's relative height according to that specific camera. Hence, this step determines the pixel height of the person in a single camera's field of view. The pixel height is defined as the highest pixel in the field of view that recorded the person over the course of the event. Like presence and direction detection, height detection determines if a person is present in a pixel using the threshold r_{thresh}. Height detection records the highest row in any frame or any column where it detects a person over the course of an event. No information beyond the current h_{pixel} is retained for height detection. Once an event is complete (as defined by presence detection) each camera reports h_{pixel} as the pixel height of that event.

Biangulation

To transform the relative pixel heights h_{pixel} reported by the thermal cameras into a height of an individual Lethe performs localization by biangulation. To perform biangulation, two thermal cameras are stacked on top of each other on one side of the doorframe with both facing the opposite side. With these two cameras,

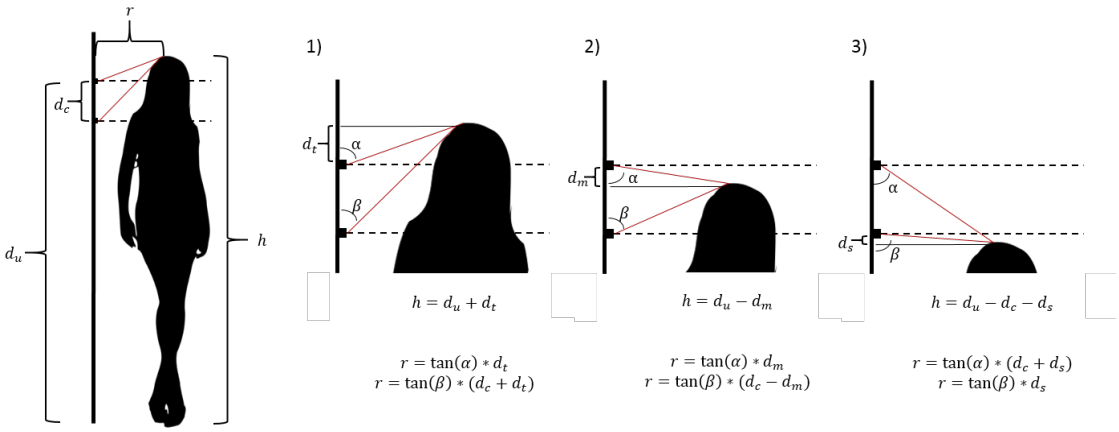


Figure 4.4: Lethe calculates the height of an individual by biangulating their distance from the camera and physical height using two cameras placed a distance of d_c apart. There are three potential positions a person can be in as it relates to the cameras: 1) above both cameras, 2) between the camera, and 3) below both cameras. For each scenario, Lethe determines the angles α and β from the relative heights detected by the cameras and solves the corresponding system of equations to determine a person's physical height.

Lethe determines how far away a person is from the cameras and then estimates their physical height from the relative pixel heights. Lethe calculates this estimate by dividing the detected pixel heights into 3 cases as seen in Figure 4.4: 1) above the cameras, 2) between the cameras, and 3) below the cameras. For this calculation, the reported pixel value h_{pixel} for the upper and lower camera is converted into h_u and h_l respectively. The conversion places the pixel that views the angle parallel to the floor (the dashed line in Figure 4.4) as zero, with pixels above that point positive and below that point negative. This conversion needs to be calibrated based on the angle of the cameras when they are installed. After the conversion, the angles α and β (in degrees) are calculated from the pixel heights of the upper, h_u, and lower, h_l, cameras:

$$\alpha = \begin{cases} 90 - p_w h_u, & \text{if } h_u > 0 \\ 90 - p_w abs(h_u), & \text{if } h_u \leq 0 \end{cases} \tag{4.1}$$

$$\beta = \begin{cases} 90 - p_w h_l, & \text{if } h_l \geq 0 \\ 90 - p_w abs(h_l), & \text{if } h_l < 0 \end{cases} \tag{4.2}$$

where p_w is the width of each pixel in degrees (e.g., 0.6375 degrees/pixel). Lethe then uses these angles to solve a system of equations in each of the three cases to determine the location of the top of a person's head relative to the cameras. When the person is above both cameras (Figure 4.4.1) we solve for d_t in:

$$r = tan(\alpha) * d_t \tag{4.3}$$

$$r = tan(\beta) * (d_c + d_t) \tag{4.4}$$

When the person is between the cameras (Figure 4.4.2) we solve for d_m in:

$$r = tan(\alpha) * d_m \tag{4.5}$$

$$r = tan(\beta) * (d_c - d_m) \tag{4.6}$$

When the person is below both cameras (Figure 4.4.3) we solve for d_s in:

$$r = tan(\alpha) * (d_c + d_s) \tag{4.7}$$

$$r = tan(\beta) * d_s \tag{4.8}$$

Finally, Lethe takes the distance values calculated in each case and calculates the person's height h:

$$h = \begin{cases} d_u + \frac{tan(90-p_w h_l)*d_c}{tan(90-p_w h_u)+tan(90-p_w h_l)}, & \text{if } h_u > 0, h_l > 0 \\ d_u - \frac{tan(90-p_w h_l)*d_c}{tan(90-p_w abs(h_u))+tan(90-p_w h_l)}, & \text{if } h_u \leq 0, h_l \geq 0 \\ d_u - d_c - \frac{tan(90-p_w abs(h_u))*d_c}{tan(90-p_w abs(h_l))+tan(90-p_w abs(h_u))}, & \text{if } h_u < 0, h_l < 0 \end{cases} \tag{4.9}$$

where d_u is the height of the upper camera and d_c is the distance between the two cameras.

For some events, both cameras are unable to get an accurate reading of height. Occasionally, this is due to a missed detection either through lost frames or a person walking too quickly. More often, one camera is unable to determine the pixel height of the person, because the top of their head is either above or below the frame. Specifically, people with heights outside of Lethe's range of vision cannot be measured in the lower or upper camera, respectively, when they are too close to the cameras. When this occurs, Lethe generates a height based on monocular vision: an estimate made from only one camera. Since we cannot calculate the distance from the camera in this scenario, the midpoint of the doorway is chosen and height is calculated based on this assumption.

Identity Mapping

Finally, we determine the identity of the person who produced height h by mapping their biangulated height to their measured standing height. This step is required since a person's height as they walk through a doorway does not exactly match their standing height. Often when a person walks, their measured height shortens proportionally to the length of their stride. Hence, a distribution of measured heights is produced

for a single person. Lethe must map the heights found by the cameras to the known heights of individuals in the space. This mapping is performed using k-means clustering, with the number of components equal to the number of people in the home. The mean biangulated height, h_m, of each cluster from k-means is mapped linearly to the known real heights, h_r. We map these heights based on the relative heights of individuals: the taller individual in the space maps to the taller cluster. We then solve this linear system:

$$h_r^1 = m * h_m^1 + b \tag{4.10}$$

$$h_r^2 = m * h_m^2 + b \tag{4.11}$$

to linearly map all biangulated heights into the real heights. This method requires a small calibration phase before tracking can be performed to ensure enough measurements have been taken of each individual in the doorway to create an accurate map. However, this calibration can be performed periodically and automatically while the system is in operation as long as all individuals in the space use the door before heights are mapped.

4.3.3 Memory, Data Transmission, and Privacy

The Lethe thermal camera requires only 21 values in memory for its operation. For the detection algorithms, 4 values are required for presence detection (t_s, t_e, r_{back}, t_last), 12 values for direction (loc_{l1}^i, loc_{r1}^i, loc_{l1}^{i-1}, loc_{r1}^{i-1}, d_{ind1}, d_{sum1}, loc_{l2}^i, loc_{r2}^i, loc_{l2}^{i-1}, loc_{r2}^{i-1}, d_{ind2}, d_{sum2}), and a single value for pixel height detection (h_{pixel}). Additionally, the thermal camera requires space to store the current pixel temperature, the current location of the pixel (row and column), and the current time of the device. Reference information, such as the height and width of the image, the lag time allowed in frames, and the r_{thresh} value can be kept statically. Many of the dynamic values, depending on the size of the image, can be represented using a single byte. Only the timestamps require a larger storage space at 4 bytes a piece. Hence the memory requirement for the Lethe thermal camera is only $M = 33$ bytes as seen in Table 4.1. In the 60×80 pixel thermal camera used in this work, that means only 33 pixels (0.69%) of the image can ever be stored on the device.

While limiting storage decreases the amount of image data that can be stored on the device, it does not guarantee privacy. If the network connection was fast enough to offload streaming data of each pixel as it was collected the image could be reconstructed at another location and could bypass local memory storage. Hence, the limited memory implementation of the thermal camera must be paired with a low data rate hardware transmitter to ensure that streaming image data is not possible. Since only events are transmitted

Figure 4.5: The two thermal cameras were placed vertically along a cubicle threshold and wired to two attached RaspberryPis (left). The FLIR imaging sensor and breakout board are shown to the right.

from the device and events are assumed to last at least 0.366 seconds and each event requires only 10 bytes for transmission, this data rate need only be $R = 30$ bytes/second. Hence, with the combination of limited memory storage and low transmission rate we can guarantee that no full thermal image can be pulled off the device (only 63 bytes at most). With this guarantee, we can mitigate concerns of privacy related to the image data.

4.4 Experimental Setup

To evaluate if Lethe can detect human crossings, filter non-crossings, and determine the identity of individuals while preserving image privacy we implement and collect data from a paired thermal camera setup. We collect two different datasets for evaluation: a mixed crossing/non-crossing dataset and height variation dataset, described below. Both experiments were conducted in an instrumented cubicle doorway in a university office building, shown in Figure 4.5. Data from the cameras was offloaded to an external server for processing and analysis. While a full implementation of Lethe would perform the crossing detection and pixel height detection steps on the physical camera, all steps were processed offline for this evaluation. Details of the thermal camera hardware, ground truth collection, two experiments, and evaluation metrics are included below.

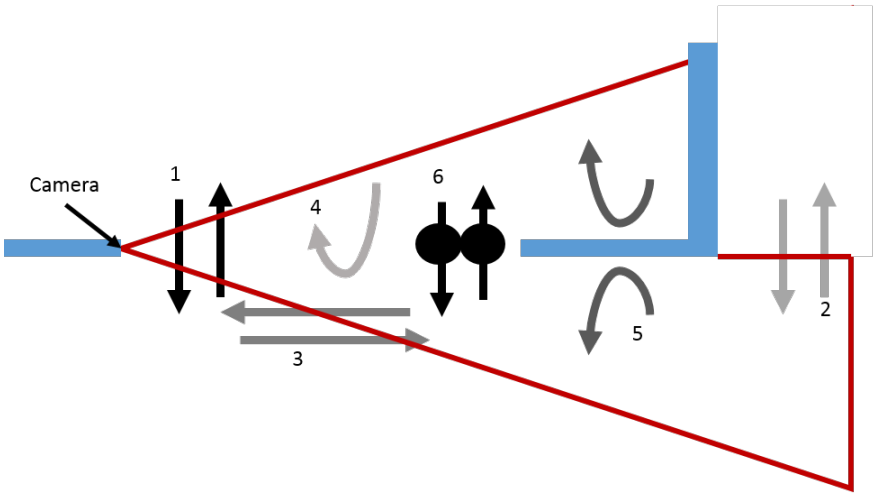


Figure 4.6: There are a variety of ways a person could be in the field of view of the camera. These events can be true crossings such as: the subject walking through the doorway (1) or walking into the doorway, standing for a period of time and then crossing (6). Or these events could be a non-crossings such as: crossing through another doorway adjacent to the doorway with the sensor (2), walking by door but not through it (3), walking into the doorway and then walking out (4), or two people walking up to doorway at the same time but not crossing (5). We collected data for each of these cases when evaluating Lethe's crossing detection.

4.4.1 Thermal Camera Implementation

The thermal camera sensors used in our experiments were assembled using a FLIR Dev Kit and a Raspberry Pi 1. An image of the setup is included in Figure 4.5. The FLIR Dev Kit consists of breakout board as well as a Lepton longwave infrared (LWIR) imager that typically consumes 150 mW - 160 mW. The thermal imager captures infrared radiation input in its nominal response wavelength band (from 8 to 14 microns) and has a thermal sensitivity of 50 milli-Kelvins. The imager has an effective frame rate of 8.6 Hz and a resolution of 60 (w) $\times$ 80 (h) active pixels (each 17 μm in size and covering a 0.6375 degree angle). Frames obtained by the thermal imager are sent over a SPI port to the Raspberry Pi, where each frame is timestamped. Once the images have been collected, we export the data to an external server for processing and evaluation.

4.4.2 Ground Truth

Ground truth for the experiments was collected using a slide advancer and keylogger logging script. Each click would be recorded with a UNIX timestamp to denote when an event took place. Just before a participant would enter the field of view of the camera, the experimenter would log the direction in which they entered the doorway. When a participant left the field of view, the experimenter would log the direction they exited the doorway. These two data points together would record the direction of movement. For each participant, a log of start time, end time, direction, and identity was created for each crossing they made during the study.

4.4.3 Height Variation Data Collection

To evaluate the accuracy of Lethe's identity detection, we collected 40 crossing events from 21 different participants, for a total of 840 crossings. Participants were asked to walk 10 times each over 4 marked lines at distances of 84cm, 63cm, 42cm, and 22cm from the thermal cameras. They walked over each marker 10 times before moving to the next, in order of greatest to least distance from the camera. Each participant performed the walking steps in the same order. No other directions were given to participants about their walking behavior, speed, or body posture. As reference, each participant was also asked to stand on each of the marked lines facing the camera for 1 second.

Participants ranged in height from 163cm to 181cm and were chosen from among graduate and undergraduate students present in the building at the time of the experiments. Of the participants, 7 were female and 14 were male. Participants were not asked to change their clothes or hairstyles for the experiment. Overall, 2 of the males had a medium-long haircut (˜3cm), 6 had a medium men's haircut, 3 a short cut, 1 a buzz cut, and 1 was bald. One male had below-shoulder-length hair pulled back into a pony tail. Of the women, 1 had shoulder length hair, 3 had hair below their shoulders, 1 wore a hijab/headscarf, and 2 had their hair pulled up into a bun. Many of the participants wore glasses. Participants walked with whatever footwear they had on at the time, and their heights were measured in that footwear.

Data collection from participants was spread out over the course of seven days. Data was collected at varying times of day: morning, afternoon, and night. We made no effort to control the temperature in the room beyond the building set thermostat (at 21°C). The thermal cameras themselves faced the large external windows as seen in Figure 4.12.

4.4.4 Mixed Crossing/Non-crossing Data Collection

To evaluate the accuracy of Lethe's crossing detection, we collected 220 crossing events and 440 non-crossing events where a person was in the field of view of the camera from 11 participants. The design of the non-crossing events was informed by a prior experiment in an Institutional Review Board approved study where we collected RGB video data of participants living in an instrumented home over a period of three weeks. In this data, we saw many instances of people walking by the doorway and leaning into the doorway but not walking through. For the evaluation of Lethe, we simulated those behaviors. As seen in Figure 4.6 these non-crossings include: (2) crossing a doorway or hallway beyond the monitored threshold, (3) walking parallel to the doorway, (4) leaning into but not crossing the threshold, and (5) two people in the field of view on either side of the camera. For this dataset we also collect normal crossings (Figure 4.6.1) and delayed crossings (Figure 4.6(6)) where a person pauses in the doorway to talk or look around before continuing

through. While we note this study is not a replacement for a full in-situ evaluation, it does provide an understanding of how Lethe might perform in a real world deployment with real human behavior. We also note that unusual crossings where two people walk through the doorways side-by-side or in quick, overlapping succession can occur. However, only 0.2% of the in-situ crossings we used to inform this study had that behavior and therefore we did not include it in this study.

Participants for this study are a subset of the height variation dataset participants. 11 of the height variation participants were willing and able to participate in this longer mixed crossing study. Of these participants, 4 were female and 7 were male. They included 2 each of the long, medium, and short men's haircuts, the one male with a long pony-tail, and 1 each of the female hair styles. Participants were directed through each of the mixed crossings in the same order, performing 10 of each type before moving on. Participants were not directed in how to perform each task, just told things like "pretend to lean in and turn on a light" or "pretend to stop in the doorway and talk to someone". Participants were told they could walk at whatever distance they chose from the thermal cameras during this experiment. Most walked near the center of the doorway while 2 chose to walk at a different location each time they walked through.

4.4.5 Evaluation Metrics

We use four main metrics to evaluate Lethe: *crossing accuracy*, *direction accuracy*, *identification accuracy*, and *transition accuracy*. Crossing accuracy is the classification accuracy of crossing detection and is used to evaluate if Lethe can differentiate crossing and non-crossing events. Direction accuracy is the percentage of correctly determined directions for all detected crossing events. We use it to evaluate direction detection. Identification accuracy is the classification accuracy of identity detection when an event is mapped to a person using identity mapping. It represents the overall ability of Lethe to identify an individual. Transition accuracy is the combination of these metrics and represents the accuracy that would be transferred to a higher level application. It is the percentage of crossings that are detected with the correct identity and direction, over the total number of ground truth events.

4.5 Results

Our results show that Lethe can detect the identity of individuals and track them in an indoor space despite the severe memory constraints used to preserve privacy. When combining both height and crossing studies. Lethe has an overall direction accuracy of 99.7% and an overall crossing accuracy of 96.9%. As identification accuracy depends on the heights of the people using the doorway (people of the same height are difficult to differentiate), we present a best case example of identity and transition accuracy for 3 of our participants in

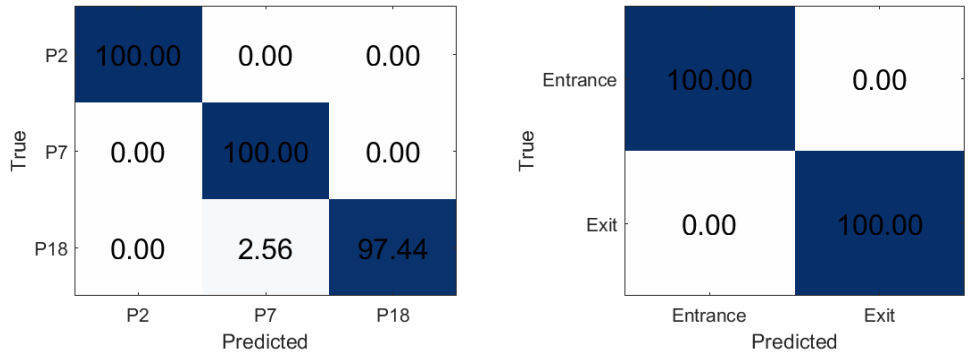


Figure 4.7: Lethe can identify 3 people with a ˜6cm difference in height with 99.1% identification accuracy (left). Combined with a direction accuracy of 100% (right) and a crossing accuracy of 95.1% (6 events went undetected), Lethe can identify who, when, and in which direction a person travels (the transition accuracy) in this best case with 94.3% accuracy.

Figure 4.7. Here, Lethe is able to identify and track these individuals with 94.3% transition accuracy, 95.1% crossing accuracy, 99.1% identification accuracy, and 100% direction accuracy. On average, Lethe has an identification accuracy of 96.0% for 2 people with a 5cm (˜2in) or greater difference in walking height and 92.9% with a 2.5cm (˜1in) or greater difference.

The following subsections discuss the results of our height variation and mixed crossing/non-crossing experiments. Additionally, we discuss how we expect a thermal camera of this type to perform in the future as frame rates and resolution increases and why memory limited identification and tracking is difficult with an RGB imager.

4.5.1 Crossing Detection

The results from our crossing data show that Lethe is able to correctly detect crossing and non-crossing events with 94.7% accuracy. A confusion matrix of the data is presented in Figure 4.8. Misclassifications were caused mainly by an insufficient vertical viewing angle in the cameras.

The majority of the misclassified events, 91.4%, were non-crossings that simulated turning on/off a light switch through the doorway, Figure 4.6(4). In this simulation, participants reached through the doorway to pretend to turn on a light switch and alternated this motion on the far wall from the camera and the wall the cameras were attached to. These non-crossings were misclassified in two ways: 1) participants left the view of the upper camera and not the lower and 2) participants left the view of both cameras when leaning through the doorway. These events covered 38% and 56% of misclassified non-crossings respectively and were caused by an insufficient vertical viewing angle in the cameras. The vertical viewing angle defines the space a camera can see above and below its location. With a viewing angle of 180°, the camera could see a person no

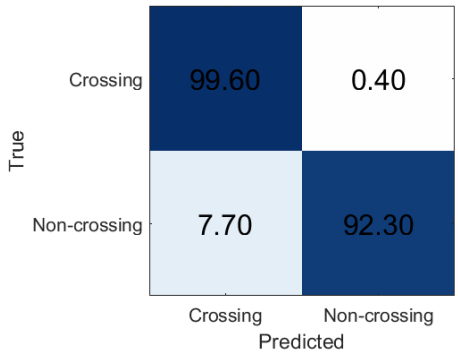


Figure 4.8: Lethe is able to identify crossing and non-crossing events with 94.7% classification accuracy.

matter how close they are to the wall the camera is attached to. However, at lesser angles a person could hide in the gap between the field view and the wall. In the first case of misclassified non-crossings, when the participants leaned through the doorway to flick on a light switch they moved into this blind spot for the upper camera. Hence, the upper camera saw two crossings rather than one non-crossing – the person leaning in and the same person leaning back out. Lethe pairs one of these events with the lower camera's non-crossing during biangulation and removes it, but retains the other as a crossing. A more complex pairing algorithm for biangulation could improve results. The second case of misclassified non-crossings had participants leave the viewing angle of both cameras when they reach down to turn on a light switch. Again, a larger viewing angle could be the solution to identifying these as non-crossings. However, as long as both events are labeled as crossings (i.e., leaning in as an entrance and leaning out as an exit) any room-level tracking algorithm could count both events as crossings without affecting the overall accuracy of a tracking application. Hence, we leave these mislabeled events to be solved in future hardware iterations where the viewing angle of the cameras can be increased or to be incorporated as normal events in a room-level tracking algorithm.

Insufficient frame rate caused the misclassification of the final 2 events – one misclassified crossing and one misclassified non-crossing. A frame rate is insufficient when a person can move through the frame faster than the thermal camera can record. This poses a problem particularly when a person is not caught moving in one of the two direction zones on either side of the doorway. Since Lethe's non-crossing detection algorithm relies on detecting the motion of a person on either side of the doorway, crossing events where the person is not detected moving on one side are labeled non-crossings. This caused one crossing event to be labeled a non-crossing. For non-crossing events, if Lethe detects that they have crossed into both detection regions, but they pull back too quickly to be detected leaving, then the event will be mislabeled a crossing. While these mislabeled events are dependent on the walking speed of a person, higher frame rate thermal cameras will be able to reduce the occurrence of this mislabeling in future implementations.

4.5.2 Identity Detection

For the height variation study, we first examine how well biangulation with the stereo cameras estimated a participant's height. Then we evaluate how well people can be differentiated with these height measurements. Finally, we discuss how an individual's standing height relates to their walking height and what this means for identification by height in the real world. Overall, Lethe can differentiate 2 individuals 96.0% of the time with a 5cm (˜2in) or greater difference in walking height and 92.9% with a 2.5cm (˜1in) or greater difference.

We evaluate Lethe's accuracy at height measurement in terms of how well the participant's detected heights are clustered. The tighter a cluster of heights produced by a single person, the closer in height two people can be and still be differentiated. Figure 4.9 shows the standard deviation in the detected heights output from biangulation (labeled Binocular Height). The binocular heights are compared to the monocular height estimated from the upper and lower cameras – where a person is assumed to be at the midpoint of the doorway threshold. For our participants 10/21 have 95% of their biangulated heights in a range of 8cm and 17/21 have 95% within 12cm. In most cases the higher standard deviation (P1, P4, P12, P13, P15) is caused by 1-5 (out of 40) outlier measurements of a person's height. For three others (P9, P16, P18) the measurements themselves are more variable, the distribution wider. This is generally caused by how each individual walks – their gait, head movements, and subsequent walking height is more variable. The final two participants with higher standard deviation in height measurements were P20 and P21. These participants were the tallest in our study at 177cm and 181cm respectively. As they moved closer to the cameras, both participant's heads moved above the viewing range of the lower camera and even above the upper camera at the closest distance for P21. In these situations, the biangulation step in identity detection defaulted to using only the upper camera to estimate the monocular height. Like the errors in crossing detection, this is caused by the vertical viewing range of the cameras. The top of participant's heads would appear in at least one of the camera's blind spots. Though we did not have any participants short enough in this study (P1 was 163cm), the same would be true of short people walking too close to the camera. Here, their binocular estimate would approach the lower monocular estimate. Unfortunately, this cannot be fixed algorithmically – if a person is out of the field of view, we do not have enough information to estimate their heights. Fortunately, future iterations of the hardware could be created with a larger viewing angle or be implemented with more than two thermal cameras to cover the full range of human heights.

Next, we evaluate how well people can be differentiated with these height measurements by performing a 21 choose 2 evaluation of the data from all participants. Here, we pair each participant with every other and evaluate their identification accuracy as compared to their walking heights (Figure 4.10). The walking height is selected at the median height measured by the cameras. Since height is a weak biometric, we don't expect

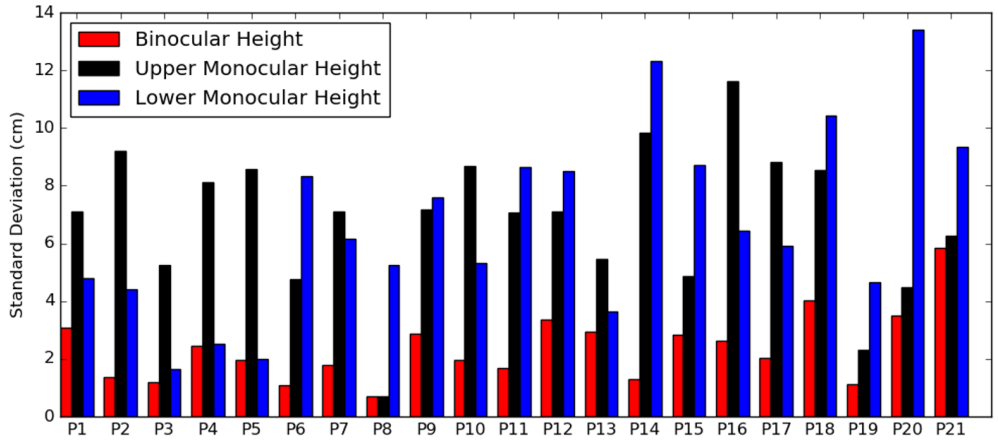


Figure 4.9: Most participants (17/21) had a standard deviation of 3cm or less in their measured walking heights. The participants are ordered by height and range from 164cm (P1) to 181cm (P20). Overall, binocular height estimation produced more consistent heights than either camera alone (monocular).

that Lethe can reliably differentiate people of the same walking height in the identity mapping stage. This is visible in Figure 4.10 in height differences below 2.5cm where the identification accuracy averages a little over 50%. However, Lethe can reliably differentiate participants when their height differences are above 2.5cm. Overall, Lethe can differentiate 2 individuals 96.0% of the time with a 5cm (˜2in) or greater difference in walking height and 92.9% with a 2.5cm (˜1in) or greater difference.

We present the identification analysis based on walking height because a person's standing height (as measured in a doctor's office) is rarely the height they exhibit when walking. As a person's stride increases, their walking height decreases since they rarely reach a full standing position while in motion. Additionally, while a person's head is perfectly upright and straight for a standing height measurement, during walking people will often lower their head to look where they are going. This behavior was often noticed in our experiments, where participants would look down to make sure they were walking over the correct mark on the floor. On average, our participants had a walking height 2cm lower than their standing height with a 1.2cm standard deviation in this average. Of those with walking heights more than 2cm below their standing height (12 total), 10 were caused by participants tilting their heads to look down at the floor. For the other two (P5 and P18), this was caused by heat insulation due to a hijab and hair respectively. We discuss this in more detail in Section 4.5.3. Overall, walking behaviors and habits do affect the difference between a person's walking and standing heights. This includes not only head movements and stride, but also height changes caused by footwear. Since Lethe uses the walking height to differentiate individuals, we can perform identification as long as a relative height difference of 2.5cm or greater exists while people are walking. In studies in the US and UK, it's been found that most couples have standing height differences greater than

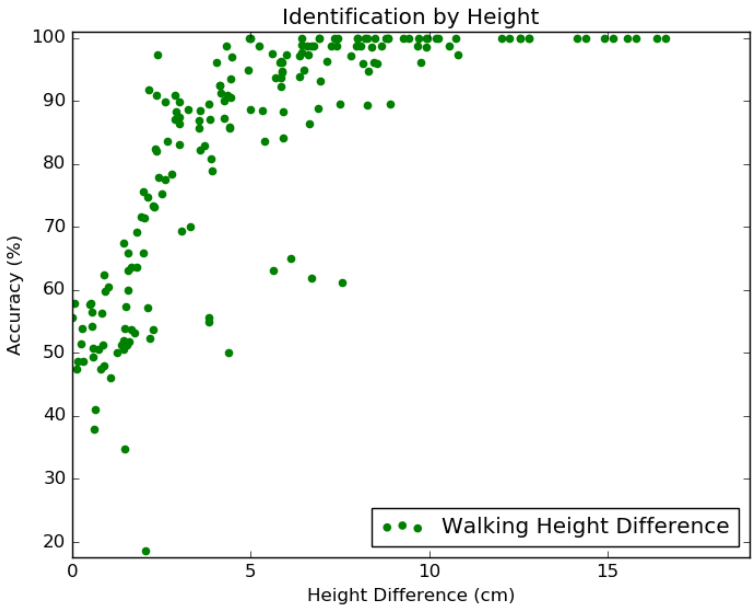


Figure 4.10: On average for our participants, Lethe can differentiate 2 individuals 96.0% of the time with a 5cm (˜2in) or greater difference in walking height and 92.9% of the time with a 2.5cm (˜1in) or greater difference.

5cm [144, 145]. This majority height difference allows for greater height variations during walking while still allowing Lethe to distinguish the majority of these couples by height

4.5.3 Hair, Head Coverings, and Background Temperature

Lethe relies on the assumption that the human body is a higher temperature than the background environment. Indoors, this is generally true. We tend to keep air temperatures in the range of 20°C to 22°C and most objects in the environment conform to this temperature. Additionally, human skin ranges from 32°C to 34°C making skin easily differentiable from the environment. However, since Lethe is measuring height, we want to measure the top of a person's head rather than their skin – which may be covered by hair or cloth. We found that hair commonly absorbed the heat from a person's head, ranging in temperature from 25°C to 30°C. For 19 of our participants, including many with shoulder length or lower hair, this was enough to differentiate their hair from the background temperature and detect their height. However, for two of our participants (P5 and P18) we were unable to consistently detect the top of their head.

Participant P5 was the only participant wearing a hijab (i.e., a scarf wrapped around the head and tied under the chin) in our dataset. For P5, we consistently measured her forehead instead of the top of her head. The consistency of this meant the standard deviation of her walking height was relatively low (2cm in Figure 4.9) but she had the highest difference between her standing height and measured walking height

at 3.7cm. Additionally, while this was exhibited infrequently in the collected data, when she turned away from the camera and only her hijab was visible Lethe was unable to detect her presence as different from the background. While P5 did inform us that most women do not wear their hijabs inside their homes (i.e., accurate height measurement is possible for her in residential spaces) this does show that Lethe is limited in detecting accurate heights when people wear head coverings.

Like P5, Lethe often measured participant P18 at their forehead rather than the top of their head. This was due to P18's thick, curly hair. P18, the only male in our study with long hair, had thick curly hair pulled back into a pony tail at the base of his neck. Like P5, P18 was measured at forehead height fairly consistently throughout the study, with a measured walking height 3.6cm lower than his standing height. P18 also had one of the largest variations in measured height (Figure 4.9), though that was more likely due to his fast, bouncy gait than Lethe's inability to measure the top of his head. Because the thermal cameras used in this study have a low frame rate (8fps) a fast gait can mean Lethe only measures a few points of a person's gait when then cross. Hence, each crossing may not measure the top of a person's gait in each crossing and overall heights can be more variable.

For both P5 and P18, the background temperature of a space may change the likelihood that their hair or head covering would be detected by Lethe. If the background was cooler, Lethe may be able to detect the top of their heads. However, this is also true in reverse: as the background increases in temperature other hair types may become harder to detect. However, we believe that a general increase in background temperature beyond that seen in our dataset to be unlikely in conditioned indoor spaces. The data collected in this work faced a background of office cubicles and large windows (Figure 4.12) where the external air temperature ranged from 14°C to 32°C and solar gain was allowed to heat objects in the environment. Hence, our dataset is likely close to the upper range in terms of general background temperature. In contrast to general background temperature, heated objects (e.g., computers, objects with running motors, etc.) are a general limitation to Lethe's person detection. These objects could either increase the average background temperature, or be detected as people if they exhibit human skin or greater temperatures. Future work may look at detecting and dismissing these fairly stationary objects while still providing human tracking.

4.5.4 The Effect of Frame Rate and Resolution on Identity

Our crossing results showed that frame rate can have an effect on Lethe's crossing detection accuracy. Here, we attempt to analyze what effect a change in frame rate or resolution could have on height and identity detection. Intuitively, a lower resolution will provide a coarser grained estimate of a person's height. Hence, at lower resolutions the distance between two people in height must be larger to differentiate them. Additionally,

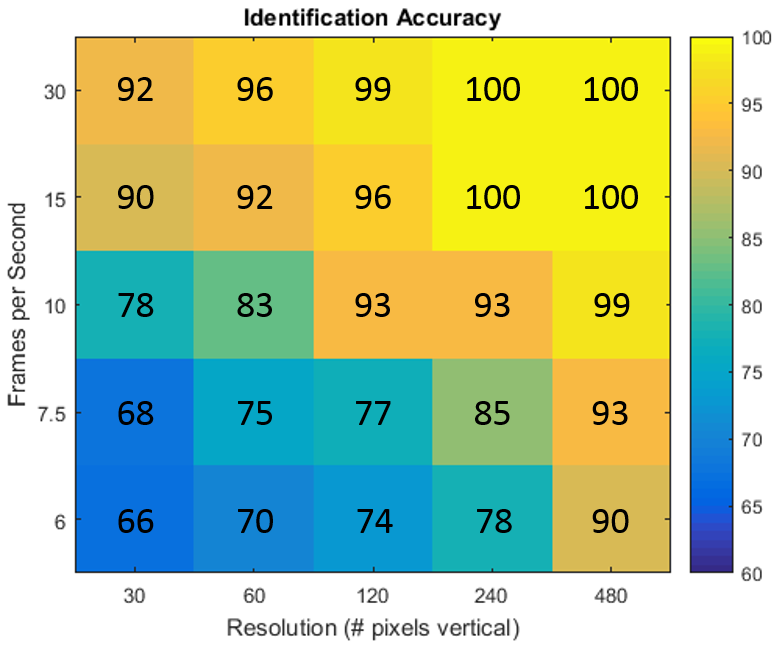


Figure 4.11: An increase in resolution increase the accuracy of identity detection as finer grained heights are measured. As frame rate increases, a more complete image of a person's heights during their stride is evaluated and measurements of their tallest height for each event becomes more consistent.

frame rate can have an effect on height accuracy though the reason is less intuitive. A person's height changes as they walk, reaching their maximum standing height when their feet are together and having shorter heights as their stride length increases. If the frame rate of a camera is low, a camera has a lesser chance of taking an image of a person when they are at their tallest, almost standing position in their stride. Additionally, the point in a person's stride where the camera takes an image could vary widely from event to event depending on the person's speed and stride in that particular crossing.

Figure 4.11 shows our estimate of how height and identification accuracy may change with resolution and frame rate. This graph was produced using an RGB camera, as we did not have a thermal camera that could operate at a higher frame rate than 8 fps or have higher resolution than 80 vertical pixels. Pixel heights were extracted from the camera using a column level background subtraction and estimated using monocular vision. The same identity mapping as Lethe was used to turn heights into identities. Figure 4.11 shows the identity accuracy for 2 participants (164cm and 180cm) as they perform 400 crossings. While we cannot make a direct comparison to Lethe since the processing and detection of heights are dissimilar, we can see that an increase in both frame rate and resolution increase the accuracy of identity detection. We predict that an increase in resolution or frame rate for Lethe will follow the same trend and future hardware will only make Lethe's identity detection more accurate.

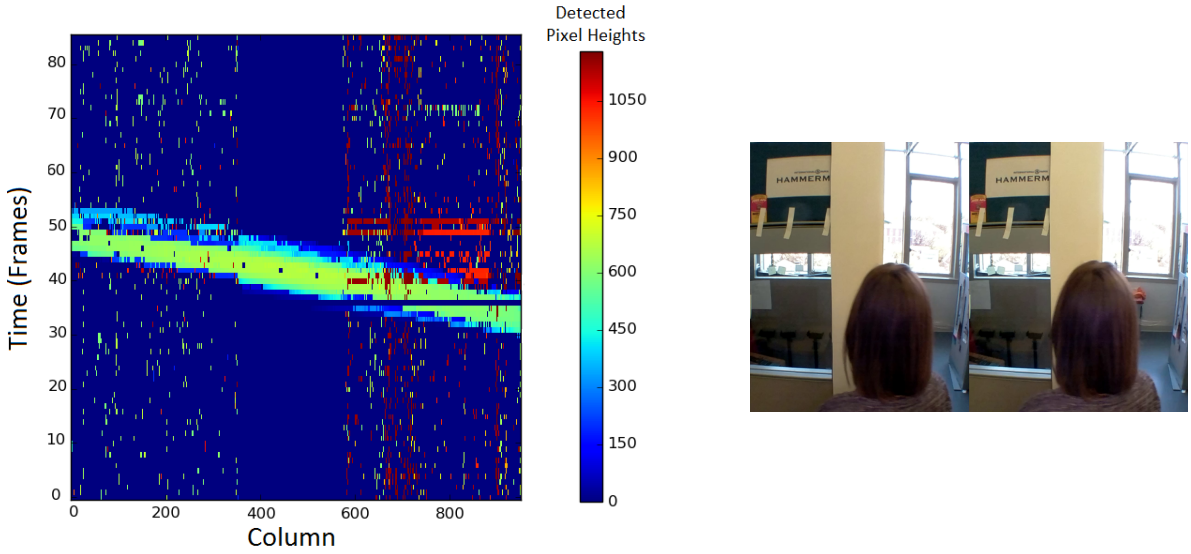


Figure 4.12: RGB cameras are sensitive to small lighting changes when processed in a memory limited fashion. Pixel heights measured from direct background subtraction in each column produce numerous noise values (left) instead of a clear image of the person's heights (left, horizontal streak). Only when the background is fully uniform are human heights alone detected (columns 380-580) and seen with the white poster board down the center of the image (right).

4.5.5 Why not use an RGB camera?

While we used an RGB camera to analyze the effect of frame rate and resolution, we found that visual cameras are not accurate when processed in a memory limited fashion. Because visual cameras do not detect body temperature, they require some type of foreground detection to identify a human presence in the pixels. While background subtraction has long been used to detect moving objects, such algorithms rely on either a perfect background (such as a green screen) or modeling to detect and filter out noise in the background data. If we perform background subtraction in a memory limited fashion (such as processing only a single column of the image) we cannot filter out these small variations. Hence, foreground pixels not caused by a human are mistakenly attributed to a person and recorded as pixel heights. Figure 4.12 shows an example of this, where each row represents a frame taken in time and each column represents a column in the image. Each column in time has been compared to the previous image of that column in time for direct background subtracting using limited memory. Only highest point of change is detected and recorded as a pixel height using this method. Here, even an unmoving background can have small changes in lighting that are detected as these pixel heights.

Typical background subtract algorithms filter out these noise pixels with a variety of methods, such as Gaussian mixture modeling. In fact, one could easily filter out the noise in Figure 4.12 to detect the horizontal streak of human heights. However, these noise filtering techniques cannot be performed in a memory limited

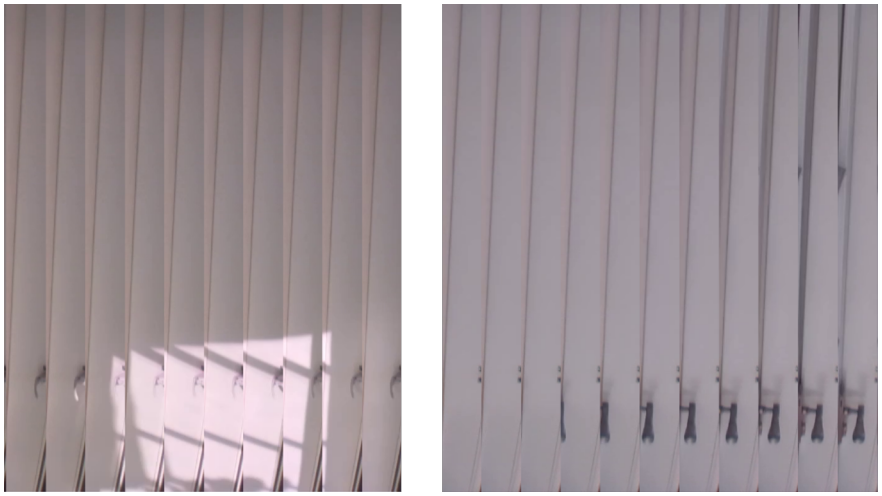


Figure 4.13: RGB cameras also suffer from lighting level changes and smaller movements that can be confused for the motion of people. In memory constrained devices, these false positives become difficult to identify and filter.

fashion. Even offloading the image in Figure 4.12 requires storing at least two full frames for background subtraction before the processed data can be transmitted off camera. Only with a near perfectly uniform background does this noise filtering cease to present a problem. This can be seen in the noiseless strip in Figure 4.12. However, this uniform background cannot be guaranteed or even expected when deploying such cameras in the wild.

Even if RGB camera data could be processed in a memory limited fashion to detect heights, they still suffer from environmental lighting problems. Figure 4.13 shows two scenarios in a doorway that will likely be detected as the movement of people when processed in a memory limited fashion. Additionally, the presence of these problems in RGB devices was a motivating reason for researchers to move to thermal cameras even without memory limitations. For these two reasons, RGB cameras are not ideal for detecting and identifying individuals in a memory limited and privacy-preserving device such as Lethe.

4.6 Limitations and Future Work

While the Lethe prototype shows the potential of privacy-preserving image processing in AmI systems, there is still work to be done. This work falls into two main categories: further development and evaluation of Lethe as a human tracking system and further exploration of the hardware based image-privacy preserving technique presented here on other applications and with other sensors.

For the Lethe prototype, future work would look at evaluating accuracy in in-situ environments. While our experiments found a fairly uniform background temperature that was not disturbed by the floor to ceiling

window in its field of view, the background may not be as uniform in all environments. Non-human hot objects, such as computers or laptops, may present a challenge in in-situ environments. They could either raise the average background temperature or mistakenly be identified as people. Future work may look at detecting and dismissing these fairly stationary objects while still providing human tracking – and how to do so while still preserving privacy with limited memory. Work in this direction may also lead towards solutions to multi-user phenomenon, such as a person sitting in the field of view of the cameras while another person crosses through the doorway. An initial two object expansion to tracking in Lethe, where the recorded values in memory are doubled, may be a first step in that direction. Future work would also include testing Lethe with a higher resolution and higher frame rate camera to see if our predictions of increased accuracy hold true. A 3 or greater camera system, as opposed to Lethe's two, could also provide an increase in accuracy. Additionally, future work will look at actually building a self-contained Lethe and measuring its form factor, efficacy, and energy consumption.

In addition to this potential future work, it must be noted that Lethe has a number of standard tracking limitations. For identification, the use of height as a biometric limits the number of households that can leverage identity information. While house level tracking algorithms leveraging floorplans can help to mitigate this for creating relative tracks (e.g., understanding that the same person went into the kitchen, bathroom, and bedroom in succession), people of the same height will always be difficult to identify [146]. Additionally, a person's height may legitimately change over the course of a day as they change their posture or footwear. This means that a person's walking height and distribution may be more variable in an in-situ environment. However, most couples have enough of a height differential to be tracked effectively by Lethe since it uses the relative height difference between these distributions to perform identification [144, 145]. Multi-user occlusion also presents a limitation for Lethe, where people walking side-by-side or in quick succession are mislabeled as only a single crossing. For Lethe, 0.366 seconds must pass before another person can cross the doorway and be accurately detected. Further sensing from multiple viewpoints may mitigate this problem. For example, a camera placed at the top of the doorway could identify when two people cross side by side. However, our in-situ study that informed the design of mixed-crossing dataset found that these occlusion behaviors occurred in only 0.2% of the crossings. While this may be larger in other homes, it does indicate that occlusion in residential homes may be a minor limitation. Finally, Lethe's use of thermal cameras limits its use to environments that are cooler than human temperatures. When background temperatures are too warm, or head coverings insulate a person's head too well, Lethe is unable to accurately measure a person's height. Determining how common such environments and head coverings are would be a direction for future work.

For privacy preserving image processing in general, future work could extend the concept to other

applications. First, we might explore whether privacy preserving thermal cameras could detect the presence of a person and determine their identity if the camera was used to monitor space other than a doorway, such as in front of an appliance or the bathroom mirror. Since the sensitivity of a thermal camera falls over distance, understanding the size and shape of a space that could be monitored in this fashion is an area of future work. Additionally, just as the two binocular cameras in Lethe can localize a person's heights, future work may look at localizing a person using privacy preserving image process in a space using cameras on two or more walls. Finally, would we be able to detect other human movements, such as coarse or fine grained gestures, with this approach to provide privacy preservation to existing thermal imaging technologies?

Future work could also look at extending the concept of memory limited image privacy preservation to other imaging sensors. We have shown that RGB cameras are not ideal for memory limited processing as they require uniform backgrounds or extensive background subtraction. However, depth cameras may be able to leverage memory limited processing for privacy preservation. Like thermal cameras, they capture a more limited view of a person's clothing and image than RGB – but are still capable of capturing privacy invasive information such as body posture and behavior when such information is irrelevant to the AmI application. Additionally, depth cameras are already being used to monitor people in tracking and gesture applications alongside thermal cameras in full image technologies [147, 121]. However, they may also be able to differentiate people on a per pixel basis. Instead of heat, they could detect people based on their distance from the camera under the assumption that the background is far and people are close. This would provide the ability to process the pixels in a streaming fashion while maintaining a memory smaller than the image size. However, processing of a single "pixel" for depth cameras is more computationally intensive than thermal processing as a point cloud must be processed for depth, and limitations on processing and frame rate speed would need to be evaluated for this sensing approach in future work. Overall, the exploration of other applications for an image-privacy preserving thermal camera and an image-privacy preserving depth camera are exciting directions for future work.

4.7 Conclusions

In this work we prototype Lethe: a memory limited, privacy-preserving, identity capable thermal camera. In a best case of our collected data, it can differentiate 3 individuals with only a 6cm height differential with 94.3% accuracy. On average, it can differentiate 2 individuals 96.0% of the time with a 5cm (˜2in) or greater difference in walking height and 92.9% with a 2.5cm (˜1in) or greater difference. Additionally, Lethe has an overall direction accuracy of 99.7% and an overall crossing accuracy of 96.9%. These results show that Lethe

can effectively track and identify people with a memory limited thermal camera, without risking image leaks and therefore preserving privacy.

As a proof of concept, Lethe demonstrates a hardware-based approach to privacy-preserving image processing in AmI systems. By limiting both the memory available onboard the camera and the data rate of camera communication, privacy-preserving image processing aims to prevent the hacking or leakage of private images by constraining the data collection and storage of data to only that consent by the users are needed for the application. At most, a hacked privacy-preserving camera could leak $M + R$ bytes per time period Δt, but can still extract useful information using M bytes of memory. We anticipate this hardware-based approach to privacy-preserving image processing can generalize beyond thermal cameras and Lethe, helping to advance the use of information-rich image sensors while adding to the portfolio of techniques used to solve the privacy vs. data collection challenge in AmI systems.

Chapter 5

Leveraging Similarity for Learning in AmI

5.1 Introduction

As the rise in sensing and control systems in our homes, offices, cars and environment has skyrocketed, more and more systems are facing off against another AmI concern: human preferences. Systems that previously concerned themselves with the average preference (such as building temperatures formulated for the average male [9]) now have the sensing and control capabilities to observed information about and control for individual preferences [148]. Knowing these preferences can help meet societal needs like buildings that save energy by tuning systems to desired temperatures and preventing a user from manually disabling or undermining a building control system because it doesn't meet their needs [149, 150]. Unfortunately, *learning* those individual preferences is not as simple as asking a system user to fill out a survey. The state space of possible environments is often huge, people can't always articulate their preferences on the spot, and those preferences may change drastically and/or frequently over time. One of the most promising technologies for learning preferences in these environments is a machine learning mechanism called reinforcement learning. Reinforcement learning (RL) can learn user preferences over time by observing their interactions with the AmI system naturally, learning preferences from these interactions, and performing actions according to a policy that best meets those preferences. However, systems tied to a person's physical environment often suffer from a *small data* problem: observations of behavior may only occur on a daily, weekly, or even monthly basis. For example, if a system wants to learn a person's preferences when it's 6 am and sunny (maybe to control the blinds or give them clothing recommendations) it can sample that data at most once a day and

significantly less frequently in the winter or in cloudy climates. Additionally, people's preferences may change due to variables the RL system is not observing–such as exercise level, fashion, climate, sun location, and age changing a person's preferred thermostat temperature. Learning these changing user preferences quickly (especially before the next preference change occurs) becomes a vital problem to solve for personalized control in AmI systems.

Reinforcement learning algorithms can learn some preferences in these dynamic environments and many RL algorithms are designed to adjust to new preferences as user responses change over time with exploration heuristics. However, current approaches do not learn fast enough in environments with a small data problem. The most popular traditional RL approach, the model-free Q-learning [151] and its variants, require that each state in an application be visited many times to learn and propagate information about the best actions to take in that state. In applications with a small data problem, this process can be prohibitively slow – especially if preferences will change in the middle of learning. While there are various techniques to increase the sample rate of experienced states (mainly through reprocessing old experiences [152, 153, 154]), model-based approaches bypass this problem by recording a model of the application and propagating information about all states for the policy at every iteration. However, both approaches are still limited by the few interactions a user is capable of experiencing. Neither systems can test out two different actions on the same user (e.g. raising the blinds or leaving them closed) when they are experiencing a particular context. The systems must wait for that context to appear again to try the other action. Additionally, when preferences change both approaches retain old, incorrect information that is propagated along with the newly updated states and preferences. A common approach to dealing with this old, incorrect information is what we'll call *forgetting*. Essentially, when a change in preferences is detected the old, incorrect Q record or model of information is thrown out and the system begins anew [155]. This can greatly benefit the learning process when most or all of the old model is outdated. However, in physical environments, much of the data in the model remains the same even across preference changes. For example, changing a person's desired temperature from 68 to 70 will not change their preferences above or below that temperature range. Hence, we need an approach that can quickly learn new preferences without discarding preferences that remain unchanged and can effectively sample the response to more actions than the single user can experience at one time.

To quickly learn changing and infrequently sampled preferences we present an approach called KindredRL. KindredRL leverages the idea that many people, particularly in physical environments, have similar preferences. Examples of this include desired air temperature, where people of similar age, body size, or fashion sense will have common temperature preferences [156, 157], common preferences for open or closed blinds over the course of the day or year for people with similar facing windows, and behavioral changes by people following the same media stories. KindredRL identifies these similar people and uses their interactions with

the RL system to update the modeled information of their similar neighbors. This collaborative approach will allow preferences that remain unchanged to persist (i.e. the algorithm does not forget), but new preferences in different states will be sampled by all the similar people in a neighborhood and quickly updated for all users. Prior approaches are designed to operate on each user independently and do not leverage any preference similarity. In contrast, KindredRL identifies similar users based when they respond similarly to the algorithm's actions and uses that similarity to effectively increase the sampling span and rate of a user with the responses of its behavioral kin.

Our results show that the collaborative approach of KindredRL improves the learning of preferences in a physical environment as preferences change. We evaluate KindredRL with three simulated building applications based on real-world data and processes: thermostat control for desired temperature, blinds control to decrease glare in office buildings, and controlling a water heater for load shifting. We evaluate the systems as they operate over a simulated year of data with up to 800 users. We show that KindredRL reduces the overall regret (i.e. the difference between the optimal actions the RL system could take and the actions it does take as it learns) below that of the independent and forgetting approaches. KindredRL reduces regret below the best corresponding baseline approach by as much as 60% and 67% for the thermostat and blinds applications respectively. It also provides an additional 4% monetary savings in the water heating application over the nearest baseline. Overall, we anticipate that KindredRL will be an important addition to help bring personalized control to physical environments and can push forward the adoption of AmI systems for meeting societal needs.

5.2 Related Work

Reinforcement learning has been studied extensively over the past few decades for use in a wide variety of applications, including robotics and biological data [158, 159]. As KindredRL focuses on using reinforcement learning in physical AmI environments with human interactions, we limit our discussion of related work applications to building and preference based systems. Additionally, we discuss work related to dealing with non-stationarity – the reinforcement learning term that includes the process of changing preferences. Finally, we discuss current work where RL uses collaboration for learning.

Many works use RL to control systems in buildings, particularly for HVAC control [160, 38]. Additionally, recent work has begun to use deep reinforcement learning for device localization and building energy management and control [161, 162, 163]. However, such systems rarely learn personal preferences and instead focus on energy and thermal reinforcement signals by instead learning the behavior of an entire building or system and aggregating any individual human behavior. RL systems that do focus on personal preferences do

so independently for each user, such as learning music playlist preferences or web recommendations [164, 165]. Additionally, most of these systems assume that preferences remain unchanging, or rely on the traditional exploration based adaptability of RL algorithms to respond to any small changes in preferences over a long period of time. None of these application works focus on dealing with changing preferences in physical environments with a small data problem.

Preference changes in the reinforcement learning realm often fall under the header of non-stationary change. Traditionally RL paradigms assume that the model they are learning is stationary – meaning that the rewards (or preferences) in that model and the underlying transitions (movement from one state to the next) of the environment remain unchanged [166]. Approaches that address non-stationarity often use exploration tactics, such as ϵ-greedy action selection, to continually look for and learn new preferences [166]. Approaches to speed up learning these changes in Q-learning often taken the form of experience replay – learning from recorded data multiple times to improve information propagation [152, 153, 154]. However, these heuristics still suffer from the small data problem as they can only evaluation one action at any given time. Additionally, these approaches do not differentiate replay data from old preferences. This means data used in experience replay may be contaminated with old preferences. Approaches that do differentiate old preferences rely on an approach called change or context detection [155, 167, 168]. These approaches detect when a change in preferences or transition probability occurs and begin a new model. Often the old model or data is discarded when the change is detected as maintaining old models is storage intensive and it is not guaranteed that preferences will repeat. This process can work well when the majority of old preferences are incorrect but can discard valuable data when portions of the old preferences remain unchanged. While all of these approaches can be beneficial in specific environments, none leverage the similarity between people and their responses for learning human preferences. We aim to complement related work on non-stationary data by leveraging similarity in applications where it is available.

Our preference based approach uses information about similar individuals to learn collaboratively. This is not to be confused with general collaborative reinforcement learning or multi-agent learning. Both paradigms handle multiple different 'people' or environments while learning, but neither approach looks to leverage their similarities for the benefit of the individual. Instead, their goal is often to predict the actions of other RL agents to improve their own learning and subsequent actions or to meet some global objective through coordinating multiple agents. Applications here include collaborative robots learning with RL [169] and multiple agents working on different components of the same problem (e.g. each agent controls a traffic light in the city), either cooperative and competitive, where joint actions of all agents become part of the state [170, 171]. Joint decisions for object tracking and search are also common in this area [172, 173]. Future iterations of KindredRL may be integrated with work in this area, both learning preferences collaboratively

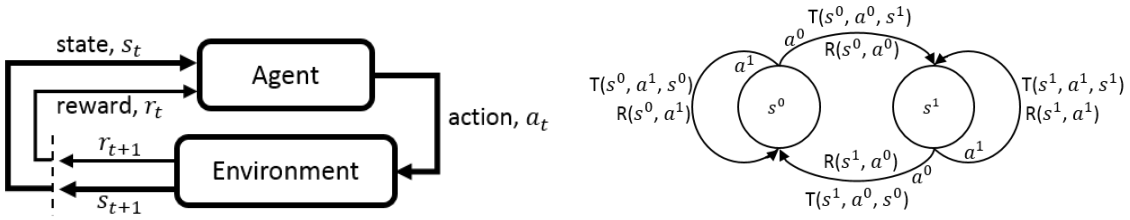


Figure 5.1: The general Reinforcement Learning Agent interacts with the environment through actions a, rewards r, and observations of states s (left) to learn the best actions to take in a Markov Decision Process environment (such as the example, right) with underlying rewards R and transition probabilities T.

and applying that learning in collaborative environments.

5.3 Reinforcement Learning Background Information

This section describes the general purpose and approach of reinforcement learning. It includes details on Markov decision processes (MDP), agents interacting with the environment, and obtaining an optimal policy from a model for model-based reinforcement learning. Readers who are already familiar with RL may wish to skim this section for the bulleted notation.

The goal of reinforcement learning is to learn what action to take in a given state to maximize the cumulative reward over the life of the algorithm's operation. Rewards are specific to the application–such as a negative reward when a user has to change the blinds. To obtain this reward, the algorithm performs an action a at given time t from its current state s and receives a reward r from the environment. This process is shown in Figure 5.1. The states, rewards, actions and transition probabilities are modeled as a Markov decision process. Formally, a Markov Decision Process (MDP) is generally used to describe this environment with the following terms:

- S is the set of states the process can be in.
- A is the set of actions that can be performed from the states, not all actions must be available in all states.
- $T_p(s, a, s')$ is the transition function, indicating the probability that you will get to state s' by taking action a in state s.
- $R_p(s, a)$ is the reward function, indicating the reward obtained by performing action a in state s. As the received reward does not always match the underlying function (e.g. when a user doesn't bother to give a response), we will use r to represent the reward received during operation.

0.1) initialize $\pi_{p,0} \leftarrow \vec{0} \; \forall \; p \in P$
0.2) initialize any model-based components
for *time step* t **do**
  for *person* $p \in P$ **do**
    1) observe state $s_{p,t}$
    2) choose action $a_{p,t}$ using the policy $\pi_{p,t}(s)$
    3) take action $a_{p,t}$, observe $r_{p,t}$, $s'_{p,t}$
    4) update any model-based components and the policy $\pi_{p,t+1}$
  end
end

Algorithm 1: Traditional RL algorithms operate independently on the users/environment and learn each model separately by interaction.

- $\gamma \in [0, 1]$ is the discount factor, indicating the priority of immediate rewards compared to later rewards. A lower discount factor allows for the selection of lower valued actions now to later gain the benefit of higher rewards in the future given that action path.

We add the subscript p to these definitions to denote which components are related to the individual person.

Reinforcement learning then uses interactions with the environment (and users) to turn this MPD into a specific policy $\pi_p(s)$ which maps states to high-value actions. This can be done by either directly accumulating information about the MDP model and performing a model-based planning process such as value iteration or interacting with the MDP to build a policy without directly storing the MPD constructs in a model-free learning process such as Q-learning [166].

All reinforcement learning algorithms interact with the environment by (1) observing the current state $s_{p,t}$ at time t, (2) choosing an action $a_{p,t}$ from the current policy $\pi_{p,t}$, (3) observing the reward for that action $r_{p,t}$ and the following state $s'_{p,t}$, and then (4) updating the policy $\pi_{p,t+1}$ and any retained model information for the next interaction. A duplicate of this process will operate on each user $p \in P$ in an independent environment (such as different offices where each user wants to control the blinds). Where the environment is not independent (such as different traffic lights sending cars to each other) different types of general collaborative reinforcement learning are used to maximize global and not just local reward [169]. These approaches focus on how different agents interact in the same environment as they perform actions that effect another agent's environment. For our work we focus on independent environments, leaving expansion into general collaborative reinforcement learning for future work. Hence, the basic algorithm for reinforcement learning on $|P|$ users is shown in Algorithm 1.

In this work, we implement our approach with both model-based and model-free learning. In the model-based approach, we build a model from our interactions with the environment and derive a policy from that model. In general, model-based approaches have been well studied in reinforcement learning. Many

of these approaches focus on learning portions of the MDP, such as modeling the transition function T to better predict the following state s' for a given state and action (i.e $\rho(s'|s,a)$) with approaches like Gaussian mixture models or neural networks [174]. Additionally, many works use various model-based techniques as function approximators to decrease the amount of data stored while learning. Our work focuses on the basic model-based approach: learning the components of an MDP and using a planning algorithm to derive a policy from that model. We do this to provide a basic starting point for KindredRL and expect that future works and applications can incorporate more complex modeling. To derive a policy $\pi_{p,t}$ from the MDP model we use standard value iteration. Value iteration takes an MDP and learns the optimal policy for that model. This will not necessarily be the optimal policy for the underlying model (i.e. the environment that the RL algorithm is interacting with) unless the recorded model matches the underlying model. It does, however, let us calculate the best action for our current level of knowledge. This process of updating the model and performing value iteration to obtain the current best policy $\pi_{p,t}$ comprise step (4) in the basic reinforcement learning process shown in Algorithm 1. Our specific approach to this will be described in more detail in the next section.

For our model-free implementation we use the classic Q-learning approach. Variations on Q-learning are the most commonly used RL in practical systems. It is fast, as it does not need to plan for a model, and requires no models of the underlying system to be created and stored. It instead updates the Q value of an action given a state by looking at the value of taking the estimated best action in the next state and adding that to the immediate value of the current action. This allows Q-learning to slowly (more so than a model based approach) learn the effect of future actions in order to best select the current action. Like model-based techniques, various Q-learning approaches use function approximators to decrease the amount of stored data while learning. Again, we will use the basic Q-learning approach to provide a starting point for KindredRL. As Q-learning does not have a model, step (4) in the basic reinforcement learning process shown in Algorithm 1 consists of just the update step for the Q vector. This Q represents the expected value of taking an action in a given state and selecting the best action for a state gives us the current policy $\pi_{p,t}$.

5.4 KindredRL: Leveraging Human Similarity to Learn Changing Behaviors

To quickly learn preferences in the face of limited preference samples and changing preferences KindredRL consists of three main components: the *learning framework*, *neighborhood creation*, and *collaborative preference estimation*. The *learning framework* describes the general RL components required to learn from the

Parameters: k, α, $horizon$, ϵ
0.1) initialize $\pi_{p,0} \leftarrow \vec{0} \; \forall \; p \in P$
0.2) initialize $R^{group}_{p,0}, T_{p,0} \leftarrow \vec{0} \; \forall \; p \in P$ or $Q^{group}_{p,0} \leftarrow \vec{0} \; \forall \; p \in P$
0.3) initialize $D_{p,p',0} \leftarrow 0 \; \forall \; p, p' \in P$
0.4) initialize $A^{p,p',0}_{(s,a)}, C^{p,p',0}_{(s,a)} \leftarrow 0 \; \forall \; p, p' \in P, (s,a) \in SxA$
for *time step* t **do**
  for *person* $p \in P$ **do**
    1) observe state $s_{p,t}$
    2) choose action $a_{p,t} \leftarrow \pi_{p,t}(s)$ with $1 - \epsilon$ probability, otherwise choose random action $a_{p,t}$
    3) take action $a_{p,t}$, observe $r_{p,t}, s'_{p,t}$
  end
  4) use Equations 5.3- 5.6 to update collaborative grouping information
  5) Model Based: update collaborative model component $R^{group}_{p,t}$ with Algorithm 3 and update non-collaborative model component $T_{p,t}$ with Equation 5.1- 5.2
  5) Q-learning: update collaborative $Q^{group}_{p,t}$ with Algorithm 3
  6) update policy to $\pi_{p,t+1}$
end

Algorithm 2: KindredRL performs actions on users independently and then uses their responses to derive neighborhoods of similar people and update a user's information according to the actions of their neighborhood.

environment. This includes all the steps seen in Algorithm 1 where the system observes the current state, selects an action based on the current policy, takes that action and observes the resulting reward and next state of the user's environment. In the case of a model based implementation, the framework also includes any non-collaborative model updates and the transformation of the model (including $R^{group}_{p,t}$ and $T_{p,t}$) into the current policy $\pi_{p,t}$. These techniques are standard to many RL approaches.

Neighborhood creation and collaborative preference estimation describe the novel collaborative components of this approach. *Neighborhood estimation* leverages actions of the learning framework and the responses of the users to estimate how similar users are. It relies on the idea that similar users will have similar responses to the same RL actions and that the RL algorithm will start to converge on similar policies for similar users. It then places users into neighborhoods, $N_{p,t}$, of users similar to user p based on this information. *Collaborative preference estimation* uses these neighborhoods to estimate the preferences of a single user by treating the information known about its neighbors as their own. It uses a replay approach that can apply to a variety of RL techniques. To show this versatility, we implement two approaches: a model-based approach and a Q-learning based approach. The model based approach creates the model component $R^{group}_{p,t}$ that is then fed back into the learning framework to generate the current policy with value iteration. The Q-learning approach directly estimates a collaborative $Q^{group}_{p,t}$ as the policy. This process then continues for every iteration–improving the model and subsequent policy–until the system is discontinued.

A typical iteration of KindredRL can be seen in Algorithm 2. Steps 0-3 and 6 are handled by the learning framework, while step 4 is described in neighborhood creation, and step 5 (for both a Model Based or

Q-learning approach) in collaborative preference estimation. These steps operate on the set of people P that are present in any given application. Details of each step are described below.

5.4.1 Learning Framework

The goal of the learning framework is to provide the method for the KindredRL's interaction with the user. It also defines how the non-collaborative model component T_p is learned and describes how the model is transformed into a policy for the model based approach. Like any reinforcement learning algorithm, KindredRL interacts with the environment and users through observing the current state, choosing an action based on the current policy, taking that action, and observing the resulting state and reward. To ensure that KindredRL does not exploit current knowledge to the point of ignoring potential greater reward, we implement selecting an action with the ϵ-greedy algorithm where the current best action for the policy is selected with $1-\epsilon$ probability and a random exploration action chosen otherwise. This allows KindredRL to slowly explore unobserved states and check lower value states to see if their reward has increased.

In the model-based approach, as KindredRL interacts with users it learns the model components $R^{group}_{p,t}$, the reward function, and $T_{p,t}$, the transition function. $R^{group}_{p,t}$ is learned in the collaborative fashion described in later sections. $T_{p,t}$ is learned independently for each user in a process similar to that defined in [155]. Here $T_{p,t}$ is continually adjusted towards a probability estimate using:

$$\Delta T_{p,t}(x) = \begin{cases} \frac{1-T_{p,t}(s,a,x)}{L^{p,t}_{(s,a)}+1} & x = s' \\ \frac{0-T_{p,t}(s,a,x)}{L^{p,t}_{(s,a)}+1} & x \neq s' \end{cases} \tag{5.1}$$

$$T_{p,t+1}(s,a,x) = T_{p,t}(s,a,x) + \Delta T_{p,t}(x), \forall x \in S \tag{5.2}$$

where $L^{p,t}_{(s,a)}$ represents the number of times that a (s,a) pair has been seen by p. Here, $L^{p,t+1}_{(s,a)} = min(L^{p,t}_{(s,a)}+1, M)$ whenever a (s,a) is seen to make the memory of past experiences finite (i.e. the current distribution of experiences will have a larger weight when M is small). The other model components S, A, and γ are defined based on the specific application being learned. Once $R^{group}_{p,t}$ and $T_{p,t}$ have been updated, the learning framework uses the standard RL approach of value-iteration to calculate a policy $\pi_{p,t+1}$ for all $p \in P$ as step 7 of Algorithm 2. This process is not special to KindredRL and any process that transforms a Markov decision process model into a policy could be used.

In the Q-learning approach, KindredRL simply turns the current $Q^{group}_{p,t}$ into the policy $\pi_{p,t+1}$ by selecting the action with the highest Q value as the current policy for a given state. Once the policy is updated, the

next iteration of KindredRL commences.

5.4.2 Neighborhood Creation

The goal of neighborhood creation is to find the neighborhood of similar people, $N_{p,t}$, for each user. While other approaches sometimes use information external to the RL process to perform grouping (e.g. metadata such as gender or age [175]), we aim to use only reward information gleaned from interactions with the environment. Hence, the only information we have with which to perform grouping are the state and actions of the RL algorithm, the (s, a) pairs, and the reward, $r_{p,t}$, given by the user p. Therefore, we define similarity in these terms as: two people are similar if they respond with a similar reward r to the same (s, a). The concern given this definition is that users may not experience the same (s, a) pair such that we can calculate this similarity. Additionally, if uses preferences can change over time then two similar users may not have similar r given a (s, a) is the rewards are collected at from the two users at very different times. Hence, we use two ideas to turn this definition of similarity into a similarity value. First, we only compare user responses within a short *simultaneous window*, w, where we have a smaller likelihood of observing a change in preferences. Given a new (s, a) pair experienced by user p at time t, if user p' experienced the same (s, a) pair in the window $[t - w, t]$ then the two users have experienced simultaneous responses and their similarity can be updated. Second, as the RL algorithm learns the users preferences the system will naturally converge on and commonly repeat the same state s and (s, a) pair for similar users. Essentially, if a user likes the temperature at 69 the RL system will converge on the state over time and perform actions to remain in that state – and it will do the same for a similar user. Hence, we say two users are dissimilar if they often respond differently to the same (s, a) within the simultaneous window and rarely share the same (s, a) pair. To calculate these values we update both the count of times they have experienced simultaneous (s, a) pairs, $C^{p,p',t}_{(s,a)}$, and the average difference in their responses, $A^{p,p',t}_{(s,a)}$. Formally,

$$C^{p,p',t+1}_{(s,a)} = C^{p,p',t}_{(s,a)} + 1 \tag{5.3}$$

$$A^{p,p',t+1}_{(s,a)} = \frac{C^{p,p',t}_{(s,a)} A^{p,p',t}_{(s,a)} + abs(r_{p,t} - r_{p',t'})}{C^{p,p',t}_{(s,a)} + 1} \tag{5.4}$$

for every simultaneous experience of p' at t' in the $[t - w, t]$ range. Here, r_p represents the current reward of user p and $r_{p',t'}$ represents the reward of user p' at the simultaneous experience time t'. An example of this process with a window of $w = 1$ by four users is shown in Figure 5.2. Similar users P3 and P4 quickly

time

P1		$(s_0, a_1)\ r_P$	$(s_1, a_0)\ r_N$	$(s_0, a_1)\ r_P$	$(s_1, a_1)\ r_P$
P2	$(s_0, a_0)\ r_N$	$(s_0, a_1)\ r_P$		$(s_1, a_1)\ r_P$	
P3	$(s_0, a_0)\ r_P$		$(s_0, a_0)\ r_P$	$(s_0, a_0)\ r_P$	$(s_0, a_0)\ r_P$
P4		$(s_0, a_1)\ r_N$	$(s_1, a_0)\ r_P$	$(s_0, a_0)\ r_P$	$(s_0, a_0)\ r_P$
Shared	P2, P3	P1, P2, P4	P1, P4	P3, P4	P3, P4
Different	P2-P3	P1-P4, P1-P2	P1-P4		

Figure 5.2: KindredRL finds similarities between users by noting when users share a state and action and recording when their response to those actions are different. Over time similar users will often converge to the same continued state action pairs (such as P3 and P4) while differentiating themselves from different users (P1 and P2) given their responses during simultaneous interactions.

converge on a similar (s, a) while accumulating difference information between themselves and users P1 and P2.

Once the count and average difference in responses of two users has been calculated, we transform those values into a distance between the two users, $D_{p,p',t}$. This transformation makes the total distance between two users a sum of the reward difference minus the RL convergence similarity. To weight the two values equally, both $A^{p,p',t}$ and $C^{p,p',t}$ are normalized to values between 0 and 1 for each (s, a) pair. Formally, the total distance between two users is

$$D_{p,p',t} = \sum_{(s,a) \in SxA} \frac{A^{p,p',t}_{(s,a)}}{max_{(}A^{p,p',t}_{(s,a)})} - \frac{C^{p,p',t}_{(s,a)}}{max_{(}C^{p,p',t}_{(s,a)})} \tag{5.5}$$

where the value of $D_{p,p',t}$ can fall in the range $[+|SxA|, -|SxA|]$. If two users have never experienced simultaneous (s, a) pair, their distance value will be zero. If two users have experienced many simultaneous (s, a) and always given the same reward, their distance value will be below zero. Note that this means that two users who have never experienced simultaneous (s, a) can be placed in the same neighborhood. Hence, their likelihood of experiencing the simultaneous (s, a) pairs in the future increase as their policies are calculated to be similar in their neighborhood – making it more likely we will discover any dissimilarities as they respond to the policy. This also prevents known dissimilar users from being placed in a neighborhood just because the similar users have not yet experienced simultaneous (s, a) pairs.

Once we have obtained the distance between users, we then group the users with other similar users. Specifically, we use the k nearest neighbor algorithm to find the k neighbors most likely to share preferences and create a neighborhood $N_{p,t}$ for each user. Formally,

$$N_{p,t} = knn(k, D_{p,\bullet,t}) \tag{5.6}$$

Table 5.1: The notation used for KindredRL.

Neighborhood Creation	
$r_{p,t}$	the reward received from person p at time t
$C^{p,p',t}_{(s,a)}$	count at time t of the simultaneous experiences of (s,a) by p and p'
$A^{p,p',t}_{(s,a)}$	the average difference in reward up to time t of the simultaneous experiences of (s,a) by p and p'
$D_{p,p',t}$	the distance between p and p' at time t
k	the size of neighborhood groups
$N_{p,t}$	the set of neighbors of a person p at time t
w	the simultaneous window

Collaborative Preference Estimation	
$horizon$	the length of the estimation period
$\pi_{p,t}$	the policy at time t for person p
Model Based	
$L^{p,t}_{(s,a)}$	the number of times that p experiences an (s,a) pair
$T_{p,t}$	the estimated transition vector for p at time t
$R^{group}_{p,t}$	the collaboratively estimated reward function for p at time t
Q-Learning	
$Q^{group}_{p,t}$	the collaboratively estimated Q function for p at time t

where $D_{p,\bullet,t}$ represents the distance between p and any other user. We select knn as our grouping algorithm to allow easy adjustment to groups of different sizes for individuals and allow a soft clustering of similar users in real-world applications. Hence, k is one of the algorithm parameters that must be selected for implementation of this approach. We discuss the effect of selecting this parameter in Section 5.6.2. While our approach currently uses the same k for all users, we anticipate future versions of the approach can create neighborhoods of varying ks. Once the current neighborhood of a user has been determined, that neighborhood is used to determine the model component $R^{group}_{p,t}$ in the following estimation step.

5.4.3 Collaborative Preference Estimation

The goal of the estimation step is to use the neighborhood groupings $N_{p,t}$ of a user to generate that user's reward model $R^{group}_{p,t}$ or Q value $Q^{group}_{p,t}$ for the model based and Q-learning implementations respectively. These components can then be used to generate the current policy. To calculate these components we treat all the users in $N_{p,t}$ as if they were the single user p. Hence, user p now has many more samples than those that they alone experienced. This allows the estimation algorithm to fill in expected (s,a) rewards of user p with responses given by similar users. Additionally, as preferences change it allows responses from neighboring users to override old, incorrect data in R_p that the current user hasn't yet re-experienced. This allows the algorithm to select the correct action for the state when user p experiences it, rather than needing to try the incorrect action independently to know preferences have changed. However, this estimation step isn't as simple as averaging the neighborhoods' responses to (s,a) since many of the neighbors may not have experienced the new preferences. Hence, they may be retaining older, incorrect data as well. Averaging

$h \leftarrow t - horizon$ //initialize the time in history to begin group estimation
for p = *1,2,...,P* **do**
  Model Based: $R^{group}_{p,t} \leftarrow R^{group}_{p,h-1}$ //initialize historical knowledge
  Q-Learning: $Q^{group}_{p,t} \leftarrow Q^{group}_{p,h-1}$ //initialize historical knowledge
  for $i = h, h+1, ..., t$ **do**
    foreach $p_n \in N_p$ **do**
      Model Based: $R^{group}_{p,t}(s_{p_n,i}, a_{p_n,i}) \leftarrow R^{group}_{p,t}(s_{p_n,i}, a_{p_n,i}) + \alpha(r_{p_n,i} - R^{group}_{p,t}(s_{p_n,i}, a_{p_n,i}))$
      Q-Learning: $Q^{group}_{p,t}(s_{p_n,i}, a_{p_n,i}) \leftarrow$
       $Q^{group}_{p,t}(s_{p_n,i}, a_{p_n,i}) + \alpha(r_{p_n,i} - Q^{group}_{p,t}(s_{p_n,i}, a_{p_n,i}) + \gamma \dot{\max}_a(Q^{group}_{p,t}(s_{p_n,i}, a)))$
    end
    if $i == h$ **then**
      Model Based: $R^{group}_{p,h} \leftarrow R^{group}_{p,t}$ //set historical knowledge given current neighbors
      Q-Learning: $Q^{group}_{p,h} \leftarrow Q^{group}_{p,t}$ //set historical knowledge given current neighbors
    end
  end
end

Algorithm 3: KindredRL learns the reward model R^{group}_{p} or Q value $Q^{group}_{p,t}$ for a person p by learning from all the samples of p's neighborhood as if they were the user's own samples.

would require a large number of neighbors (e.g. up to half) to record a change in preferences before it can be corrected for the current user.

To make sure estimation can quickly adapt to changing preferences we adapt a concept called experience replay, where some historical responses by users are recorded to be later replayed by the RL algorithm for various objectives. In our case, we retain all experiences from the users over an *estimation horizon*, *horizon*, and rebuild our information about them ($R^{group}_{p,t}$ for the model based implementation and $Q^{group}_{p,t}$ for the Q-learning implementation) based on their neighborhood over this horizon.

We replay all the (s, a, r) triples experienced by each user in $N_{p,t}$ over the estimation horizon and apply them to reconstruct $R^{group}_{p,t}$ or $Q^{group}_{p,t}$. We apply each experience in the order they were originally experienced. If neighbor one experienced (s, a) at $t1$, $t2$ and $t4$, and neighbor two experienced (s, a) at $t3$, then the user p would have the (s, a) replay first updated with neighbor one's $t1, t2$, then neighbor two's $t3$, and then neighbor one's $t4$ response. This means that the newest information for the group from any neighbor for each (s, a) is applied last, overriding or improving the estimate of any older data for the neighborhood. Depending on the learning parameter α that is used, this process can quickly change the collaborative value with only a few neighborhood members recording the preference change. The update equation is described formally in Algorithm 3 for each implementation type.

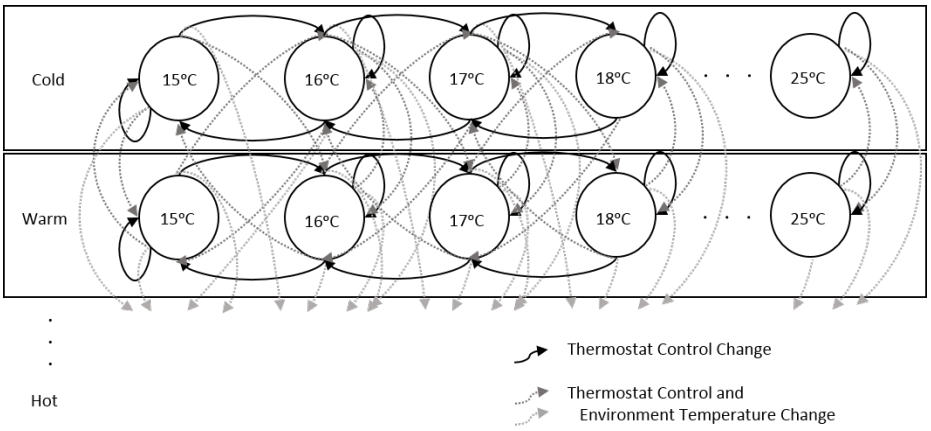
The group estimation of $R^{group}_{p,t}$ or $Q^{group}_{p,t}$ must be continuously replayed at each iteration because the distance measures and subsequent neighbors $N_{p,t}$ are also continuously updated. Because the estimation of $R^{group}_{p,t}$ is replayed over the estimation horizon with the current set of $N_{p,t}$ it effectively uses the future to correct the past. Previous neighborhoods that may have been incorrect for user p only affect values in $R^{group}_{p,t}$

or $Q_{p,t}^{group}$ before the estimation horizon begins, and are often corrected by the current neighbors during replay. To ensure that the replay begins with the most recent neighborhood knowledge, the earliest $R_{p,t}^{group}$ or $Q_{p,t}^{group}$ in the replay is chosen as the next initialization point at time h for $R_{p,h}^{group}$. Hence, there is a lag time the length of the estimation horizon before information based on grouping is set in stone. This allows collaborative learning to learn groupings in a cold start scenario without overly propagating the initial errors in neighborhoods. Once $R_{p,t}^{group}$ or $Q_{p,t}^{group}$ has been learned for the current iteration it and the learned T are fed into value iteration to create the current policy. The RL algorithm then proceeds as normally as shown in Algorithm 2 and continually updates neighborhoods and policy until the process is discontinued.

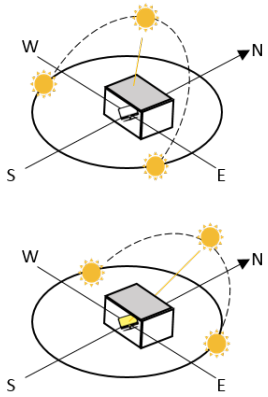
5.5 Experimental Setup

To evaluate if KindredRL can leverage the similarity between individuals to quickly learn preferences in an AmI environment with the small data problem we simulate two common building applications and one uncommon one. Two common applications exhibit both groups of people with similar preferences and changes in those preferences over the course of a year. These applications are: thermostat control for desired temperature and blinds control to prevent glare in individual offices. Both of these simulations use individual offices to directly tie human interactions with the building and RL algorithm to desired preference - shared office spaces with internally conflicting temperature and blinds preferences are not covered in this paper. For both applications we learn preferences by considering a person's interaction with the system to be a form of negative feedback (e.g. moving the blinds up when the system has moved the down indicates that 'down' was a wrong action) and their non-interaction with the system to be a form of positive feedback (e.g. if the system increases the temperature and the user does not indicate the new temperature is uncomfortable, this must be an acceptable temperature). Both applications use real-world external temperature and solar levels from NREL's National Solar Radiation Database for the year 2017 in Washington DC, USA (ID# 1144614) as the basis for the simulations [176]. Both applications are simulated on an hourly basis (i.e. an action can be taken by the RL system every hour) and operate between the hours of 8 A.M. and 5 P.M. Weekends are treated the same as weekdays for our simulations – providing a slightly larger, conservative estimate on total learning samples over a year.

We selected thermostat control and blinds control as our simulation applications because each is one of the most common control problems in office buildings. Additionally, both confront the preferences of individuals in their operation and must adjust to those preferences so the occupants of a building will be comfortable – and won't try to circumvent the system with space heaters, disabling methods, or manual control. Thermostat preferences place people in a variety of groups based on factors like their body size,



(a) The Markov Decision Process for the thermostat application moves between states consisting of the tuple (external temperature, internal temperature) by either increasing, decreasing, or not changing the internal temperature as an action. The external temperature changes independently of the internal temperature control, so actions can move from a state in the cold range to a state in the warm or hotter ranges.



(b) The blinds application calculates the glare in a room based on the azimuth and altitude of the sun over the course of a year. Depending on the direction a window is facing, each person will experience a change in blind preferences at different times over the year.

Figure 5.3: Application Simulations.

gender, chosen fashion, metabolic rate, age, proximity of working location to airflow, and acclimation level to outdoor temperatures [156, 157]. Blinds preferences place people in groups based most often on their physical location, and that of their visitors or monitors, in relation to the sun. Both applications can have the preferences of a group change over time, such as when the external temperature decreases and people begin to wear heavier clothing both inside and outside the office or when the angle of the sun changes throughout the year suddenly producing glare at 8 A.M. in all east facing windows.

In addition to thermostat and blinds control, we select load shifting with a water heater (based on the work on ThermalThrift in Chapter 3) to represent an uncommon building application that deals with preferences. In this case, reinforcement feedback isn't based on the negative reward when a person interacts with the system, but the cost amount of water heating. Hence, we get a higher negative reward for overheating the tank with more energy and a higher cost than maintaining the tank. Additionally, the preferences in this application effects both the reward (e.g. we pay more when maintaining a temp during usage) and the transition function (e.g. we lose temperature fast during higher usage). We describe this process in more detail below. Additionally, details on all the applications (such as the selection of preferences and change times), the various simulation parameters used for sensitivity analysis, baselines for comparison, and evaluation metrics are included below.

5.5.1 Thermostat Application

In our thermostat application, we choose to have 6 groups of similar individuals. Within a group, each person has an identical preferred temperature to their group members. We choose to use identical preferred temperatures, rather than a mix of almost identical temperatures, because at a large scale there will always be groups of people with identical temperature preferences. For example, if 100 people prefer temperatures in a uniform distribution of the comfortable range 20°C to 22°C, then 33 of those people will have identical preferences. This is obviously affected by granularity, where larger temperature buckets will have more people with "identical" preferences. We look at the effect of including people that do not match this preference in the neighborhood in Section 5.6.2. In our simulation, temperature preferences range from 15°C to 25°C.

When an action by the RL system produces a temperature that does not match the preferred temperature an individual can give a negative response to the new temperature though whatever feedback mechanism is provided by the building. Like many commercial building thermostat systems, we do not allow a user's interaction to directly modify the temperature of a room. Users can provide negative feedback, but only the RL system can change the temperature. For simplicity, our simulation allows the system to change the temperature of a room by 1 degree in 1 hour. Each hour the system also samples the current external temperature for that user. We choose four times throughout the year to have the desired temperature preference of each group change corresponding to seasonal changes. The first preference period is considered a "cold start" where we have no prior information about the individual's neighborhoods or preferences. When a preference period ends each individual transitions to the next preference period at a random time over the next two weeks, corresponding to the average time it takes for a person to acclimatize to a new temperature.

Formally, the set of states, S, for the thermostat application consists of the (external temperature, internal temperature) tuple. The set of actions, A, consists of increasing the temperature by 1 degree, decreasing the temperature by one degree, and leaving the temperature the same. The external temperature changes independently of the chosen action. Overall, the thermostat application follows the Markov Decision Process depicted in Figure 5.3a. We simulate a range of internal temperatures from 10°C to 30°C. The internal temperatures of the day begin in the same state as the previous day. External temperatures range from 10°C to 30°C, where temperatures above and below that from the real world data set are truncated to fit in that range. With three possible actions, there are 1323 (s, a) pairs in the thermostat application. The reward, R, for any (s, a) pair is $-|internal\ temperature - external\ temperature| - 10$ if the resulting temperature of the state action pair is not the individual's desired temperature and $-|internal\ temperature - external\ temperature|$ if it is.

5.5.2 Blinds Application

In our blinds application, we select 8 groups of similar individuals based on the cardinal (N, S, E, W) and primary intercardinal (NE, SE, NW, SW) directions their office windows face. While real-world applications might have subgroups within these preference groups, such as different monitor locations, for simplicity we assume each direction has the same blinds preferences and that the dimensions of all rooms are the same. An inclusion of the subgroups would likely require more people to learn preferences, but would still show a similar learning trend given a larger population. All people in our application simulation share the same light level preference, where light entering their room above that level produces glare and light below that level does not. Again, diversifying this parameter would necessitate more people to learn, but is possible. In our simulation, preferences change over the course of the year based on the level of light into that room given the angle of the window, the dimensions of the room, and the altitude and azimuth of the sun as seen in Figure 5.3b. For our application, that means preferences change a total of 41 times over the course of the year and the majority of preferences changed affect only one group at a time.

Formally, the set of states, S, for the blinds application consists of the (hour, external light, blinds location) triple. The set of actions, A, consists of opening, closing, or leaving the blinds unchanged. The external light changes independently of the chosen action. Unlike the thermostat application, users have direct control of the blinds in this simulation. If the RL algorithm moves the blinds up that user can press a button to immediately move the blinds back down. In this case, the subsequent state s' will have the blinds down based on the user action rather than the algorithm action. Overall, this will speed the algorithms progress towards the correct blind state but will prevent less preferential (s, a) pairs from being sampled. Because we limit the simulation to the work hours between 8 A.M. and 5 P.M and discretize the external light levels to 5 values, the total number of (s, a) pairs is 300. The reward, R, for any (s, a) pair is -10 if the resulting blind state is not the user's preference and 0 when it is. Additionally, any action the moves the blinds has an added -5 to the reward to minimize the annoyance of moving blinds.

5.5.3 Water Heating Application

In our water heating application we use a subset of the data collected in Section 3.4.1. For each of the 6 pairs of participants we randomly select 2 days of usage data and permute and repeat them over a period of two weeks to represent users that are similar in their usage patterns, but not exactly the same. We do not exhibit changing preferences in this application, instead focusing on the variations in usage for an individual and its simulated group. This allows us to show the use of KindredRL to increase the exploration of the state space regardless of changing preferences.

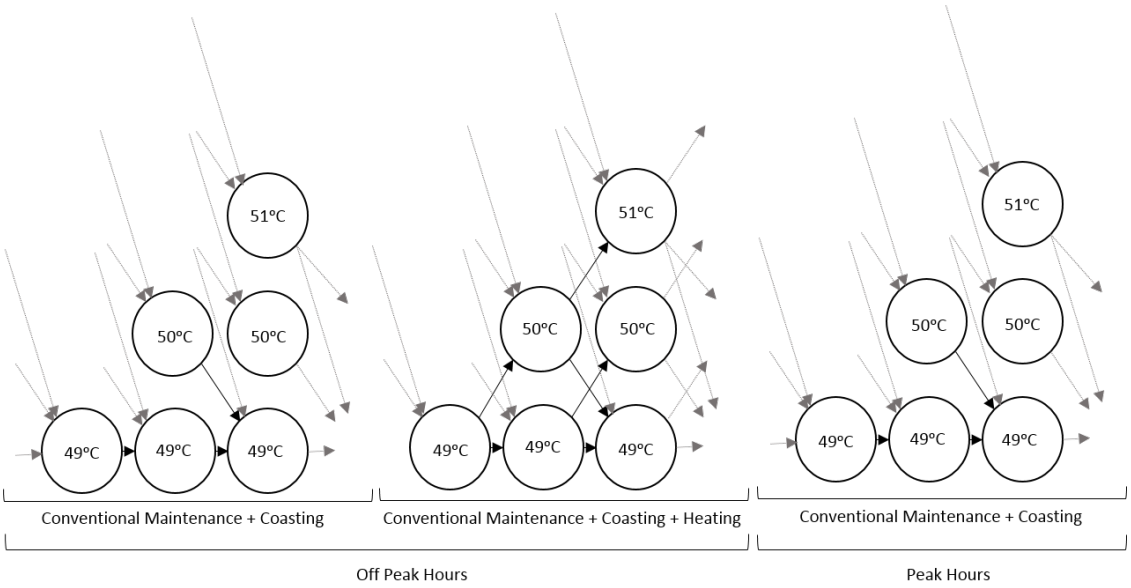


Figure 5.4: The water heating application allows heating to occur for 4 hours before the peak period. Only coasting (allowing the temperature to fall due to thermal loss or usage) is allowed in the peak and early off-peak periods. All periods require the maintenance action to be taken at 49˚C during all periods to maintain hot water in the tank.

The water heating application works by potentially selecting the action to heat during the end of the off-peak period and allowing the temperature to fall during the on-peak period. The reward for any action is the cost of any heating following that action. The price of electricity changes according to one of the TOU schedules (P2) in Section 3.4.2. This cost includes extra heating to raise the temperature and maintenance heating to maintain a 49˚C tank from either usage or thermal loss. If KindredRL does not perform heating and allows the temperature to fall, then change from s to s' represents the temperature change in the tank due to either thermal loss or usage. A diagram of the application actions as they relate to the time is shown in Figure 5.4. Formally, the set of states S for the water heating application consists of the (hour,tank temperature) tuple. The set of actions, A, consist of heating, maintaining a 49˚C temperature, or letting the temperature fall in a coasting action. Not all actions are applicable to all states. The heating action can only occur for the 4 hours before the peak period and the maintenance action is only available when the tank temperature is 49˚C. Each action is run on an hourly basis. We use a few simplifying assumptions in the application, including that any usage can be recovered from, heat wise, within one hour and that the thermal loss stays the same over a one hour period.

Unlike the blinds and thermostat applications, the personalization portion of the water heater application can affect the transition function T. This means that a usage of 5˚C of water can change the state from (6 am, 70˚C) to (7 am, 64˚C) for one person while another without that usage has a transition to (7 am, 69˚C)

assuming 1°C of thermal loss. Hence, it's not just the reward function that is affected by a person's use, but the transition function as well. Because the water heater application looks at the long term effect of actions (the benefit of storing heat may come many hours later) the transition function is an important piece of estimating the value of actions. Hence, for the water heater application we modify collaborative estimation slightly to make $T_{p,t}$ the average of a user's neighbors $T_{p_n,t} \forall p_n \in P$. This step was not necessary for the blinds or thermostat applications since their actions didn't have a far reaching effect on future actions, but was necessary to the water heater application. Any rewards recorded by neighbors needed to be combined with potential transitions to all a user to estimate values given future rewards.

5.5.4 Approach Parameters

For our evaluation, we select the approach parameter values in accordance with the application. These parameters include α, ϵ, and γ from the learning framework, w and k from neighborhood creation, and the estimation *horizon* from collaborative estimation. Any application that uses KindredRL must select these parameters for operation. For any analysis that is not checking the sensitivity of a particular parameter (we specifically look at k, the *horizon* and α), the values of the approach parameters match those discussed here.

The size of the neighborhood k is chosen to match the ground truth size of the number of people in a group for each application. The estimation *horizon* is 280 iterations or about 2 weeks of the year. The exploration parameter ϵ is 0.05 for the thermostat and water heating applications and 0.01 for the blinds application. We chose a slightly higher exploration rate for the thermostat and water heating applications because of their larger and more complex state space. We select γ as 0.5 for the blinds and thermostat applications and 0.99 for the water heating application. The water heating application's γ is much higher since the selection of an action depends on the result of that action many hours later towards the end of the peak period. The window for detecting simultaneous experiences for grouping, w, is 1. This means that we conservatively only check the similarity of users that experience a (s, a) pair at the exact same time. The learning parameter α should be a different value depending on various factors of the application (we discuss this extensively in Section 5.6.1). Hence, for all approaches we do a parameter sweep of alpha and select the best result to present.

5.5.5 Simulation Parameters

Alongside the algorithm parameters, we use a variety of simulation parameters to perform a sensitivity analysis with the blinds and thermostat applications. These parameters include: *occupancy*, *response rate*, *staggered preference change period*, and *preference length*. *Occupancy* is used to denote whether or not a

user's lack of negative feedback should be attributed to positive feedback. Essentially, we assume we have some form of occupancy detection available to make sure the system doesn't learn a particular action is acceptable just because someone is not in the room to respond negatively. If the room is not occupied, the simulation does not perform or learn from any action. We use the separate *response rate* to define when a person is present but chooses not to respond to an action, either due to general data noise, preference tolerance/noise, or laziness. Both occupancy and response rate are represented as an overall percentage in the simulation (e.g. users respond 60% of the time) but each user is present or responds individually given that percentage. The *staggered preference change period* represents the potential for noise at the beginning of a group's preference change – such as the 2 week acclimatization period in the thermostat application. The staggering period defines when the changes across the groups occur and each individual changes preferences at a randomly selected point in that period. Finally, the *preference length* represents the length of a period where a group's preferences remain unchanged. The smaller this length is, the less time an RL algorithm has to learn and exploit a policy.

For both blinds and thermostat applications we use an occupancy percentage of 60% for the main result, corresponding to the amount of time middle and upper management workers spend potentially out of the office in meetings [177]. We set response rate to a low 60% for the thermostat application, as people often have a tolerance for small changes in temperature and little interest in responding every hour to report temperature comfort. We set the response rate to 90% for the blinds application, as a person's response to glare is often immediate. The 10% lack of response in this application is to cover any noise in tasks that would affect tolerance for glare, such as sometimes reading a printed document when the person usually reads on their monitor. We set the staggering preference change period and preference length for the thermostat application alone, as these parameters are calculated from the location of the sun for the blinds simulation. For the thermostat simulation, the staggered preference change period is set to 2 weeks and the preference length is set to 3 months. A sensitivity analysis of all of these parameters is presented in the results.

5.5.6 Baselines

We present two variations of our approach in the results: the model based KindredRL and the Q-learning based Q-KindredRL. We compare our approach against four main baselines: traditional Q learning, Q-learning with forgetting, an independent model-based approach using value iteration, and an independent model-based approach using value iteration and forgetting. Traditional *Q learning* is one of the most widely used variations of RL and is designed to handle slow changes in the model, including with preferences, over time. Q learning and its variations are commonly used since they required no model of the system in order to learn a policy.

Q-learning w/Forgetting uses ground truth knowledge of the preference change time to forget all learned knowledge of the model. This represents the ideal case of any change or context detection RL algorithm where there is no delay in detecting the change. Additionally, it throws out any old data that no longer matches the new preferences preventing any incorrect data from propagating through the new policy. The third baseline, called *Independent*, learns the model components, R and T, in the same way our model-based approach KindredRL does, but without any form of collaboration (i.e. $N_p = p$). Each user has an independent RL agent that learns their model and derives a policy with value iteration. The final baseline performs the same as Independent with the addition of forgetting. Lin Q-learning w/Forgetting, *Independent w/Forgetting* throws out any old data that no longer matches the new preferences preventing any incorrect data from propagating through the new policy.

In addition to the baselines, we also compare against an upper bound for our collaborative approach. We call this bound *KindredRL Bound* or *Q-KindredRL Bound* for the model-based and Q-learning variants respectively. Simply put, the bounds use the ground truth neighborhoods as the output of neighborhood creation. Here, N_p is always 100% according to the simulated data. Hence, these bounds represent the accuracy we can achieve in KindredRL if the calculation and creation of neighborhoods is perfect.

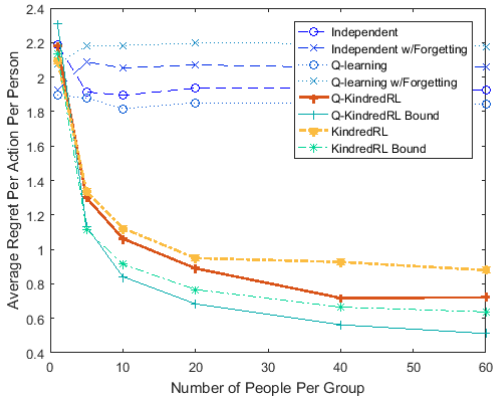
5.5.7 Metrics

To evaluate the success of our approach, we look at one main metric: regret. *Regret* represents the difference between the reward obtained by the optimal policy and the learned policy over the course of the system's operation. Remember that the goal of any RL algorithm is to maximize the sum of rewards over time. However, depending on the values of the reward for any given action even the optimal policy's reward could be less than zero. Hence, we present regret to show how close a policy is to optimal. Particularly, we calculate the average regret per action the algorithm performs per person in the simulation. This allows us to compare regret across our various sensitivity analyses. Formally, the average regret per action per person is:

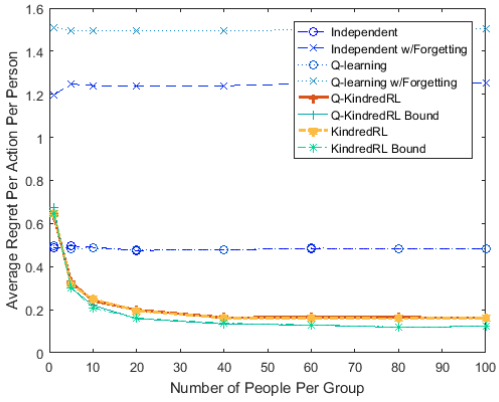
$$(\sum_{p=1}^{P}\sum_{t=1}^{T}(|R_{p,t}^{opt} - R_{p,t}|)/I_p)/P \tag{5.7}$$

where P is the number of people, T is the total number of iterations performed by the system, R is the PxT matrix of reward, R^{opt} is the PxT matrix of optimal reward, and I is the P length vector of the number of interactions between the system and the person though policy actions.

To calculate the optimal policy we run value iteration on the ground truth model for each application. Since the model changes with each preference change, we perform this optimal policy calculation for each period of static preferences. This set of piecewise optimal policies then becomes the full optimal policy



(a) Thermostat application.



(b) Blinds application.

Figure 5.5: The regret (the difference between the reward from our policy and the optimal one) decreases as the number of people in a group increases for the thermostat (left) and blinds (right) applications. At 100 people KindredRL reduces regret by 60% and 67% from the next best baseline for the thermostat and blinds applications respectively.

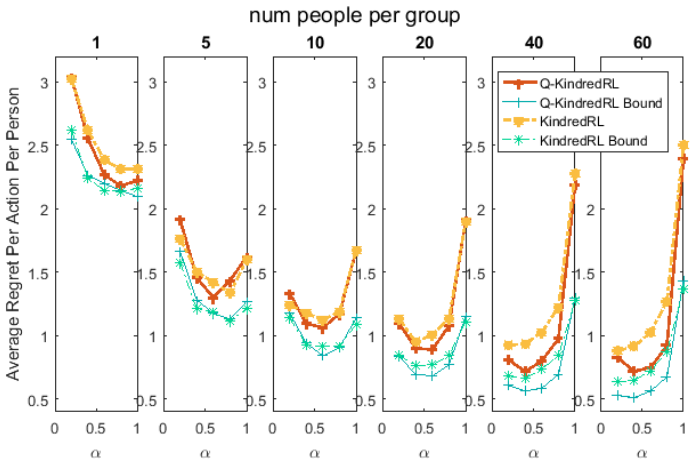
for the simulation. The optimal policy is then applied to the application in the same way using the same exploration rate as KindredRL and the other baselines to focus the difference on the effect of learning and not the exploration rate that all baselines share.

We do not use regret for the water heating application. Instead, we choose to represent the savings in monetary cost over the running of a conventional water heater that always maintains a 49°C temperature in the tank.

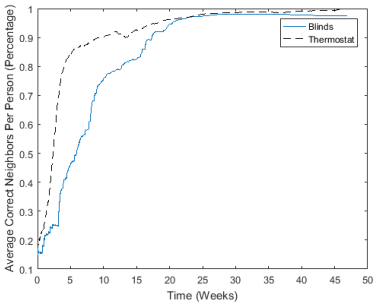
5.6 Results

Our results show that both collaborative variants, KindredRL and Q-KindredRL, can outperform independent approaches in physical environments where preferences change over time. In our two simulated application, blinds and thermostat, the approach begins to improve regret with as few as 5 similar people in a group. Each application has a different relative regret for each baseline, with forgetting severely increasing regret in both applications. The Q-learning baseline is slightly better in the thermostat application and about the same in the blinds application. As shown in Figure 5.5, 80 and 100 people per group decreases regret over the *best* independent baseline by at least 60% the thermostat application and 67% in the blinds application respectively. Overall, leveraging people that have similar preferences can diminish the consequences of changing preferences and subsequently lower regret.

With perfect neighborhood creation, this reduction in regret can reach as much as 73% or 75% for the thermostat and blinds applications respectively as shown by the Q-KindredRL bound in Figure 5.5.



(a) Because the thermostat application has noise in the form of a lower response rate, increases in the number of people in a group also increase the noise for learning preferences in a neighborhood. Hence, lower values of the learning parameter α become necessary as the group size increases.



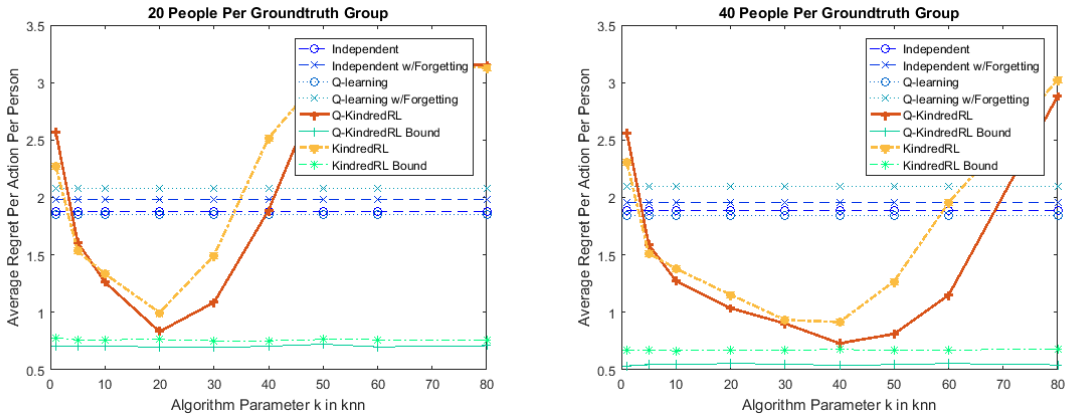
(b) Neighborhood accuracy grows with number of iterations performed, particularly as users have different responses to the same actions.

Figure 5.6

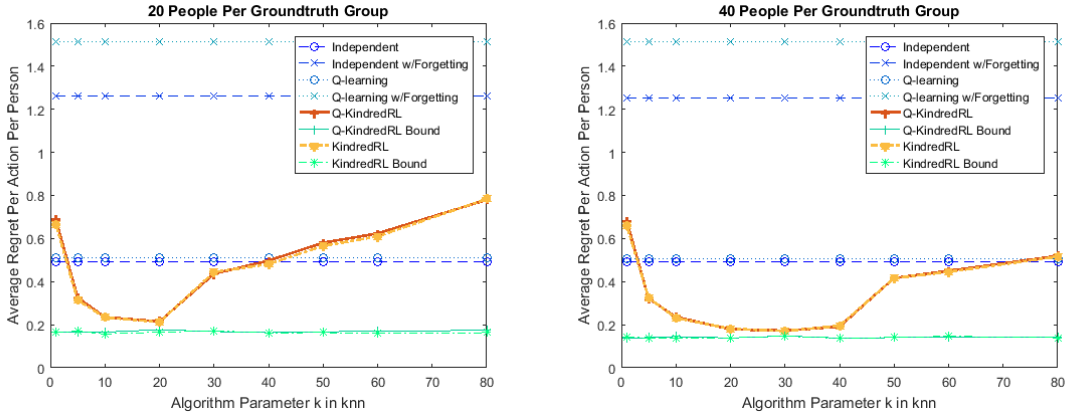
Additionally, the majority of the difference between the variants of KindredRL and their bounds is due to the initial learning process for neighborhoods – the longer KindredRL runs, the longer it can leverage learned neighborhoods. Neighbors for both applications were learned quickly over the year of simulation as shown in an example of learning for neighborhoods of 40 people in Figure 5.6b. Both applications had a 90% average accuracy in neighborhoods within 20 weeks of operation. The thermostat application was able to reach this accuracy before any preference change occurred and the blinds application experienced 19 preference changes across the different neighborhoods during this period. The following sections discuss the effects of the various approach and simulation parameters on the learning capabilities of KindredRL. All parameters below are the same as those used for Figure 5.5, except the sensitivity parameter. As the blinds and thermostat applications show little differences between the behaviors of KindredRL and Q-KindredRL, we will refer just to the model-based KindredRL in their analysis sections.

5.6.1 Selection of Learning Rate

One of the most important approach parameters to select when applying KindredRL to an application is the learning rate α. This parameter defines how quickly the RL algorithm learns from new samples when used in tradition independent approaches. In general, the more noise that is present in the application, the lower the learning rate should be to prevent erroneous data from having a large effect. In KindredRL, this parameter learns not only from the individual's samples but also from those in their neighborhood. Thus the selection of α depends not only on the general noise of the application but also on the number of people



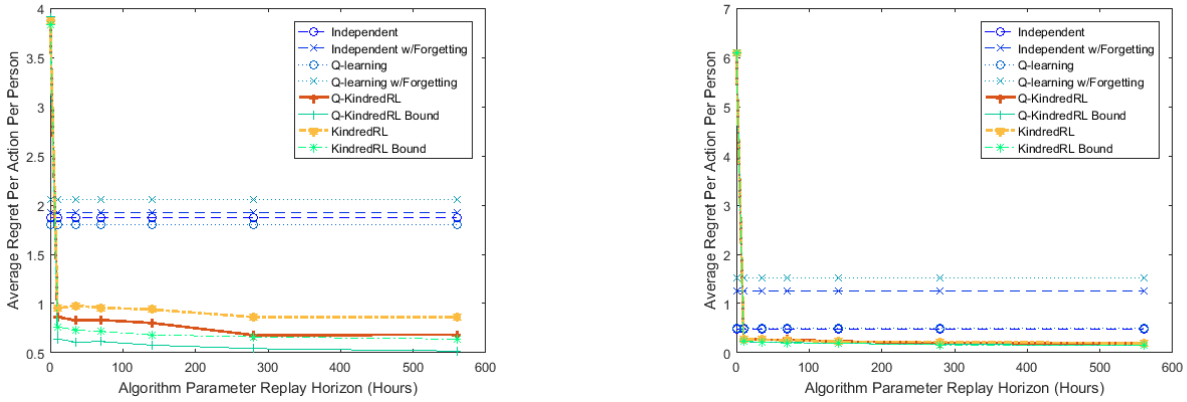
(a) Thermostat application.



(b) Blinds application.

Figure 5.7: Values of k smaller than the ground truth number of people in a group still produces a decrease in regret. Additionally, the inclusion of some members from a different group with a too large k only increases regret slightly in the thermostat application as KindredRL selects similar out-group members. However, too large a k will increase regret above an independent algorithm as seen in the blinds application.

k that are grouped into a neighborhood. Figure 5.6a shows the effect of increasing k corresponding to the number of people in a group for various values of α. These results are shown for the thermostat application and have a 60% response rate – the application based equivalent for noise in reward. Here, larger values of α increase regret as the number of people in a group increases. This same increase is seen in the KindredRL bound, showing that while imperfect neighborhoods exacerbate the issue the problem exists regardless of how accurate groups are. However, when the response rate is increased to 100% (e.g. no noise in the reward data) this behavior essentially disappears. For example, the blinds application, which has a response rate of 90% does not show this behavior at all in the KindredRL bound and only shows a slight (¡0.1 per action per person) increase in regret at 100 people per group. Hence, as with independent RL approaches, the selection of α will be application dependent for KindredRL.



(a) Thermostat application.

(b) Blinds application.

Figure 5.8: For both applications, an estimation *horizon* that is too small can severely impact the performance of KindredRL. However, for both applications an estimation *horizon* as small as 100 iterations, or 1.5 weeks, maintains the good performance of a longer horizon.

5.6.2 Selection of Neighborhood Parameter

The selection of the number of people in a neighborhood k for KindredRL is dependent on the people present in an application. Overestimating this number too much can increase the regret created by KindredRL as it applies samples from non-similar people to a user based on this overlarge neighborhood. However, as seen in Figure 5.7, a small increase in this number beyond perfect neighborhoods can still let KindredRL leverage similarity to reduce regret beyond independent approaches. This shows that neighborhoods do not need to be exactly the same in their preferences, or resulting policies, to gain benefit from a collaborative approach. However, if the dissimilarity between neighbors is too large, then the policy suffers. As you can see in Figures 5.7 doubling the group size for either approach is the largest a group can be overestimated before accuracy suffers. Hence, a few dissimilar group members do not remove the benefit of KindredRL. Since we are using knn as our grouping mechanism, rather than a hard clustering method, this means that users placed incorrectly into another user's neighborhood will still have a neighborhood of their own made up of mostly similar people. Hence, slightly overestimating k, or alternately including a small relative number of dissimilar people, is something that KindredRL can tolerate. Additionally, setting k smaller than the group size in the blinds application produces very minimal increases in regret down to a k of 5. This implies that application designers can select k conservatively and still gain a benefit from the collaborative approach of KindredRL.

5.6.3 Selection of Estimation Horizon

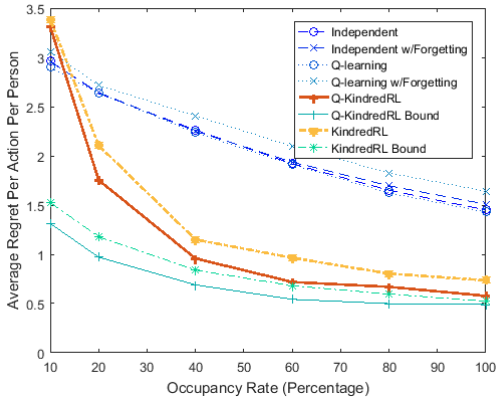
The *estimation horizon* defines how both how long KindredRL replays data for estimation and how many responses must be stored by KindredRL. The smaller the estimation horizon is, the less storage and time

KindredRL uses to perform its calculations. However, the shorter the estimation horizon the more likely errors in neighborhoods are likely to be recorded as fact and propagated as learning moves forward as KindredRL has less time to adjust neighborhoods. Hence, there is an inherent trade-off in the selection of the estimation horizon. However, for our two applications, the length of the estimation horizon can be quite small without severely affecting the overall regret. As shown in Figure 5.8, both applications can have an estimation horizon as small as 10 hours with little effect on regret. This exact value may depend on the application. However, this shows that while the selection of the estimation horizon must be large enough to allow a lag time in neighborhood creation, it can be chosen to limit storage or performance issues.

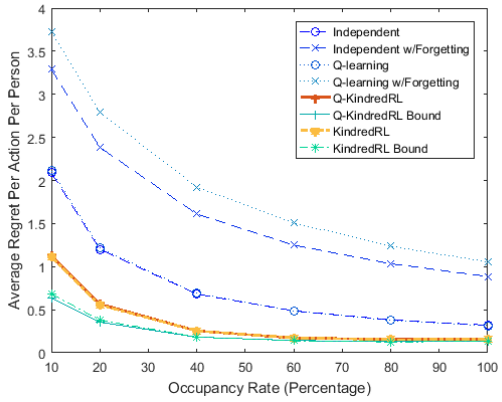
5.6.4 Effect of Occupancy and Response Rate

The underlying behaviors of an application can have a profound effect on the learning capabilities of an RL algorithm. For our applications, we selected two application parameters common to physical environments to discuss for KindredRL. The first, occupancy, represents the amount of time a person is available to be measured over the course for an RL algorithm's application. In general, it shortens the amount of time an algorithm has to learn before a preference change when occupancy is low. With 100% occupancy, and the subsequent higher sample frequency, the Q-learning and Independent baselines are able to reduce the relative difference in regret from KindredRL in both the thermostat and blinds applications. For all the approaches, including KindredRL, a decrease in occupancy means an ever-growing increase in regret. While the KindredRL approach is generally able to slow this increase, for the thermostat application, Figure 5.9a shows that at 10% occupancy KindredRL has a higher regret than the independent baselines. This is in part because with so little occupancy KindredRL is unable to effectively learn groups. With perfect grouping, the KindredRL bound is able to maintain low regret even at low occupancy rates. This implies that, given enough time to at least partially learn neighborhoods, KindredRL could handle short periods of low occupancy. Similar behavior is seen for the blinds application in Figure 5.9b, though KindredRL maintains lower regret than the best baseline in this application.

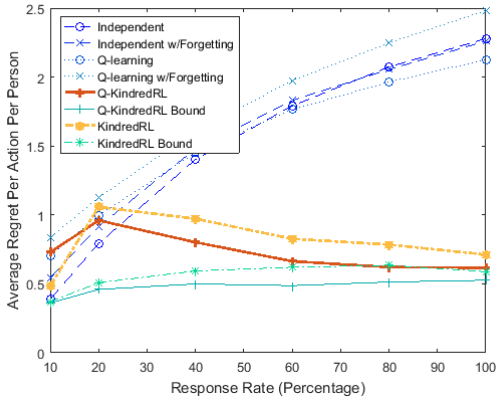
Response rate represents how frequently a person is both present for learning and actively responding to the system. Regardless of whether a person's true response would be positive, if a person doesn't respond due to the response rate their action is assumed to be negative. This represents a type of noise typical in physical environments that tends to skew the results towards higher reward. Figure 5.9c shows that this is particularly true for Q-learning, the independent baselines, and the 10% response rate for KindredRL. As response rate goes down, so does regret as users fail to respond negatively even when actions do not match their preferences. For the thermostat application, regret increases slightly in KindredRL until 20% as



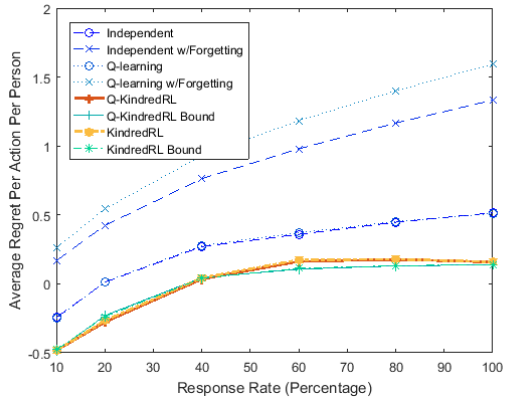
(a) Thermostat application occupancy.



(b) Blinds application occupancy.



(c) Thermostat application response.



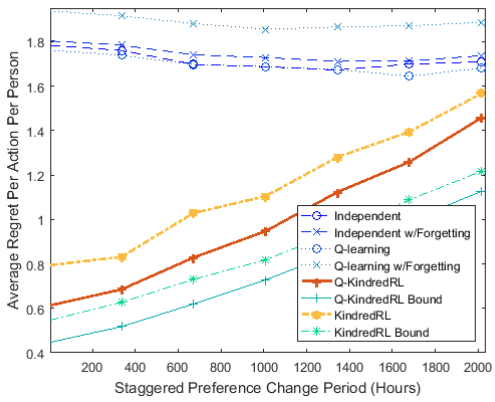
(d) Blinds application response.

Figure 5.9: Both the occupancy rate and response rate of individuals can affect the performance of KindredRL. Lower occupancy rates decrease the overall number of samples available to all approaches, increasing regret. Lower response rates decrease regret overall by getting little negative feedback from users as they don't respond, but KindredRL maintains its performance gains over a wide range of response rates.

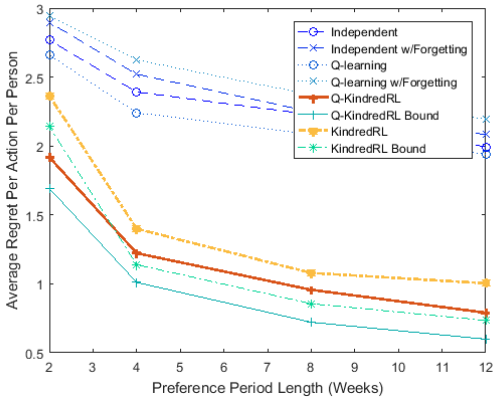
neighbor creation becomes harder with users responding differently to the same (s, a) pair due to the response rate even when their preferences are similar. Figure 5.9c shows little change in the KindredRL bound with response rate as perfect neighborhoods are not affected by this problem. This analysis of response rate also shows that applications where users given different responses (either due to response rate difference or slight differences in hour to hour preferences) are still benefited by KindredRL.

5.6.5 Effect of Staggered Preference Change Period and Preference Length

The length of time before preferences change and how distributed or staggered when the change of each person in a neighborhood is, can also affect the accuracy of KindredRL. As we only control these parameters for the thermostat application, only the results of changing these parameters for the thermostat application are shown in Figure 5.10. Shorter lengths of time before preferences changes tend to increase regret in all



(a) Staggered Preference Change Period.



(b) Preference Length

Figure 5.10: With control over when preference changes occur in the thermostat application, we can see that both shorter stable preference periods and longer periods of active change produce higher levels of regret in KindredRL. Shorter stable preference periods also increase regret in the baseline RL algorithms.

approaches. Even KindredRL sees an increase in regret, particularly as the preference period gets closer to the length of the staggered change period. At a 2 week preference period users are constantly changing preferences at random and independently of their neighborhood. However, even in this case KindredRL outperforms the best baseline (Figure 5.10b).

Changes in the staggered preference change period have essentially no effect on the independent baselines. This is because each user is treated independently here, including for forgetting. Essentially, Independent w/Forgetting only discards old data when the current user changes preferences and does so independently of any other user in the group. However, both KindredRL and it's bound increase regret as the staggered period gets larger as shown in Figure 5.10a. This is because for that period of time samples learned from the neighborhood can in fact not be the same preferences as the user. This is true even with the perfect groups of the KindredRL bound. While it takes a period of over 24 weeks before this issue causes KindredRL's regret to increase to the independent baseline, it does show that KindredRL relies on the common change in a neighborhoods' preferences to occur around the same time. This will be important to keep in mind when deciding to use KindredRL for an application.

5.6.6 Water Heating

The water heating application shows a distinct difference in the results of KindredRL and Q-KindredRL. As you can see from Figure 5.11, KindredRL and the Independent baseline are able to save more than 30% of the cost of a conventional water heater, while Q-learning and Q-KindredRL save less than 20%. This is because, unlike the thermostat and blinds applications, the total reward in the water heating application

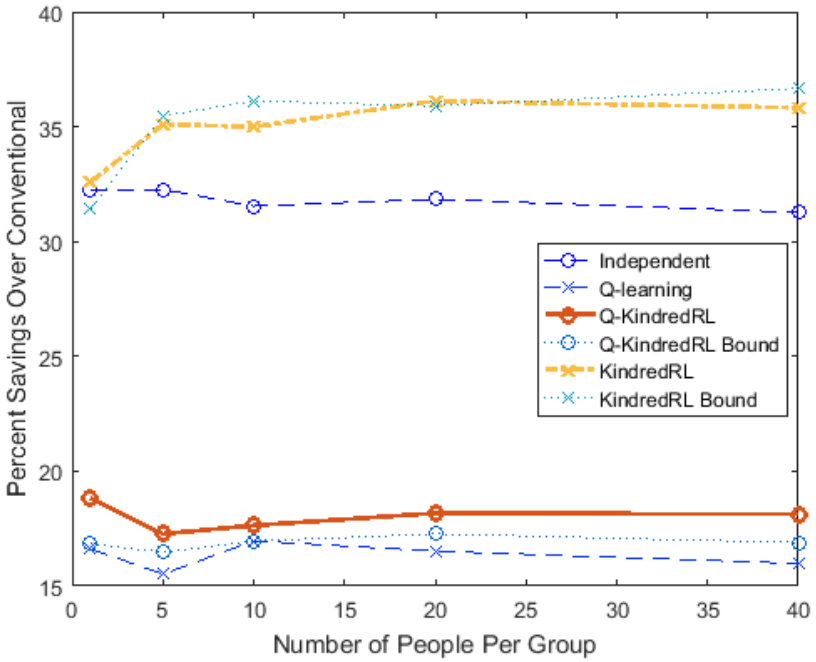


Figure 5.11: The water heating application allows heating to occur for 4 hours before the peak period. Only coasting (allowing the temperature to fall due to thermal loss or usage) is allowed in the peak and early off-peak periods. All periods require the maintenance action to be taken at 49˚C during all periods to maintain hot water in the tank.

depends on the effect of current actions on the long term. Heating might cost much more than maintaining during off-peak hours, but the coasting that is the result of heating might save more heating and cost during expensive peak hours. However, Q-learning must slowly back propagate this information through its update procedure by sampling backward from these savings multiple times. Since time is moving the sampling forward through normal operation, this sampling takes a significant amount of time and the Q-learning variants were not able to learn when to heat in the 2 week period of this test. In fact, they generally learned not to heat, as the full expense of maintaining through peak hours also takes a long time to propagate back to when heating decisions are being made. Savings occurred as a byproduct of exploring heating tracks that are better than conventional in reality, but were not repeated since long term savings were not propagated. It is likely that Q-KindredRL does better than its Bound for that reason, since its incorrect neighborhoods do more exploring and inadvertently gain more reward.

Despite the problems with Q-KindredRL, KindredRL shows that our approach still works in this situation. By using a model-based approach, KindredRL bypasses the propagation issues and is able to learn when to heat from the interactions with its neighbors. This gives KindredRL an additional 4% savings compared to the Independent approach for a total of 35% monetary savings over a conventional water heater. This

number includes the time spent learning and trying less rewarding heating actions. Hence, KindredRL can still exhibit a benefit in learning on more complex applications aimed at meeting societal needs.

5.7 Limitations and Future Work

While KindredRL shows the potential of using people with preference similarity to learn quickly in environments with changing preferences, there is still much work to be done. The main limitation of the above evaluation is that it is run on simulated people and applications. Though the simulated applications are based on real-world data and environment models, they are not a full substitute for an in-situ evaluation. Future work would look to apply KindredRL to a building and evaluate if and when the system breaks down. Additionally, identifying other applications with the preference behaviors needed for KindredRL is a direction of future work. Currently, KindredRL requires an application with users that share similarities in preferences in various groups and change those preferences as a group around the same time. Finding other applications that fit those assumptions and evaluating KindredRL on them is a large area of future work.

In addition to further evaluation, there are a number of things to explore with the KindredRL algorithm. At the moment, KindredRL requires the manual selection of various parameters related to its operation, such as k and α. One direction for future work is to look at automatic processes to select and adjust these parameters, potentially per person, to decrease the burden on algorithm designers when they use KindredRL. This is particularly important for the neighborhood sizes k as they play are large role in the accuracy of KindredRL. Additionally, further work looking at similarity in sub-preferences, such as two people sharing preferences in winter months but not summer months, may be a direction of future work. While KindredRL's current design easily allows people to leave a neighborhood, joining one that a user was previously dissimilar to is much harder as similarity considers the user's full history. Techniques to phase out old similarity information and allow neighborhood merges might also be a direction for future work.

In addition work on the KindredRL algorithm itself, future work may include integrating this technique with other advancements in RL. Q-KindredRL would likely perform just as well as KindredRL in the water heating application if combined with an experience replay technique that could more quickly propagate long term reward information. Exploring how well KindredRL integrates into an agent-collaborative environment or handles function approximation to reduce state space are both potential future work. Integrating KindredRL with a deep RL algorithm, where neighborhood estimation will remain unchanged but collaborative preference estimation must be adapted to a new model, would be an exciting direction for future work.

5.8 Conclusions

In this work, we present KindredRL: a reinforcement learning algorithm that leverages similarity between users to quickly learn changing preferences in a physical AmI environment. In our two common application simulations, we are able to reduce regret (the difference between the actions of an optimal policy and the learning one) by 60% and 67% over the next best baseline for thermostat and blind control respectively. For our water heating application, we were able to save 35% of the monetary cost of a conventional water heater, 4% more than the independent based RL approach. These results show that KindredRL can quickly learn preferences and respond accordingly even when preferences are changing over time.

KindredRL demonstrates that leveraging similarity between users is a promising direction for learning preferences with reinforcement learning when faced with infrequent interactions in an AmI environment. By leveraging the similarity between users KindredRL can more quickly explore the state space of states and their potential actions in an AmI system and use the exploration of a user's behavioral kin to speed learning. We anticipate that this approach could help bring personalized control based on natural human interactions to AmI systems being developed in physical environments today. With the lower participation requirement of natural interactions, and the faster learning of KindredRL, this technique could provide a learning method in lower personal value AmI systems designed to meet societal needs.

Chapter 6

Conclusion

6.1 Summary

In this work, we show that including human preferences to create AmI systems targeting societal concerns, such as shifting the energy peak, can greatly increase the individual benefit of the system to the users and therefore the potential for adoption. We show in the load sifting work called ThermalThrift, that including individual preferences and usage patterns can be used to shift 47% of the peak load with saving consumers 25% of the cost. An approach that does not use personalization can shift more load, but costs three of our evaluated consumer 11-12% more than a conventional heater. This approach of using average behaviors or not considering behavior at all can greatly reduce the adoption of systems meeting societal needs as users decide not to pay the monetary or frustration cost of non-personalized systems. Because societal concerns are often of low personal value, making these systems no more effort than living normally is necessary for their success. This means these systems must meet the AmI vision: meeting objectives with highly adaptive, anticipatory, personalized, and contextually aware systems that rely only on implicit interaction that the user would practice naturally. However, because such systems rely on understanding the people in an environment, AmI has two major challenges: privacy and keeping people in the loop of learning.

In Lethe, we present a novel approach to privacy of restricting hardware capabilities in a camera to preserve image privacy by severely limiting the data that can be obtained by hacking. These limitations prevent a person from removing ever removing the full image recorded by the camera, but also severely limit our processing power and memory. To show that AmI sensing can still be performed, we design Lethe to detect and identify people as they walk through a doorway–a common AmI sensing application. Even with limited hardware, we are able to track three individuals by height with 94.3% accuracy and only a 6 cm

height differential. With the limited hardware, only 1.32% of an image could even be extracted even if all security measures failed. We anticipate that this privacy preserving technique could be added to the portfolio of privacy solutions and help increase the use of one of the most information-rich sensors: cameras.

In KindredRL we keep humans in the loop by leveraging the similarity in user preferences to speed learning when only natural interactions are used to obtain preference information. We base this work on the machine learning paradigm best designed for natural interaction, reinforcement learning. Because natural interactions are often infrequent, particularly given a specific context, we show that using the interactions of similar user's to update one user's information can increase the number and diversity of interaction samples and speed learning. In particular, this can help learn new preferences as a group of similar people's preferences change over time. We show that we can reduce regret in RL by up to 67% over the next best independent RL approach. With this increase in accuracy, and the lower participant requirement of natural interactions, we anticipate that this technique could provide a learning method that could increase the adoption of lower personal value AmI systems that meet societal needs.

While the work in this dissertation was designed to target meeting societal needs, the approaches presented could easily scale to AmI systems with other objectives. Reducing the concerns of privacy and the load on the user of participation are necessary for lower personal value applications, but could also improve the adoption and general satisfaction of other higher personal value applications. For example, an image privacy preserving camera for tracking could easily be used in an ambient assisted living situation and a similarity based reinforcement learning process could be used to more effectively learn personal systems such as clothing recommendation. While we hope that our contributions will promote the use of AmI system to meet societal needs, we anticipate that they will be just as effective at meeting personal ones.

6.2 Limitations

There are a variety of limitations present in the three works of this dissertation:

- ThermalThrift was evaluated only on one house with a rotation of participants. We do not know if these results extend across a wider variety of buildings and houses.

- ThermalThirft was evaluated in simulation where we were unable to evaluate if the existence of more hot water present in the tank would change user behavior, like taking longer showers.

- We did not evaluate the effect of or mitigation of a possible secondary "pre-peak" of hot water heaters storing energy in preparation for peak hours at large scale.

- Lethe was evaluated in only one doorway and therefore one thermal background. Other thermal backgrounds may present unanticipated challenges for Lethe.

- Lethe was evaluated in a controlled setting where we did not evaluate or mitigate the effect of hot object sitting in the field of view.

- Like all height based tracking system, Lethe is limited to accurate identification in small groups of people with variable heights. While this is common in a residential setting, Lethe is unlike to be accuracy in a commercial one. Additionally, a person's height might legitimately change over the course of a day as they change posture or footwear.

- The difference between the background temperature and the top of a participant's head is critical to Lethe. When background temperatures are too warm or head coverings are too insulating, Lethe is unable to accurately measure heights.

- Hardware limiting privacy preservation is necessarily limiting in the applications it can support. Applications that require extensive process or the full image to evaluate all at once likely will not be able to use Lethe's approach.

- KindredRL was mainly evaluated on simulated data. The applications themselves were simplified to focus on learning the preferences of the users. We do not evaluate the effect of meeting a non-personalization objective, such as energy savings, in the blinds and thermostat applications.

- KindredRL was evaluated on only a subset of the water heater data. This was mainly due to time constraints and the evaluation should be further expanded.

- KindredRL leverages similarity in such a way that it works best when a user's preferred policies are identical to the majority of their neighbors. Their underlying preferences and responses can be different, such as preferring different temperatures within the system granularity or using different amounts of water during peak hours, but the system learns best when the preferred policies are the same.

- KindredRL currently does not automatically determine its application parameters, particularly the neighborhood size k. Additionally, the neighborhood size for all users is currently the same despite the fact that real users may have different size neighborhoods.

6.3 Future Research Directions

This dissertation does not solve all problems in AmI, nor does it present a complete and total solution to the two AmI challenges. Instead, it aims to provide novel approaches to be included in the portfolio of solutions needed to bring AmI to wide enough adoption rates to begin impacting societal needs. Future work might look at improving each of the ideas independently, such as addressing the limitations presented above. Particularly, study each of these works in an in-situ environment could further our understanding of the accuracy, benefits, and drawbacks of each system. Additionally, it could provide the opportunity to find and deal with corner cases – such as having a hot background object in Lethe's field of view. ThermalThrift could be evaluated on a wider variety of buildings, Lethe could be evaluated on more diverse doorways, and KindredRL could be run on real building systems and evaluated for accuracy and user satisfaction.

Further technical research in each area is also a direction for future work. ThermalThrift could look at incorporating objects to level the pre-peak cause by storage on a neighborhood scale or perform internal objectives like using the heat changing capabilities to kill bacteria. Additionally, further energy savings might be obtained by allowing the tank temperature to fall below a conventional tank and using the learned preferences to minimize or avoid any comfort loss in that scenario. ThermalThrift could also be extended to dealing with renewable energy production by allowing heat storage when energy is produced in excess. This might even be performed on a neighborhood scale to direct storage of this excess energy into homes most likely to use it quickly after storage or during peak hours.

Future work for Lethe includes applying this technique to other applications. While we perform tracking in doorways, other applications might try to perform tracking with rooms, or identify users as they interact with specific objects. Dealing with thermal background noise and objects may also be valuable and could introduce a technical trade-off between privacy and accuracy or robustness. Testing with higher resolution and frame rate cameras may improve accuracy. For Lethe specifically, future work could look at building a self-contained camera and measuring its form factor, efficacy, and energy consumption. For the general concept of privacy-preserving image processing, future work could look at applying the hardware limiting technique to other image-based and image-like sensors. Depth cameras are an interesting step in this direction with the possibility of using distance to detect people on a per pixel basis. Processing a depth camera's point cloud in a memory limited way would be an interesting technical challenge.

Future work in KindredRL includes expanding the model of similarity to include possible subgroups where users share some preferences but not all, designing techniques to automatically select application parameters such as the neighborhood size, and potentially varying the effect of each neighbor on the policy of the main user to account for differences within the neighborhood. Additionally, KindredRL could be integrated with various

other reinforcement learning techniques. This includes using alternate models, such as neural networks, and integrating traditionally experience relay techniques to boost the performance of Q-KindredRL in particular. Future work might also look at integrating KindredRL with traditionally collaborative reinforcement learning and evaluate the effect of similarity based learning on meeting a global, multi-agent objective. KindredRL could also be evaluated for its potential in other domains, such as recommendation systems tied to the physical world like clothing or recipe recommendation. Testing these techniques in an in-situ environment on a variety of applications would be an exciting direction for future work.

Beyond individual extensions to the approach in this work, future work might evaluate their intersection. In particular, examining the security and privacy risks of using similarity a learning system where the actions of one can affect the control of another provides an interesting direction. Additionally, using Lethe to provide better learning in the ThermalThrift through knowledge of an individual's water usage, rather than the household's, might improve prediction accuracy. Lethe could also allow KindredRL to learn individual preferences in multi-user environments. Future work might also apply these techniques to meet a variety of needs, including further societal needs and extending the solutions to meet personal needs as well.

Bibliography

[1] Derek Anderson, James M Keller, Marjorie Skubic, Xi Chen, and Zhihai He. Recognizing falls from silhouettes. In *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE*, pages 6388–6391. IEEE, 2006.

[2] Koen Steemers and Geun Young Yun. Household energy consumption: a study of the role of occupants. *Building Research & Information*, 37(5-6):625–637, 2009.

[3] Jie Zhao, Bertrand Lasternas, Khee Poh Lam, Ray Yun, and Vivian Loftness. Occupant behavior and schedule modeling for building energy simulation through office appliance power consumption data mining. *Energy and Buildings*, 82:341–355, 2014.

[4] Raymond De Young. Changing behavior and making it stick: The conceptualization and management of conservation behavior. *Environment and behavior*, 25(3):485–505, 1993.

[5] Emile HL Aarts and José Luis Encarnação. *True visions: The emergence of ambient intelligence*. Springer Science & Business Media, 2006.

[6] Current Population Reports. An aging nation: The older population in the united states. `https://www.census.gov/prod/2014pubs/p25-1140.pdf`, 2014.

[7] Ken Parsons. *Human thermal environments: the effects of hot, moderate, and cold environments on human health, comfort, and performance*. Crc Press, 2014.

[8] Norbert Streitz, Dimitris Charitos, Maurits Kaptein, and Marc Böhlen. Grand challenges for ambient intelligence and implications for design contexts and smart societies. *Journal of Ambient Intelligence and Smart Environments*, 11(1):87–107, 2019.

[9] Boris Kingma and Wouter van Marken Lichtenbelt. Energy consumption in buildings and female thermal demand. *Nature climate change*, 5(12):1054, 2015.

[10] Mark Weiser, Rich Gold, and John Seely Brown. The origins of ubiquitous computing research at parc in the late 1980s. *IBM systems journal*, 38(4):693–696, 1999.

[11] Timothy W Hnat, Erin Griffiths, Ray Dawson, and Kamin Whitehouse. Doorjamb: unobtrusive room-level tracking of people in homes using doorway sensors. In *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems*, pages 309–322. ACM, 2012.

[12] Avinash Kalyanaraman, Dezhi Hong, Elahe Soltanaghaei, and Kamin Whitehouse. Forma track: tracking people based on body shape. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(3):61, 2017.

[13] Yan Li and Ting Zhu. Using wi-fi signals to characterize human gait for identification and activity monitoring. In *2016 IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*, pages 238–247. IEEE, 2016.

[14] Chao Wang, Siwen Chen, Yanwei Yang, Feng Hu, Fugang Liu, and Jie Wu. Literature review on wireless sensing-wi-fi signal-based recognition of human activities. *Tsinghua Science and Technology*, 23(2):203–222, 2018.

[15] Fei Zuo and PHN de With. Real-time embedded face recognition for smart home. *IEEE transactions on consumer Electronics*, 51(1):183–190, 2005.

[16] Xiang Li, Daqing Zhang, Qin Lv, Jie Xiong, Shengjie Li, Yue Zhang, and Hong Mei. Indotrack: Device-free indoor human tracking with commodity wi-fi. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(3):72, 2017.

[17] Mohamad Daher, Ahmad Diab, Maan El Badaoui El Najjar, Mohamad Ali Khalil, and François Charpillet. Elder tracking and fall detection system using smart tiles. *IEEE Sensors Journal*, 17(2):469–479, 2016.

[18] Jake K Aggarwal and Michael S Ryoo. Human activity analysis: A review. *ACM Computing Surveys (CSUR)*, 43(3):16, 2011.

[19] Heba Abdelnasser, Moustafa Youssef, and Khaled A Harras. Wigest: A ubiquitous wifi-based gesture recognition system. In *2015 IEEE Conference on Computer Communications (INFOCOM)*, pages 1472–1480. IEEE, 2015.

[20] Shwetak N Patel, Matthew S Reynolds, and Gregory D Abowd. Detecting human movement by differential air pressure sensing in hvac system ductwork: An exploration in infrastructure mediated sensing. In *International Conference on Pervasive Computing*, pages 1–18. Springer, 2008.

[21] Edison Thomaz, Vinay Bettadapura, Gabriel Reyes, Megha Sandesh, Grant Schindler, Thomas Plötz, Gregory D Abowd, and Irfan Essa. Recognizing water-based activities in the home through infrastructure-mediated sensing. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pages 85–94. ACM, 2012.

[22] Andrej Grgurić, Miran Mošmondor, and Darko Huljenić. The smarthabits: An intelligent privacy-aware home care assistance system. *Sensors*, 19(4):907, 2019.

[23] Ana Cristina Bicharra Garcia, Adriana Santarosa Vivacqua, Nayat Sanchez-Pi, Luis Marti, and José M Molina. Crowd-based ambient assisted living to monitor the elderlys health outdoors. *IEEE Software*, 34(6):53–57, 2017.

[24] Abdur Rahim Mohammad Forkan, Ibrahim Khalil, Zahir Tari, Sebti Foufou, and Abdelaziz Bouras. A context-aware approach for long-term behavioural change detection and abnormality prediction in ambient assisted living. *Pattern Recognition*, 48(3):628–641, 2015.

[25] Marek Novák, Miroslav Biňas, and František Jakab. Unobtrusive anomaly detection in presence of elderly in a smart-home environment. In *2012 ELEKTRO*, pages 341–344. IEEE, 2012.

[26] UABUA Bakar, Hemant Ghayvat, SF Hasanm, and SC Mukhopadhyay. Activity and anomaly detection in smart home: A survey. In *Next Generation Sensors and Systems*, pages 191–220. Springer, 2016.

[27] Jingyu Liu, Wen Zhang, Xiaodong Chu, and Yutian Liu. Fuzzy logic controller for energy savings in a smart led lighting system considering lighting comfort and daylight. *Energy and Buildings*, 127:95–104, 2016.

[28] Fariba Sadri. Ambient intelligence: A survey. *ACM Computing Surveys (CSUR)*, 43(4):36, 2011.

[29] Matjaz Gams, Irene Yu-Hua Gu, Aki Härmä, Andrés Muñoz, and Vincent Tam. Artificial intelligence and ambient intelligence. *Journal of Ambient Intelligence and Smart Environments*, 11(1):71–86, 2019.

[30] I Chandra, N Sivakumar, Chandra Babu Gokulnath, and P Parthasarathy. Iot based fall detection and ambient assisted system for the elderly. *Cluster Computing*, pages 1–9, 2018.

[31] Parisa Rashidi and Alex Mihailidis. A survey on ambient-assisted living tools for older adults. *IEEE journal of biomedical and health informatics*, 17(3):579–590, 2012.

[32] Mohammed F Alhamid, Majdi Rawashdeh, Haiwei Dong, M Anwar Hossain, Abdulhameed Alelaiwi, and Abdulmotaleb El Saddik. Recam: a collaborative context-aware framework for multimedia recommendations in an ambient intelligence environment. *Multimedia Systems*, 22(5):587–601, 2016.

[33] Pervez Hameed Shaikh, Nursyarizal Bin Mohd Nor, Perumal Nallagownden, Irraivan Elamvazuthi, and Taib Ibrahim. Intelligent multi-objective control and management for smart energy efficient buildings. *International Journal of Electrical Power & Energy Systems*, 74:403–409, 2016.

[34] Israr Ullah and DoHyeun Kim. An improved optimization function for maximizing user comfort with minimum energy consumption in smart homes. *Energies*, 10(11):1818, 2017.

[35] Pervez Hameed Shaikh, Nursyarizal Bin Mohd Nor, Perumal Nallagownden, and Irraivan Elamvazuthi. Intelligent multi-objective optimization for building energy and comfort management. *Journal of King Saud University-Engineering Sciences*, 30(2):195–204, 2018.

[36] Alexandra Queirós, Anabela Silva, Joaquim Alvarelhão, Nelson Pacheco Rocha, and António Teixeira. Usability, accessibility and ambient-assisted living: a systematic literature review. *Universal Access in the Information Society*, 14(1):57–66, 2015.

[37] Thanos G Stavropoulos, Efstratios Kontopoulos, Nick Bassiliades, John Argyriou, Antonis Bikakis, Dimitris Vrakas, and Ioannis Vlahavas. Rule-based approaches for energy savings in an ambient intelligence environment. *Pervasive and Mobile Computing*, 19:1–23, 2015.

[38] Tianshu Wei, Yanzhi Wang, and Qi Zhu. Deep reinforcement learning for building hvac control. In *Proceedings of the 54th Annual Design Automation Conference 2017*, page 22. ACM, 2017.

[39] Alan F Westin and Oscar M Ruebhausen. *Privacy and freedom*, volume 1. Atheneum New York, 1967.

[40] Florian Kirchbuchner, Tobias Grosse-Puppendahl, Matthias R Hastall, Martin Distler, and Arjan Kuijper. Ambient intelligence from senior citizens perspectives: Understanding privacy concerns, technology acceptance, and expectations. In *European Conference on Ambient Intelligence*, pages 48–59. Springer, 2015.

[41] Charith Perera, Chang Liu, Rajiv Ranjan, Lizhe Wang, and Albert Y Zomaya. Privacy-knowledge modeling for the internet of things: A look back. *Computer*, 49(12):60–68, 2016.

[42] Eva-Maria Schomakers and Martina Ziefle. Privacy perceptions in ambient assisted living. 2019.

[43] José L Hernández-Ramos, M Victoria Moreno, Jorge Bernal Bernabé, Dan García Carrillo, and Antonio F Skarmeta. Safir: Secure access framework for iot-enabled services on smart buildings. *Journal of Computer and System Sciences*, 81(8):1452–1463, 2015.

[44] Mohammed Riyadh Abdmeziem and Djamel Tandjaoui. An end-to-end secure key management protocol for e-health applications. *Computers & Electrical Engineering*, 44:184–197, 2015.

[45] Pawani Porambage, An Braeken, Andrei Gurtov, Mika Ylianttila, and Susanna Spinsante. Secure end-to-end communication for constrained devices in iot-enabled ambient assisted living systems. In *2015 IEEE 2nd World Forum on Internet of Things (WF-IoT)*, pages 711–714. IEEE, 2015.

[46] Mohamed Elhoseny, Xiaohui Yuan, Hamdy K El-Minir, and Alaa Mohamed Riad. An energy efficient encryption method for secure dynamic wsn. *Security and Communication Networks*, 9(13):2024–2031, 2016.

[47] Petra Grd, Igor Tomičić, and Miroslav Baca. Privacy improvement model for biometric person recognition in ambient intelligence using perceptual hashing. In *Proceedings of the Central European Cybersecurity Conference 2018*, page 18. ACM, 2018.

[48] Fadele Ayotunde Alaba, Mazliza Othman, Ibrahim Abaker Targio Hashem, and Faiz Alotaibi. Internet of things security: A survey. *Journal of Network and Computer Applications*, 88:10–28, 2017.

[49] Madhav Sharma and David Biros. Information privacy concerns in the age of internet of things. 2018.

[50] Chimezie Leonard Oguego, Juan Carlos Augusto, Andrés Muñoz, and Mark Springett. A survey on managing users preferences in ambient intelligence. *Universal Access in the Information Society*, 17(1):97–114, 2018.

[51] Mark Feldmeier and Joseph A Paradiso. Personalized hvac control system. In *2010 Internet of Things (IOT)*, pages 1–8. IEEE, 2010.

[52] Hani Hagras, Victor Callaghan, Martin Colley, Graham Clarke, Anthony Pounds-Cornish, and Hakan Duman. Creating an ambient-intelligence environment using embedded agents. *IEEE Intelligent Systems*, 19(6):12–20, 2004.

[53] Michael Mozer, Robert Dodier, Debra Miller, Marc Anderson, J Anderson, D Bertini, M Bronder, M Colagrosso, R Cruickshank, B Daugherty, et al. The adaptive house. In *IEE Seminar Digests*, volume 11059, pages v1–39. IET, 2005.

[54] Eduardo Velloso, Markus Wirth, Christian Weichel, Augusto Esteves, and Hans Gellersen. Ambigaze: Direct control of ambient devices by gaze. In *Proceedings of the 2016 ACM conference on designing interactive systems*, pages 812–817. ACM, 2016.

[55] Abdul-Rahman Al-Ali, Imran A Zualkernan, Mohammed Rashid, Ragini Gupta, and Mazin Alikarar. A smart home energy management system using iot and big data analytics approach. *IEEE Transactions on Consumer Electronics*, 63(4):426–434, 2017.

[56] Valérian Guivarch, Valérie Camps, and André Péninou. Context awareness in ambient systems by an adaptive multi-agent approach. In *International Joint Conference on Ambient Intelligence*, pages 129–144. Springer, 2012.

[57] Annette Evans, Vladimir Strezov, and Tim J Evans. Assessment of utility energy storage options for increased renewable energy penetration. *Renewable and Sustainable Energy Reviews*, 16(6):4141–4147, 2012.

[58] R Parameshwaran and S Kalaiselvam. Energy conservative air conditioning system using silver nano-based pcm thermal storage for modern buildings. *Energy and Buildings*, 69:202–212, 2014.

[59] Nabil Nassif, C Hall, and D Freelnad. Optimization of ice thermal storage system design for hvac systems. *Cell*, 336:740–1601, 2013.

[60] MJ Sebzali, B Ameer, and HJ Hussain. Comparison of energy performance and economics of chilled water thermal storage and conventional air-conditioning systems. *Energy and Buildings*, 69:237–250, 2014.

[61] Eurostat. Number of private households. http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=lfst_hhnhtych&lang=en, 2015.

[62] BSRIA. Trends in the world traditional and renewable heating markets. http://www.iea.org/media/workshops/2014/buildingwebinars/ webinar4/4bsriakrystynapresentationforieafinal.pdf, 2014.

[63] Wikipedia. "List of energy storage projects". Wikipedia, The Free Encyclopedia, Accessed online: 24-June-2016.

[64] P Ravi Babu and Sree Divya Vadlapudi Prasanna. Intelligent water heater demand controller for consumer satisfaction and energy conservation. *i-Manager's Journal on Circuits & Systems*, 2(4):21, 2014.

[65] Peter Kepplinger, Gerhard Huber, and Jörg Petrasch. Autonomous optimal control for demand side management with resistive domestic hot water heaters using linear optimization. *Energy and Buildings*, 100:50–55, 2015.

[66] Arnaldo Sepulveda, Liam Paull, Walid G Morsi, Howard Li, CP Diduch, and Liuchen Chang. A novel demand side management program using water heaters and particle swarm optimization. In *Electric Power and Energy Conference (EPEC), 2010 IEEE*, pages 1–5. IEEE, 2010.

[67] Khalid Elgazzar, Howard Li, and Liuchen Chang. A centralized fuzzy controller for aggregated control of domestic water heaters. In *Electrical and Computer Engineering, 2009. CCECE'09. Canadian Conference on*, pages 1141–1146. IEEE, 2009.

[68] NewsChannel 10. Newschannel 10 investigates: Xcel energy saver switch boxes. http://www.newschannel10.com/story/19185280/ newschannel-10-investigates-xcel-energy-savor-switch-boxes, 2012.

[69] Ammi Amarnath Harshal Upadhye, Ron Domitrovic. Evaluating peak load shifting abilities and regulation service potential of a grid connected residential water heater. *ACEEE Summer Study on Energy Efficiency in Buildings*, 2012.

[70] Marcel Lacroix. Electric water heater designs for load shifting and control of bacterial contamination. *Energy conversion and management*, 40(12):1313–1340, 1999.

[71] M Perlman and BE Mills. Development of residential hot water use patterns. *ASHRAE transactions*, 91(2A):657–679, 1985.

[72] Sarada Kuravi, Jamie Trahan, D Yogi Goswami, Muhammad M Rahman, and Elias K Stefanakos. Thermal energy storage technologies and systems for concentrating solar power plants. *Progress in Energy and Combustion Science*, 39(4):285–319, 2013.

[73] Yongjun Sun, Shengwei Wang, Fu Xiao, and Diance Gao. Peak load shifting control using different cold thermal energy storage facilities in commercial buildings: a review. *Energy Conversion and Management*, 71:101–114, 2013.

[74] N Soares, JJ Costa, AR Gaspar, and P Santos. Review of passive pcm latent heat thermal energy storage systems towards buildings energy efficiency. *Energy and buildings*, 59:82–103, 2013.

[75] Dan Zhou, Chang-Ying Zhao, and Yuan Tian. Review on thermal energy storage with phase change materials (pcms) in building applications. *Applied energy*, 92:593–605, 2012.

[76] Massimo Fiorentini, Paul Cooper, and Zhenjun Ma. Development and optimization of an innovative hvac system with integrated pvt and pcm thermal storage for a net-zero energy retrofitted house. *Energy and Buildings*, 94:21–32, 2015.

[77] Jan Širokỳ, Frauke Oldewurtel, Jiří Cigler, and Samuel Prívara. Experimental analysis of model predictive control for an energy efficient building heating system. *Applied Energy*, 88(9):3079–3087, 2011.

[78] Petru-Daniel Moroşan, Romain Bourdais, Didier Dumur, and Jean Buisson. Building temperature regulation using a distributed model predictive control. *Energy and Buildings*, 42(9):1445–1452, 2010.

[79] Frauke Oldewurtel, Alessandra Parisio, Colin N Jones, Manfred Morari, Dimitrios Gyalistras, Markus Gwerder, Vanessa Stauch, Beat Lehmann, and Katharina Wirth. Energy efficient building climate control using stochastic model predictive control and weather predictions. In *American control conference (ACC), 2010*, pages 5100–5105. IEEE, 2010.

[80] Yudong Ma, Francesco Borrelli, Brandon Hencey, Brian Coffey, Sorin Bengea, and Philip Haves. Model predictive control for the operation of building cooling systems. *Control Systems Technology, IEEE Transactions on*, 20(3):796–803, 2012.

[81] Jiakang Lu, Tamim Sookoor, Vijay Srinivasan, Ge Gao, Brian Holben, John Stankovic, Eric Field, and Kamin Whitehouse. The smart thermostat: using occupancy sensors to save energy in homes. In *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems*, pages 211–224. ACM, 2010.

[82] Varick L Erickson, Miguel Á Carreira-Perpiñán, and Alberto E Cerpa. Observe: Occupancy-based system for efficient reduction of hvac energy. In *Information Processing in Sensor Networks (IPSN), 2011 10th International Conference on*, pages 258–269. IEEE, 2011.

[83] Alexander Belov, Nirvana Meratnia, Berend Jan van der Zwaag, and PJM Havinga. An efficient water flow control approach for water heaters in direct load control. 2014.

[84] University of Wisconsin-Madison. Solar Energy Laboratory and Sanford A Klein. *TRNSYS, a transient system simulation program*. Solar Energy Laborataory, University of Wisconsin–Madison, 1979.

[85] University of Wisconsin–Madison. Solar Energy Laboratory and Klein, Sanford A. A transient systems simulation program. http://sel.me.wisc.edu/trnsys/, 2013.

[86] M Shaad, R Errouissi, CP Diduch, ME Kaye, and L Chang. Aggregate load forecast with payback model of the electric water heaters for a direct load control program. In *Electrical Power and Energy Conference (EPEC), 2014 IEEE*, pages 214–219. IEEE, 2014.

[87] Linas Gelažanskas and Kelum AA Gamage. Distributed energy storage using residential hot water heaters. *Energies*, 9(3):127, 2016.

[88] Kelcey Lajoie, Douglas A Halamay, and Ted KA Brekken. Residential water heaters as a grid-scale energy storage solution using model predictive control. In *Technologies for Sustainability (SusTech), 2013 1st IEEE Conference on*, pages 62–69. IEEE, 2013.

[89] S Ali Pourmousavi, Stasha N Patrick, and M Hashem Nehrir. Real-time demand response through aggregate electric water heaters for load shifting and balancing wind generation. *Smart Grid, IEEE Transactions on*, 5(2):769–778, 2014.

[90] Fang Tang and Amine Bermak. An 84 pw/frame per pixel current-mode cmos image sensor with energy harvesting capability. *IEEE Sensors Journal*, 12(4):720–726, 2012.

[91] D Sabuncuoglu Tezcan, Selim Eminoglu, and Tayfun Akin. A low-cost uncooled infrared microbolometer detector in standard cmos technology. *IEEE Transactions on electron devices*, 50(2):494–502, 2003.

[92] Zhengyou Zhang. Microsoft kinect sensor and its effect. *IEEE multimedia*, 19(2):4–10, 2012.

[93] Shihong Lao and Masato Kawade. Vision-based face understanding technologies and their applications. In *Advances in Biometric Person Authentication*, pages 339–348. Springer, 2004.

[94] Paul Smith, Mubarak Shah, and Niels da Vitoria Lobo. Monitoring head/eye motion for driver alertness with one camera. In *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, volume 4, pages 636–642. IEEE, 2000.

[95] U.S. Energy Information Administration. "Use of Energy in the United States"., Accessed online:19-June-2019.

[96] Neil E Klepeis, William C Nelson, Wayne R Ott, John P Robinson, Andy M Tsang, Paul Switzer, Joseph V Behar, Stephen C Hern, and William H Engelmann. The national human activity pattern survey (nhaps): a resource for assessing exposure to environmental pollutants. *Journal of Exposure Science and Environmental Epidemiology*, 11(3):231, 2001.

[97] Carman Neustaedter, Saul Greenberg, and Michael Boyle. Blur filtration fails to preserve privacy for home-based video conferencing. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 13(1):1–36, 2006.

[98] Eun Kyoung Choe, Sunny Consolvo, Jaeyeon Jung, Beverly Harrison, and Julie A Kientz. Living in a glass house: a survey of private moments in the home. In *Proceedings of the 13th international conference on Ubiquitous computing*, pages 41–44. ACM, 2011.

[99] Roberto Hoyle, Robert Templeman, Steven Armes, Denise Anthony, David Crandall, and Apu Kapadia. Privacy behaviors of lifeloggers using wearable cameras. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 571–582. ACM, 2014.

[100] John Vines, Stephen Lindsay, Gary W Pritchard, Mabel Lie, David Greathead, Patrick Olivier, and Katie Brittain. Making family care work: dependence, privacy and remote home monitoring telecare systems. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, pages 607–616. ACM, 2013.

[101] Alexandra König, Carlos Fernando Crispim Junior, Alexandre Derreumaux, Gregory Bensadoun, Pierre-David Petit, François Bremond, Renaud David, Frans Verhey, Pauline Aalten, and Philippe Robert. Validation of an automatic video monitoring system for the detection of instrumental activities of daily living in dementia patients. *Journal of Alzheimer's Disease*, 44(2):675–685, 2015.

[102] Florian Schaub, Bastian Konings, and Michael Weber. Context-adaptive privacy: Leveraging context awareness to support privacy decision making. *Pervasive Computing, IEEE*, 14(1):34–43, 2015.

[103] Tim Coughlan, Kerstin Leder Mackley, Michael Brown, Sarah Martindale, Stephan Schlögl, Becky Mallaband, John Arnott, Jettie Hoonhout, Dalila Szostak, Robin Brewer, et al. Current issues and future directions in methods for studying technology in the home. *PsychNology Journal*, 11(2):159–184, 2013.

[104] Anjan Chakravorty, Tomasz Wlodarczyk, and Chunming Rong. Privacy preserving data analytics for smart homes. In *Security and Privacy Workshops (SPW), 2013 IEEE*, pages 23–27. IEEE, 2013.

[105] Tiffany Hyun-Jin Kim, Lujo Bauer, James Newsome, Adrian Perrig, and Jesse Walker. Challenges in access right assignment for secure home networks. In *HotSec*, 2010.

[106] Joshua D. Neuheisel Diego A. Socolinsky, Andrea Selinger. Face recognition with visible and thermal infrared imagery. *Computer Vision and Image Understanding*, Volume 91, 2003.

[107] Atousa Torabi, Guillaume Massé, and Guillaume-Alexandre Bilodeau. An iterative integrated framework for thermal–visible image registration, sensor fusion, and people tracking for video surveillance applications. *Computer Vision and Image Understanding*, 116(2):210–221, 2012.

[108] Alex Leykin, Yang Ran, and Riad Hammoud. Thermal-visible video fusion for moving target tracking and pedestrian classification. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–8. IEEE, 2007.

[109] Chenglong Li, Shiyi Hu, Sihan Gao, and Jin Tang. Real-time grayscale-thermal tracking via laplacian sparse representation. In *International Conference on Multimedia Modeling*, pages 54–65. Springer, 2016.

[110] Hessam Mohammadmoradi, Sirajum Munir, Omprakash Gnawali, and Charles Shelton. Measuring people-flow through doorways using easy-to-install ir array sensors.

[111] Alex Beltran, Varick L Erickson, and Alberto E Cerpa. Thermosense: Occupancy thermal based sensing for hvac control. In *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings*, pages 1–8. ACM, 2013.

[112] Alex Edgcomb and Frank Vahid. Privacy perception and fall detection accuracy for in-home video assistive monitoring with privacy enhancements. *ACM SIGHIT Record*, 2(2):6–15, 2012.

[113] Pavel Korshunov, Claudia Araimo, Francesca De Simone, Carmelo Velardo, J-L Dugelay, and Touradj Ebrahimi. Subjective study of privacy filters in video surveillance. In *Multimedia Signal Processing (MMSP), 2012 IEEE 14th International Workshop on*, pages 378–382. Ieee, 2012.

[114] Timothy W Hnat, Erin Griffiths, Ray Dawson, and Kamin Whitehouse. Doorjamb: unobtrusive room-level tracking of people in homes using doorway sensors. In *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems*, pages 309–322. ACM, 2012.

[115] Fei Peng, Xiao-wen Zhu, and Min Long. An roi privacy protection scheme for h. 264 video based on fmo and chaos. *Information Forensics and Security, IEEE Transactions on*, 8(10):1688–1699, 2013.

[116] Rikke Gade and Thomas B Moeslund. Thermal cameras and applications: a survey. *Machine vision and applications*, 25(1):245–262, 2014.

[117] Andrew Sixsmith and Neil Johnson. A smart sensor to detect the falls of the elderly. *IEEE Pervasive computing*, 3(2):42–47, 2004.

[118] Eslam Mostafa, Riad Hammoud, Asem Ali, and Aly Farag. Face recognition in low resolution thermal images. *Computer Vision and Image Understanding*, 117(12):1689–1694, 2013.

[119] Zhan Lu, Yichen Xu, Zuoxiao Dai, and Bei Ma. A gait recognition based on link model of infrared thermal imaging. In *Control Science and Systems Engineering (ICCSSE), 2016 2nd International Conference on*, pages 165–168. IEEE, 2016.

[120] Sa-Mun Kim, Dae-Jong Lee, Ho-Hyun Lee, and Myung-Geun Chun. Person recognition using gait and face features on thermal images. *The Transactions of the Korean Institute of Electrical Engineers P*, 65(2):130–135, 2016.

[121] Elliot N Saba, Eric C Larson, and Shwetak N Patel. Dante vision: In-air and touch gesture sensing for natural surface interaction with combined depth and thermal cameras. In *Emerging Signal Processing Applications (ESPA), 2012 IEEE International Conference on*, pages 167–170. IEEE, 2012.

[122] Eric Larson, Gabe Cohn, Sidhant Gupta, Xiaofeng Ren, Beverly Harrison, Dieter Fox, and Shwetak Patel. Heatwave: thermal imaging for surface user interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2565–2574. ACM, 2011.

[123] Jörg Appenrodt, Ayoub Al-Hamadi, and Bernd Michaelis. Data gathering for gesture recognition systems based on single color-, stereo color-and thermal cameras. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 3(1):37–50, 2010.

[124] Zui Zhang, Oscar Perez Concha, and Massimo Piccardi. Tracking people under heavy occlusions by layered data association. *Multimedia Tools and Applications*, 74(17):7239–7259, 2015.

[125] Hongyang Xue, Yao Liu, Deng Cai, and Xiaofei He. Tracking people in rgbd videos using deep learning and motion clues. *Neurocomputing*, 204:70–76, 2016.

[126] Adam Gudyś, Jakub Rosner, Jakub Segen, Konrad Wojciechowski, and Marek Kulbacki. Tracking people in video sequences by clustering feature motion paths. In *Computer Vision and Graphics*, pages 236–245. Springer, 2014.

[127] Horesh Ben Shitrit, Jérôme Berclaz, François Fleuret, and Pascal Fua. Multi-commodity network flow for tracking multiple people. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 36(8):1614–1627, 2014.

[128] Matteo Munaro, Filippo Basso, and Emanuele Menegatti. Tracking people within groups with rgb-d data. In *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, pages 2101–2107. IEEE, 2012.

[129] D. Gafurov and E. Snekkenes. Gait recognition using wearable motion recording sensors. *EURASIP Journal on Advances in Signal Processing*, 2009:7, 2009.

[130] Z. Liu and S. Sarkar. Outdoor recognition at a distance by fusing gait and face. *Image and Vision Computing*, 25(6):817–832, 2007.

[131] Shreyas Venugopalan, Unni Prasad, Khalid Harun, Kyle Neblett, Douglas Toomey, Joseph Heyman, and Marios Savvides. Long range iris acquisition system for stationary and mobile subjects. In *Biometrics (IJCB), 2011 International Joint Conference on*, pages 1–8. IEEE, 2011.

[132] Chiraz BenAbdelkader, Ross Cutler, and Larry Davis. Person identification using automatic height and stride estimation. In *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, volume 4, pages 377–380. IEEE, 2002.

[133] Tianzhu Zhang, Kui Jia, Changsheng Xu, Yi Ma, and Narendra Ahuja. Partial occlusion handling for visual tracking via robust part matching. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1258–1265. IEEE, 2014.

[134] Savvas Papaioannou, Andrew Markham, and Niki Trigoni. Tracking people in highly dynamic industrial environments. *IEEE Transactions on Mobile Computing*, 2016.

[135] R Rajeev and SS Sridhar. Tracking people motion based on lloyds clustering algorithm. *Int. J. Comput. Sci. Inf. Technol*, 5(2):2289–2292, 2014.

[136] Francois Fleuret, Jerome Berclaz, and Richard Lengagne. Multicamera people tracking with a probabilistic occupancy map. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30:267–282, 2008.

[137] Andreas Savvides Thiago Teixeira. Lightweight people counting and localizing in indoor spaces using camera sensor nodes. *First ACM/IEEE International Conference on Distributed Smart Cameras*, 2007.

[138] M.D. Addlesee, A. Jones, F. Livesey, and F. Samaria. The orl active floor [sensor system]. *Personal Communications, IEEE*, 4(5):35–41, 1997.

[139] Y.L. Shen and C.S. Shin. Distributed sensing floor for an intelligent environment. *Sensors Journal, IEEE*, 9(12):1673–1678, 2009.

[140] Yunze Zeng, Parth H Pathak, and Prasant Mohapatra. Wiwho: Wifi-based person identification in smart spaces. In *2016 15th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, pages 1–12. IEEE, 2016.

[141] Wei Wang, Alex X Liu, and Muhammad Shahzad. Gait recognition using wifi signals. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 363–373. ACM, 2016.

[142] Nacer Khalil, Driss Benhaddou, Omprakash Gnawali, and Jaspal Subhlok. Nonintrusive occupant identification by sensing body shape and movement. In *Proceedings of the 3rd ACM International Conference on Systems for Energy-Efficient Built Environments*, pages 1–10. ACM, 2016.

[143] The Physics Factbook. Temperature of a healthy human (skin temperature). `http://hypertextbook.com/facts/2001/AbantyFarzana.shtml`, 2014.

[144] FiveThrityEight. How common is it for a man to be shorter than his partner? `https://fivethirtyeight.com/datalab/how-common-is-it-for-a-man-to-be-shorter-than-his-partner/`, 2014.

[145] The Atlantic. Why it's so rare for a wife to be taller than her husband. `https://www.theatlantic.com/sexes/archive/2013/01/why-its-so-rare-for-a-wife-to-be-taller-than-her-husband/272585/`, 2013.

[146] Avinash Kalyanaraman, Erin Griffiths, and Kamin Whitehouse. Transtrack: Tracking multiple targets by sensing their zone transitions. In *Distributed Computing in Sensor Systems (DCOSS), 2016 International Conference on*, pages 59–66. IEEE, 2016.

[147] Sirajum Munir, Ripudaman Singh Arora, Craig Hesling, Juncheng Li, Jonathan Francis, Charles Shelton, Christopher Martin, Anthony Rowe, and Mario Berges. Real-time fine grained occupancy estimation using depth sensors on arm embedded platforms. In *Real-Time and Embedded Technology and Applications Symposium (RTAS), 2017 IEEE*, pages 295–306. IEEE, 2017.

[148] Tianzhen Hong, Sarah C Taylor-Lange, Simona DOca, Da Yan, and Stefano P Corgnati. Advances in research and applications of energy-related occupant behavior in buildings. *Energy and buildings*, 116:694–702, 2016.

[149] Bernt Meerbeek, Thijs van Druenen, Mariëlle Aarts, Evert van Loenen, and Emile Aarts. Impact of blinds usage on energy consumption: automatic versus manual control. In *European Conference on Ambient Intelligence*, pages 158–173. Springer, 2014.

[150] J Fergus Nicol and Michael A Humphreys. Adaptive thermal comfort and sustainable thermal standards for buildings. *Energy and buildings*, 34(6):563–572, 2002.

[151] Christopher JCH Watkins and Peter Dayan. Q-learning. *Machine learning*, 8(3-4):279–292, 1992.

[152] Sander Adam, Lucian Busoniu, and Robert Babuska. Experience replay for real-time reinforcement learning control. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 42(2):201–212, 2011.

[153] Sherief Abdallah and Michael Kaisers. Addressing environment non-stationarity by repeating q-learning updates. *The Journal of Machine Learning Research*, 17(1):1582–1612, 2016.

[154] Shi-Yong Chen, Yang Yu, Qing Da, Jun Tan, Hai-Kuan Huang, and Hai-Hong Tang. Stabilizing reinforcement learning in dynamic environment with application to online recommendation. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1187–1196. ACM, 2018.

[155] Bruno C Da Silva, Eduardo W Basso, Ana LC Bazzan, and Paulo M Engel. Dealing with non-stationary environments using context detection. In *Proceedings of the 23rd international conference on Machine learning*, pages 217–224. ACM, 2006.

[156] Renata De Vecchi, Roberto Lamberts, and Christhina Maria Candido. The role of clothing in thermal comfort: how people dress in a temperate and humid climate in brazil. *Ambiente Construído*, 17(1):69–81, 2017.

[157] Chungyoon Chun, Alison Kwok, Teruaki Mitamura, Norie Miwa, and Akihiro Tamura. Thermal diary: Connecting temperature history to indoor comfort. *Building and Environment*, 43(5):877–885, 2008.

[158] Athanasios S Polydoros and Lazaros Nalpantidis. Survey of model-based reinforcement learning: Applications on robotics. *Journal of Intelligent & Robotic Systems*, 86(2):153–173, 2017.

[159] Mufti Mahmud, Mohammed Shamim Kaiser, Amir Hussain, and Stefano Vassanelli. Applications of deep learning and reinforcement learning to biological data. *IEEE transactions on neural networks and learning systems*, 29(6):2063–2079, 2018.

[160] Konstantinos Dalamagkidis, Denia Kolokotsa, Konstantinos Kalaitzakis, and George S Stavrakakis. Reinforcement learning for energy conservation and comfort in buildings. *Building and environment*, 42(7):2686–2698, 2007.

[161] Mehdi Mohammadi, Ala Al-Fuqaha, Mohsen Guizani, and Jun-Seok Oh. Semisupervised deep reinforcement learning in support of iot and smart city services. *IEEE Internet of Things Journal*, 5(2):624–635, 2018.

[162] Elena Mocanu, Decebal Constantin Mocanu, Phuong H Nguyen, Antonio Liotta, Michael E Webber, Madeleine Gibescu, and Johannes G Slootweg. On-line building energy optimization using deep reinforcement learning. *IEEE Transactions on Smart Grid*, 2018.

[163] Zhiang Zhang, Adrian Chong, Yuqi Pan, Chenlu Zhang, Siliang Lu, and Khee Poh Lam. A deep reinforcement learning approach to using whole building energy model for hvac optimal control. In *2018 Building Performance Analysis Conference and SimBuild*, 2018.

[164] Elad Liebman, Maytal Saar-Tsechansky, and Peter Stone. Dj-mc: A reinforcement-learning agent for music playlist recommendation. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*, pages 591–599. International Foundation for Autonomous Agents and Multiagent Systems, 2015.

[165] Nima Taghipour, Ahmad Kardan, and Saeed Shiry Ghidary. Usage-based web recommendations: a reinforcement learning approach. In *Proceedings of the 2007 ACM conference on Recommender systems*, pages 113–120. ACM, 2007.

[166] Leslie Pack Kaelbling, Michael L Littman, and Andrew W Moore. Reinforcement learning: A survey. *Journal of artificial intelligence research*, 4:237–285, 1996.

[167] Andrei Marinescu, Ivana Dusparic, and Siobhán Clarke. Prediction-based multi-agent reinforcement learning in inherently non-stationary environments. *ACM Transactions on Autonomous and Adaptive Systems (TAAS)*, 12(2):9, 2017.

[168] Qingyun Wu, Naveen Iyer, and Hongning Wang. Learning contextual bandits in a non-stationary environment. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pages 495–504. ACM, 2018.

[169] Kfir Arviv, Helman Stern, and Yael Edan. Collaborative reinforcement learning for a two-robot job transfer flow-shop scheduling problem. *International Journal of Production Research*, 54(4):1196–1209, 2016.

[170] Monireh Abdoos, Nasser Mozayani, and Ana LC Bazzan. Traffic light control in non-stationary environments based on multi agent q-learning. In *Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference on*, pages 1580–1585. IEEE, 2011.

[171] Liane Okdinawati, Togar M Simatupang, and Yos Sunitiyoso. Multi-agent reinforcement learning for value co-creation of collaborative transportation management (ctm). *International Journal of Information Systems and Supply Chain Management (IJISSCM)*, 10(3):84–95, 2017.

[172] Liangliang Ren, Jiwen Lu, Zifeng Wang, Qi Tian, and Jie Zhou. Collaborative deep reinforcement learning for multi-object tracking. In *European Conference on Computer Vision*, pages 605–621. Springer, 2018.

[173] Xiangyu Kong, Bo Xin, Yizhou Wang, and Gang Hua. Collaborative deep reinforcement learning for joint object search. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[174] Kai Arulkumaran, Marc Peter Deisenroth, Miles Brundage, and Anil Anthony Bharath. Deep reinforcement learning: A brief survey. *IEEE Signal Processing Magazine*, 34(6):26–38, 2017.

[175] Qingyun Wu, Huazheng Wang, Quanquan Gu, and Hongning Wang. Contextual bandits in a collaborative environment. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, pages 529–538. ACM, 2016.

[176] NREL. National solar radiation database, 2017.

[177] Nicholas C Romano and Jay F Nunamaker. Meeting analysis: Findings from research and practice. In *Proceedings of the 34th Annual Hawaii International Conference on System Sciences*, pages 13–pp. IEEE, 2001.