Control of Electric Load Aggregations for Power System Services

by

Stephanie Crocker Ross

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Electrical Engineering: Systems)
in the University of Michigan
2020

Doctoral Committee:

Assistant Professor Johanna Mathieu, Chair
Professor Ian Hiskens
Associate Professor Shelie Miller
Associate Professor Necmiye Ozay

Stephanie C. Ross
sjcrock@umich.edu
ORCID iD: 0000-0001-8776-414X

Dedication

To Eric.

Acknowledgments

Thank you to the community that has supported me throughout the last five and a half years. I could not have done it without you.

Thank you most of all to my advisor Professor Johanna Mathieu. Thank you for your kindness and encouragement when I hit stressful times, and thank you for your flexibility and trust as I finished my degree remotely. I have learned so much from you and am thankful to have had you as my advisor.

Thank you to the National Science Foundation (NSF) and the Rackham Graduate School for funding support. Specifically, my work has been supported by the NSF Graduate Research Fellowship Program under Grant No. DGE 125626, by the NSF Cyber-Physical Systems Program under Grant No. CNS-1837680, and by the Rackham Graduate School through the Rackham Predoctoral Fellowship.

Table of Contents

List of Figures

Figure

List of Tables

Table

Abstract

In electrical power systems, when the supply from wind or solar-powered generation fluctuates, other resources adjust their power to maintain the system's balance between demand and supply. Traditionally, gas, coal, and hydro-powered generators have provided this balancing service. In the future, as the proportion of renewable power generation increases, additional balancing resources will be needed. In this work, we develop methods that enable a new resource—aggregations of flexible loads—to provide energy balancing.

Load aggregations are a promising resource for transmission-level energy balancing, but this service should not come at the expense of lower-level services and requirements. Specifically, an aggregator's control should not compromise the loads' service to the end-user and should not cause operational issues on the distribution network. Thermostatically controlled loads (TCLs), such as air conditioners and water heaters, have user-set temperature limits and cycling constraints that must be satisfied. Distribution networks have loading and voltage constraints to ensure reliable operation. When providing balancing services, aggregators partially synchronize loads, which can cause constraint violations on the distribution network. Third-party aggregators are unaware of conditions on the network and must coordinate with the distribution operator to ensure network reliability.

The objective of this dissertation is to develop control methods by which a third-party aggregator can provide energy balancing without disrupting consumers and without causing unsafe conditions on the distribution network.

Multiple methods are proposed for identifying and protecting distribution constraints that are at risk of violation due to load control. We conduct a simulation study of realistic distribution networks and find only a small subset of network constraints is at risk of violation. This result implies that network-safe control strategies may need to account for only a subset of network constraints, enhancing computational efficiency. We propose using a "mode-count algorithm'' to control a group of TCLs to minimize their impact on an at-risk network constraint. Results show that the mode-count algorithm can effectively reduce the variability of voltage at a constrained distribution node. Developing an online method to identify the

set of at-risk constraints is non-trivial; towards this end, we propose an optimization-based method that identifies the network's most at-risk individual constraint and provides a conservative, global safety constraint on power deviations caused by the aggregator. Because the method is computationally intensive, we develop techniques based on power-flow analysis to reduce the problem size; we are able to reduce the problem size by more than 60% for a test network.

Two network-safe control strategies for energy balancing are proposed. Both strategies are hierarchical: the aggregator controls loads to track an energy-balancing signal, and the operator removes particular TCLs from the aggregator's control when necessary for network safety. The strategies differ in terms of modeling and communication requirements. In a case study, the more complex strategy achieves a root-mean-square tracking error of 0.10% of the TCLs' baseline power consumption while removing fewer than 1% of TCLs from the aggregator's control; the other strategy achieves a 0.70% tracking error while removing approximately 15% of TCLs. The two strategies provide options—one better performing, one less costly—for operators and aggregators with different capabilities and preferences. Overall, these strategies enable third-party aggregators to control larger proportions of distribution-network load, enhancing competition in wholesale markets and providing the greater balancing capacities that will be needed by future, low-carbon power systems.

Chapter 1

Introduction

1.1 Motivation

The electric power system must transition to clean energy sources to reduce society's carbon emissions and mitigate the effects of climate change. In 2017, the electricity sector accounted for 28% of the U.S.'s greenhouse gas emissions, just one percentage point less than the sector with the most emissions—transportation [85]. One strategy for reducing emissions is to transition the electricity sector to 100% clean energy sources and then "electrify everything", that is, convert other sectors, such as transportation and space heating, to electrical power as much as possible. The U.S. electricity sector has a long way to go before it is fully decarbonized: in 2018, the sources for electricity generation were 62% natural gas and coal, 20% nuclear, and 17% renewable [84].

Increasing the proportion of generation from renewable sources poses challenges to reliable power system operation because of the intermittency and uncertainty of wind- and solar-powered generation. A fundamental aspect of reliable power system operation is energy balancing: energy supplied must equal energy demanded, on the time scale of seconds, or else the system may become unstable. Traditionally, supply has been controlled to follow variations in demand, and the system's major balancing resources have been gas-, coal-, and hydro-powered generators. As carbon-emitting generators are retired, the grid's balancing requirements will need to be met by other resources. Two such resources are batteries and demand-side resources, frequently referred to as demand response. Not only are these resources zero-emission but they may also perform better, responding faster and with higher accuracy, than traditional balancing resources [8].

One promising resource for balancing—and the subject of this dissertation—is aggregations

of small, flexible electrical loads. By aggregating thousands of kilowatt (kW) scale loads, an aggregator can provide balancing at the megawatt (MW) scale that is needed by the greater power system. Prime candidates for aggregation include electric vehicles (EVs), home back-up battery systems, and thermostatically controlled loads (TCLs) (e.g., heat pumps, air conditioners, and electric water heaters). These loads all have some flexibility in when they draw power from the power system. EVs and home battery systems have flexibility due to their electrical energy storage capacity; TCLs have flexibility due to their thermal energy storage capacity—a house does not instantly become hot when an air conditioner cycles off. (Note, here we refer to residential batteries and EVs as loads because that is how they are most often operated.)

The capacity of loads that are suitable for balancing is already large [50] and is likely to grow due to electrification trends and policies [47]. EV adoption has been increasing in the U.S. [35], and ten states support the sales of EVs through Zero Emission Vehicle mandates [80]. Policies supporting electrification of home appliances are nascent. In 2019, Berkeley, CA became the first U.S. city to pass an ordinance banning natural gas connections for new buildings [29]. If other cities follow suit, there will be a substantial increase in electric space and water heating loads available for aggregation.

This dissertation addresses three technical challenges of providing balancing with load aggregations. The first challenge, which inspires much of the research in this area, is to develop a control strategy that coordinates loads to provide energy balancing while also ensuring the end-users of the loads are not inconvenienced. The second challenge is to ensure that the control strategy works well given the complexities of load behavior in the real-world. The third challenge—and the challenge that has inspired the majority of this dissertation—is to ensure that aggregator-controlled loads do not cause operational issues on the distribution network. (The distribution network is the low-voltage network that distributes power to buildings, as shown in Fig. 1.1.)

We address these three challenges using a variety of methods. To design control algorithms for tracking a balancing signal, we use model-based control and estimation methods. To analyze effects on distribution networks, we run detailed power flow simulations. To design aggregator-operator coordination strategies, we draw upon policy and industry knowledge as well as hierarchical control concepts. To design control algorithms that guarantee safety, we apply correct-by-construction control techniques. Finally, to determine a global safety constraint for a distribution network, we employ optimization-based methods.

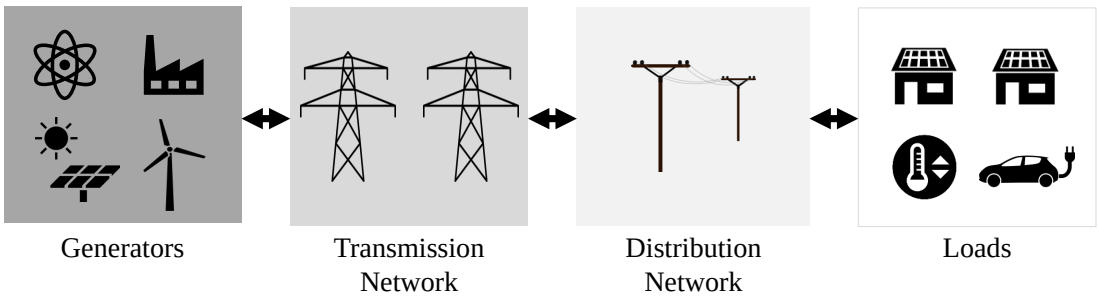


Figure 1.1: Structure of the power system.

1.2 Background

This section provides relevant background information on power system operation, thermostatically controlled loads, and policies for wholesale electricity markets.

1.2.1 Power System Operations

Figure 1.1 shows the general structure of the power system: large generators supply power to the transmission network; the high-voltage transmission network transmits the power, sometimes over hundreds of miles, to the lower voltage distribution network; and the distribution network distributes power to loads, which are devices that consume power (e.g., electronics, home appliances, and lighting). Increasingly, resources capable of supplying power are connecting to the distribution network, such as roof-top photovoltaic (PV) systems and home battery systems. Given high penetrations of these resources, it is possible for power to flow in the reverse direction (i.e., from distribution to transmission) during certain hours of the day.

Power system operation is effectively divided into two parts: 1) transmission and wholesale market operation, and 2) distribution operation. In regions of the U.S. with a competitive electricity market, an independent system operator (ISO) or regional transmission organization (RTO) operates the transmission network and wholesale electricity market. Separate entities, referred to as distribution network operators, or distribution operators for short, operate the distribution networks.

One of the fundamental objectives of ISOs and RTOs is to ensure power system reliability (i.e., to avoid power outages). An important aspect of ensuring reliability is maintaining the system's energy balance. An imbalance in energy can cause the frequency of the system's voltage and current waveforms to become unstable, which can result in large-scale outages.

The wholesale electricity market has a variety of products that are designed to match supply and demand in an economically efficient manner. The day-ahead energy market and real-time energy market schedule dispatchable resources so that their net supply matches the forecasted net demand of non-dispatchable resources. Whereas the energy markets schedule resources to match forecasts, regulation (an ancillary service market product) retroactively corrects for errors in the energy balance. Regulation-providing resources are dispatched at 2 or 4 second intervals to correct for measured area-control error (ACE) which is a function of measured error in system frequency and in scheduled power flows between ISOs/RTOs. Regulation, as it is known in the U.S., is referred to as secondary frequency control in many other countries.

In this work, we use the term *energy balancing* to refer to fast changes in a resource's demand or supply to help balance energy at the transmission level. We consider fast changes to be on the time scale of 5 minutes to a few seconds. In market terms, energy balancing includes regulation, where changes occur every few seconds, and real-time energy markets, where changes occur every 5 minutes. We will also refer to these services as transmission-level services to distinguish them from distribution-level services that are designed to improve distribution operation.

1.2.2 Access to Wholesale Markets for Demand Resources

Demand response resources have not always been able to participate in wholesale electricity markets, and still face some barriers to participation today. The Federal Energy Regulatory Committee (FERC), which regulates interstate ISOs and RTOs, has made efforts to open the wholesale electricity markets to nontraditional resources such as load aggregations. Listed below are FERC orders that have made it easier for load aggregators, as well as other nontraditional resources, to participate in the wholesale markets and to be fairly compensated for energy balancing services.

- FERC Order No. 719: In the ancillary service market, ISOs/RTOs must compensate demand response resources at the market price. Furthermore, they must allow aggregators of retail consumers to participate as demand response providers in wholesale markets [87].
- FERC Order No. 745: In the wholesale day-ahead and real-time energy markets, ISOs/RTOs must compensate demand response providers at the market price, referred to as the locational marginal price (LMP) [88].

- FERC Order No. 755: ISOs/RTOs must compensate frequency regulation resources based on their performance, i.e., the accuracy with which they follow the regulation signal [89].
- FERC Order No. 841: ISOs/RTOs must open up their wholesale markets (capacity, energy, and ancillary services) to participation by electric storage [90]. This order specifically does not address aggregations of electric storage resources.
- FERC Notice of Proposed Rulemaking (November 17, 2016): In this notice, FERC proposed rules to facilitate the participation of aggregations of distributed energy resources (DERs) in the wholesale markets [86], but these rules have been contentious and have not yet been finalized. (DER aggregations differ from aggregations of demand resources in that DER aggregations can both receive power from and inject power into the power system.)

1.3 Literature Review and Research Gaps

In this section, a review of the literature on load aggregation and control for energy balancing is presented, and research gaps are identified.

1.3.1 Nondisruptive, High-Performance Load Control

A TCL aggregation must be controlled to provide high performance energy balancing without inconveniencing the end-users. Providing energy balancing with small, distributed loads is a challenge in and of itself: loads, whose parameters are not fully known, must be coordinated such that their aggregate power consumption matches a fast-moving energy balancing signal. In addition, each load has a service it provides to the end-user that should not be compromised. For example, an air conditioner cannot be kept on or off for too long before occupants become uncomfortable. Ensuring that end-users are not disrupted is considered crucial for user acceptance [8] but also adds constraints, and therefore complexity, to the control problem.

The disruptiveness of a program depends, in part, on the mechanism used to control TCLs. A TCL can be controlled either by adjusting its temperature setpoint (e.g., [3], [40], [96]) or by directly switching it on/off (e.g., [28], [52]). Adjusting a consumer's setpoint may be perceived as inconvenient, even if the change in setpoint (and effect on user comfort) is small.

In contrast, on/off switching can be implemented while respecting user-set temperature constraints.

Another way in which load control may inconvenience users is from an increase in the frequency of on/off cycling. Increased cycling may reduce the lifetime of the loads due to strain on mechanical components, and may disrupt users due to changes in noise upon turn-on and turn-off. In [52] and [28], different methods are proposed for switching TCLs in terms of temperature priority, which indirectly increases an aggregation's minimum cycling time compared to other control methods. More recently, researchers have been including explicit constraints that ensure a minimum on-time and off-time for each TCL [14] [99].

A TCL's individual temperature and cycling constraints can make a TCL unresponsive to external control. If enough TCLs become unresponsive, the tracking performance of the aggregation can suffer; this phenomenon is referred to as saturation. To predict and avoid saturation, aggregate models of TCLs' flexibility, referred to as virtual battery models, have been proposed. The virtual battery models proposed in [28], [49], and [83] account for TCLs' temperature constraints only; the more recent models of [14] and [99] account for TCLs' cycling constraints in addition to temperature constraints.

The research community is in need of better data on TCLs' actual minimum on- and off-time constraints. Recent work assumes TCLs require both minimum on-time and off-time constraints to protect physical components, but this may not be the case. For example, compressors have a minimum off-time but not a minimum on-time [67]. Research is also needed on how consumers perceive increased cycling. Does the noise or sensation of TCLs switching on/off bother consumers? Is this effect worse for certain types of TCLs (e.g., forced air versus resistive heating)? And how much might consumers need to be reimbursed to compensate for this inconvenience?

1.3.2 Performance in the Real World

Another important topic is the development of load control strategies that will perform well when subject to real-world disturbances and changing conditions. In the literature, it is common to test control methods on an aggregation of simulated loads (referred to as the simulated plant) rather than an aggregation of actual loads because of the prohibitive costs of doing the latter. In many cases, the simulated plant is substantially simplified; it is unclear if control methods tested in such simplified environments would perform well when implemented on actual loads.

In the literature, control strategies have not been adequately designed for, or tested on, loads with realistic disturbances. As an example of a disturbance, consider a cooled house: if doors to the outside are opened or the number of occupants increases, the house's temperature will deviate from that of a simple thermal model. In much of the literature, disturbances are modeled in the simulated plant by adding a stationary, zero-mean Gaussian noise term to each TCL's temperature dynamics (e.g., [52] and [28]). Instead of this arbitrary noise term, more accurate models of disturbances would be based on real load data. A statistical model of disturbances to refrigerators was developed in [37] based on real data; similar models are needed for other types of TCLs, and in particular space heating/cooling systems. In addition, other types of disturbances should also be considered, such as the voltage dependence of TCLs' power consumption.

In model-based control strategies, the controller's model of the plant must be identified. Typically, this identification is done empirically by observing the simulated plant under ideal, constant conditions; identification can also be done analytically, as proposed in [52] and [60]. Additional research is needed on identifying models of physical plants for varying conditions. For example, it is known that [52]'s aggregate model of air conditioners must be re-identified for different outdoor temperatures. But what other changing conditions could cause the aggregate model to deviate from actual behavior? One might expect humidity and incident solar radiation to have large effects.

Finally, real control systems also experience delay (due to communication and computation) and loss of messages. Control algorithms designed to perform well given these disturbances are proposed in [42] and [95].

1.3.3 Impacts on Distribution Networks

Most research has neglected the effect that aggregator-controlled loads can have on distribution networks. By definition, aggregators partially synchronize loads to track energy balancing signals. On a distribution network with a high concentration of aggregator-controlled loads, it is possible for this synchronization to cause operational issues such as under- or over-voltages. Separately, distribution operation is becoming more challenging due to increasing concentrations of PV systems [15] and uncoordinated EVs [43]. On such networks, the provision of energy balancing with load aggregations could exacerbate operational issues.

When the load aggregator is also the distribution operator, the problem—controlling loads to provide energy balancing while respecting distribution constraints—can be solved in a

centralized manner as a constrained optimization problem, as in [93]. Or the problem can be solved in a distributed manner through the use of Lagrange multipliers, as in [17] and [30].

More research is needed for the case in which the aggregator is a third-party, i.e., distinct from the distribution operator. This problem is different because third-party aggregators and distribution operators have private information that they may be unwilling to share with one another. The operator may want to keep its network parameters private and would need to keep its load data private for consumer privacy. A third-party aggregator will likely want to keep its proprietary control algorithm private.

The impacts of load-control on distribution networks as well as the challenges posed by third-party aggregators have not been studied in detail; specific related research questions include:

- What are the effects on the distribution network when a large portion of the network's loads are aggregated and controlled to provide transmission-level energy balancing?
- If only a few distribution constraints are at risk of violation due to energy balancing, can these constraints be satisfied using a targeted intervention (rather than, for example, solving a centralized optimization problem)?
- How can a distribution network operator ensure third-party aggregators do not cause unsafe operation without overly restricting aggregators' actions?
- What are the benefits of coordination (i.e., information sharing) between an operator and third-party aggregator?

1.3.4 Economic Viability

Another challenge is making the technology economically viable. When providing energy balancing with a load aggregation, the revenue per load is low, so the costs per load must also be low. In [50], the potential revenue for different types of TCLs participating in California's regulation market was estimated; the study found a mean revenue of \$24/year for a water heater, \$0-32/year for an air conditioner, and \$22-56/year for a heat pump, where the ranges reflect different climate zones in California. Because these revenues are small, the annualized per-load costs must be kept low in order to make the technology economically attractive to consumers and aggregators.

One way to reduce costs is to reduce the cost of infrastructure required to implement load control. This infrastructure includes communication, sensing, and computation. Researchers have proposed estimation-based methods to reduce the number of load measurements (e.g.,

temperature and power measurements) that are sensed and communicated (see [6], [9], [41], and [52]). Although these papers have shown that an aggregation can perform well with fewer measurements, ISOs'/RTOs' metering and telemetry requirements may not allow such reductions [46].

In this setting, the biggest economic improvements are likely to come from the development of new ISO/RTO rules. To improve their economic viability, aggregators may need to propose changes to ISO/RTO metering and telemetry requirements that will help them reduce the associated costs without diminishing the quality of the provided services.

1.3.5 Adapting to New Technologies

Most proposed strategies that aggregate and control space heating/cooling systems assume relatively simple systems (e.g., systems with single-speed compressors and single-zone temperature control). More complex systems—with two-speed or variable-speed compressors, occupant sensing, and multi-zone temperature control—are growing in popularity. Control strategies should be developed that are compatible with the constraints of these more complex systems.

1.3.6 Gaps Addressed by the Dissertation

This dissertation addresses many, but not all, of the aforementioned research areas and gaps. Chapter 2 addresses some of the real-world performance concerns raised in Section 1.3.2. The majority of the dissertation (Chapters 3-7) is devoted to the gaps identified in Section 1.3.3, i.e., assessing and preventing the negative impacts of load control on the distribution network. Throughout the dissertation, TCL controllers are designed to be non-disruptive and high performance, as described in Section 1.3.1. In addition, the controllers proposed in Chapters 5 and 6 include TCL cycling constraints and contribute to the recent literature on control solutions that have minimum cycle-time guarantees.

1.4 Organization and Contributions of the Dissertation

Chapter 2 proposes two online methods for updating aggregate models of TCLs when conditions (e.g., temperatures) are changing or when biased disturbances are active. The proposed methods are incorporated into a model-based control system and are shown to improve the control system's tracking performance compared to two benchmark methods.

Chapter 3 presents a simulation study of distribution networks with high proportions of aggregator-controlled loads. The goal of the study is to determine what negative effects aggregated loads (controlled to provide energy balancing) can have on distribution operation.

Chapter 4 presents a hierarchical control strategy for providing network-safe load control with a third-party aggregator and distribution operator. In this strategy, the operator blocks TCLs from receiving the aggregator's commands when necessary to ensure safe network operation. An estimation algorithm for the aggregator is proposed that estimates the percentage of TCLs that are blocked; this estimate is used within the aggregator's tracking controller to compensate for blocking. In this chapter, a simplified plant model is used—that is, TCLs' cycling constraints are neglected and the distribution network is not modeled. These simplifications are removed in Chapter 6.

Chapter 5 presents mode-count control, a control algorithm designed to satisfy constraints on the power consumption of groups of TCLs. The algorithm can be used in a targeted fashion to relieve particular constraints on a distribution network. We demonstrate the algorithm with a group of 25 TCLs and show that it can reduce local voltage variability. Finally, we extend the algorithm to account for TCLs' cycling constraints.

Chapter 6 presents two control strategies for providing network-safe load control with a third-party aggregator and distribution operator. In the first strategy, the distribution operator blocks TCLs in certain areas of the network to ensure safety (as in Chapter 4), and the aggregator uses an aggregate-model based controller for tracking. A new aggregate model is proposed for TCLs with cycling constraints. In the second strategy, the operator uses targeted mode-count control to ensure network safety, and the aggregator uses a priority-stack controller that relies on individual TCL models. The two strategies are compared in terms of performance as well as their measurement, communication, and computational requirements.

Chapter 7 presents a method of constraining an aggregator's control actions such that they do not cause violations of distribution network constraints. We design a safety constraint as a function of aggregator-induced power deviations, and formulate a set of optimization problems to determine the safety constraint's limit.

Chapter 8 concludes the dissertation and discusses future avenues of investigation.

Note that the each chapter is self-contained in terms of mathematical notation except when an equation from a different chapter is explicitly referenced.

Chapter 2

Load Control to Reject Biased Disturbances

This chapter is largely based on the published work [16][1].

2.1 Chapter Introduction

Power system operators are looking to new sources of flexibility to support the growing number of wind and solar installations on the grid [27, 48]. The power production of these renewable energy resources is intermittent and uncertain, requiring flexible generators and loads to quickly change their production and consumption to ensure real-time energy supply/demand balance. Thermostatically controlled loads (TCLs), such as air conditioners, heat pumps, and electric water heaters, are a vast, flexible resource [50] because the power consumption of these sorts of loads can be manipulated without substantially impacting their end use function [8]. Residential TCLs generally operate within a temperature deadband, i.e., the consumer sets the temperature set point and the load modulates temperature to plus/minus some small range around that set point by cycling the load on and off. Non-disruptive load control techniques exploit inherent TCL flexibility by changing the timing of TCL on/off switches while ensuring that temperatures remain within the deadband; see e.g., [52, 94]. Coordinating the switching of hundreds to thousands of TCLs enables them to provide services to the power system such as regulation or load following.

[1] S. Crocker and J.L Mathieu. "Adaptive state estimation and control of thermostatic loads for real-time energy balancing". In: *Proceedings of the American Control Conference (ACC)*. Boston, MA, July 2016, (Invited).

A key challenge of controlling a TCL aggregation is maintaining accurate tracking performance when TCLs are subject to disturbances and changing conditions. Most related work assumes that noise affects each TCL's temperature dynamics, and that this noise is an independent identically distributed, zero-mean Gaussian, stationary random process (e.g., [28, 52]). This noise is meant to represent stochastic processes that increase/decrease the internal temperature of a load, such as steam from a shower warming a house. In practice, we would expect some disturbances to depend on the time of day and to have a bias. For example, these types of disturbances might occur as a result of cyclic human behaviors, such as everyone cooking dinner at approximately the same time of day. Some research has considered the effect of changing ambient temperatures on TCL control [51]; this is a changing condition that we will refer to as a disturbance in this chapter.

The contribution of this chapter is the development of two online methods for updating an aggregate TCL model to account for biased disturbances. The first system uses infrequent state measurements to update the aggregate model, and the second uses aggregate power measurements. We add these model-update methods to the model-based control system of [52] and compare their performance to two benchmark methods.

2.2 Modeling of Thermostatically Controlled Loads

2.2.1 Individual TCL Model

A model of a house's temperature dynamics was developed in [34, 58, 79] and has been frequently used in the literature [7, 19, 74]. It is a lumped-element model and can be represented by the equivalent electrical circuit shown in Fig. 2.1. In this model, the interior temperature of the house is θ, the ambient outdoor temperature is θ_{amb}, the heat capacity of the house's interior is c, and the resistance of the house's thermal envelope is r. When heat is actively being transferred in or out of the house, the conditioning device's power status σ is *on*, otherwise it is *off*. If $p_\theta < 0$ the house is being cooled, if $p_\theta > 0$ the house is being heated. Given this model, the ith TCL's temperature θ^i evolves according to the hybrid dynamics,

$$\frac{\mathrm{d}}{\mathrm{d}t}\theta^i(t) = \begin{cases} \frac{1}{\tau^i}(\theta_{\text{amb}}(t) - \theta^i(t) + p_\theta^i r^i), & \text{if } \sigma^i(t) = \textit{on}, \\ \frac{1}{\tau^i}(\theta_{\text{amb}}(t) - \theta^i(t)), & \text{if } \sigma^i(t) = \textit{off}. \end{cases} \tag{2.1}$$

where $\tau^i = r^i c^i$ is the thermal time constant of the house.

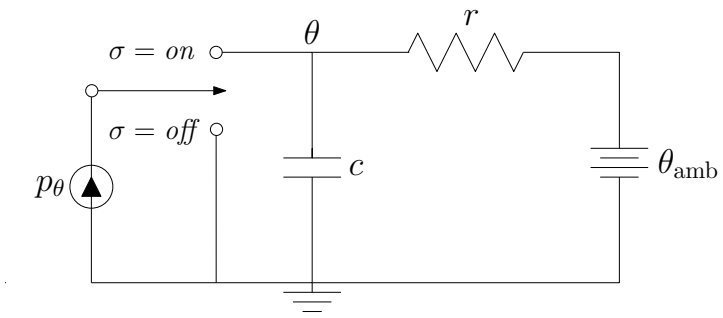


Figure 2.1: Electrical circuit equivalent of the thermal of an individual TCL. Figure based on [58].

In this chapter, we use a discrete time model of a TCL's temperature dynamics. The model, derived in [58], is just the discretized solution to (2.1) with a time step duration of h. Using the discrete model, the temperature of the ith TCL at time step $k+1$ is given by

$$\theta^i(k+1) = \begin{cases} a^i\theta^i(k) + (1-a^i)(\theta_{\text{amb}}(k) + r^i p_\theta^i), & \text{if } \sigma^i(k) = on \\ a^i\theta^i(k) + (1-a^i)\theta_{\text{amb}}(k), & \text{if } \sigma^i(k) = off, \end{cases} \tag{2.2}$$

where $a^i = \exp(-h/(r^i\, c^i))$.

In this dissertation, we restrict attention to TCLs with a fixed heat transfer rate p_θ. For this type of TCL, the thermostat controls temperature by toggling the heating/cooling device *on* and *off*. For a cooling TCL, the thermostat's control is modeled as

$$\sigma^i(k) = \begin{cases} off, & \theta^i(k) < \underline{\theta}^i \\ on, & \theta^i(k) > \overline{\theta}^i \\ \sigma^i(k-1), & \text{otherwise.} \end{cases} \tag{2.3}$$

where $\underline{\theta}^i = \theta^i_{\text{set}} - \delta^i/2$, $\overline{\theta}^i = \theta^i_{\text{set}} + \delta^i/2$, θ^i_{set} is the user-set temperature, and δ is the allowed range in temperature and is typically less than 1°C. The temperature range between $\underline{\theta}^i$ and $\overline{\theta}^i$ is referred to as the temperature deadband, or simply as the deadband.

The coupled set of equations (2.2) and (2.3) describe the trajectory of an autonomous TCL through its state space. Ranges of parameter values for residential air conditioning systems, sourced from [50, 52], are listed in Table 2.1.

Finally, the rated electrical power consumption of a TCL is given by

$$P^i_{\text{R}} = |p^i_\theta|/\zeta^i, \tag{2.4}$$

Table 2.1: Cooling House Parameters

Parameter	Range of Values	Unit
Setpoint temperature (θ_{set})	18 to 27	°C
Width of temperature range (δ)	0.25 to 1	°C
Thermal resistance (r)	1.2 to 2.5	°C/kW
Heat capacity of interior (c)	1.5 to 2.5	kWh/°C
Heat transfer rate (p_θ)	-18 to -10	kW
Coefficient of performance (ζ)	2.5	–

where ζ^i is the coefficient of performance. If there are N TCLs in the aggregation, their total power consumption is given by

$$P_{\text{total}}(k) = \sum_{i=1}^{N} \mathbb{1}_{\text{on}}\Big(\sigma^i(k)\Big) P_{\text{R}}^i. \tag{2.5}$$

We use the notation $\mathbb{1}_{\{a\}}(b)$ to represent the indicator function of the singleton set containing a: if $b = a$ the function has value 1 and if $b \neq a$ the function has value 0.

2.2.2 Aggregate TCL Model

Within the proposed control systems, we use a linear time-invariant (LTI) model to describe the aggregate dynamics of the TCL population. The model, developed in [52] and similar to the models of [3, 40, 45], is of the standard LTI form:

$$\mathbf{x}(k+1) = \mathbf{A}\mathbf{x}(k) + \mathbf{B}\mathbf{u}_{\text{model}}(k) \tag{2.6}$$

$$y(k) = \mathbf{C}\mathbf{x}(k). \tag{2.7}$$

This model makes the control problem more tractable. We briefly describe the derivation of this model but refer the reader to [52] for full details.

First, each TCL's temperature deadband is normalized so that the state of all TCLs can be described by the same deadband. Then, the normalized deadband is discretized by uniformly dividing it into $N_{\text{B}}/2$ intervals. A TCL's state is described by which "bin" it is in, where each bin corresponds to a distinct temperature interval and power status (on or off). There are a total of N_{B} bins. A TCL's dynamics can be approximately model with a Markov chain, where transitions from bin to bin are modeled probabilistically. Fig. 2.2 illustrates

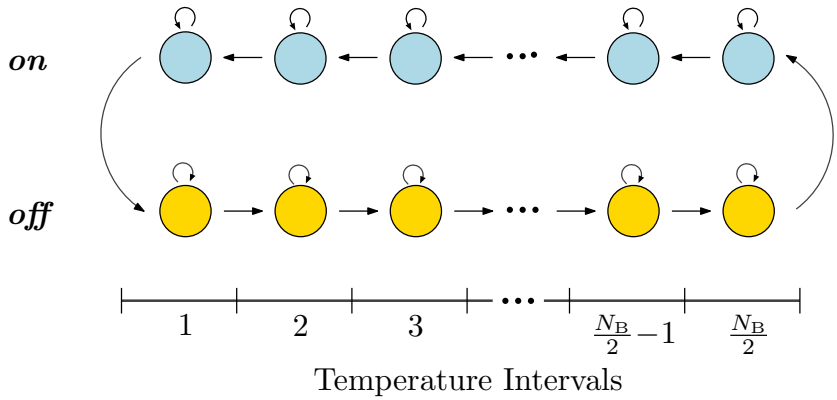


Figure 2.2: Representation of TCL state dynamics as a Markov chain. Bins (indicated by circles) are defined by a temperature interval and power status (*on* or *off*). A subset of possible bin transitions are shown with arrows.

the Markov chain; bins are represented by circles and transitions by arrows.

The aggregate state $\mathbf{x}$ represents the fraction of the population that is in each bin. The $\mathbf{A}$-matrix governs the autonomous dynamics of $\mathbf{x}$ and is the transpose of the Markov chain's transition matrix. Let $p_{i,j}$ be the probability of transitioning from bin i to bin j, then the $\mathbf{A}$ is given by

$$\mathbf{A} = \begin{bmatrix} p_{1,1} & p_{2,1} & \cdots & p_{N_B,1} \\ p_{1,2} & p_{2,2} & \cdots & p_{N_B,2} \\ \vdots & \vdots & \ddots & \vdots \\ p_{1,N_B} & p_{2,N_B} & \cdots & p_{N_B,N_B} \end{bmatrix}. \tag{2.8}$$

Assuming 40 bins and $h = 2$ seconds, as we do in this chapter, the transition probabilities above the diagonal, and below the lower off-diagonal are very small or zero.

The control input vector $\mathbf{u}_{\text{model}}$, which is half the length of $\mathbf{x}$, specifies the fraction of the TCL population that is designated to switch from each bin. Positive entries of $\mathbf{u}_{\text{model}}$ represent commands to switch *on* and negative entries represent commands to switch *off*. The $\mathbf{B}$-matrix is given by

$$\mathbf{B} = \left[\begin{array}{ccc} -1 & & 0 \\ & \ddots & \\ 0 & & -1 \\ \hline 0 & & 1 \\ & \iddots & \\ 1 & & 0 \end{array}\right]. \tag{2.9}$$

The structure of $\mathbf{B}$ ensures that the fraction that switches out of an *on* bin is moved into the corresponding *off* bin of the same temperature.

The output y represents the total power consumption of the TCLs. The $\mathbf{C}$-matrix is defined as

$$\mathbf{C} = \bar{P}_{\text{on}} N \, [\underbrace{0, \cdots, 0}_{\frac{N_{\text{B}}}{2}}, \, \underbrace{1, \cdots, 1}_{\frac{N_{\text{B}}}{2}}], \tag{2.10}$$

where the binary vector sums the fraction of the TCLs that are *on* and the scaling term converts this fraction into the approximate power consumption of all TCLs that are *on*. It is approximate because $\bar{P}_{\text{on}}$ is a constant parameter that estimates the average power consumption of all TCLs that are *on* in the current time step; in practice, this value is time-varying.

2.3 Original Control System

A look-ahead proportional controller with a Kalman filter for estimation was developed in [52]. In this chapter, we refer to this control system as the *original control system*, because it serves as the base for the systems that we propose in Section 2.4.

The block diagram of the original control system is shown in Fig. 2.3. The power signal to be tracked is P^*_{total}. The controller computes the control input $\mathbf{u}_{\text{plant}}$ which is sent to the plant. The plant consists of a population of TCLs that switch *on*/*off* based on the control input. In aggregate, the TCLs consume real power with magnitude P_{total}, which is supplied by the distribution substation. The substation also supplies power to other loads. Two forms of feedback are shown in Fig. 2.3: output feedback $P_{\text{total,meas}}$ and state feedback $\mathbf{\Theta}$, $\mathbf{\Sigma}$. The output feedback is determined from the substation's measurement of the total power consumption of its distribution system P_{DS} and a forecast of the power consumption of 'other' loads (i.e., loads not in the plant) $\hat{P}_{\text{other,forecast}}$. The forecast is subtracted from P_{DS} to determine $P_{\text{total,meas}}$, a noisy measurement of the TCLs' total power consumption. State feedback consists of TCLs' temperature histories $\mathbf{\Theta}(k) = \{\boldsymbol{\theta}(k), \boldsymbol{\theta}(k-1), ..., \boldsymbol{\theta}(k-j)\}$ and *on*/*off* histories $\mathbf{\Sigma}(k) = \{\boldsymbol{\sigma}(k), \boldsymbol{\sigma}(k-1), ..., \boldsymbol{\sigma}(k-j)\}$. This feedback is used within a state estimator, which provides the state estimate $\hat{\mathbf{x}}$ to the controller.

Note the dashed line indicates that the state histories are transmitted periodically. For example, this data could be sent from a smart meter that communicates with the TCL through a home energy management system. Because of communication protocols, smart meters are generally unable to send high-frequency data, but can collect data and transmit

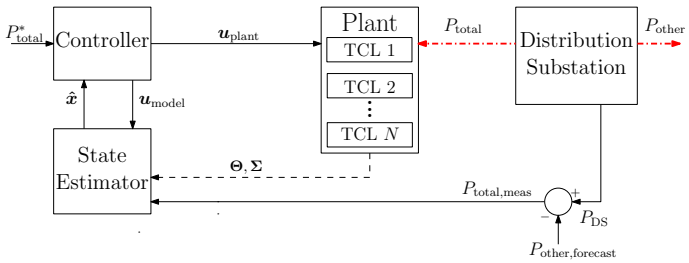


Figure 2.3: Block diagram of the original control system and plant with output feedback. Red dash-dot lines indicate physical power flows.

packets of data on timescales of minutes to hours [1]. When available, state feedback is used to directly update the state estimate.

2.3.1 Controller

The controller first predicts what the output would be one time step ahead if the TCLs were allowed to move through their deadbands unforced:

$$P_{\text{total,pred}}(k+1) = \mathbf{CA}\hat{\mathbf{x}}(k) \tag{2.11}$$

$P_{\text{total,pred}}$ is used to compute the fraction of TCLs that need to be switched on or off to achieve P^*_{total} in the next time step. This fraction is divided between the bins; here we divide it equally among all bins excluding those on the edges of the temperature band:

$$u^j_{\text{model}}(k) = \frac{1}{(N_{\text{B}}/2-2)} \times \frac{K_{\text{P}}\big(P^*_{\text{total}}(k+1) - P_{\text{total,pred}}(k+1)\big)}{\bar{P}_{\text{on}} N} \tag{2.12}$$

where $j \in \{2, ..., (N_{\text{B}}/2-1)\}$ and u^j_{model} is the j^{th} entry of $\mathbf{u}_{\text{model}}$. Note the first and last elements of $\mathbf{u}_{\text{model}}$ are set to zero. K_{P} is a control gain; if the prediction is perfect, then the optimal value of K_{P} is 1.

Before the control input is sent to the plant, it is converted from fractions based on the total number of TCLs to fractions based on the number of TCLs in the bins that the TCLs are being switched out of. These "switching probabilities" are computed by dividing the entries of $\mathbf{u}_{\text{model}}(k)$ by the associated entries of $\hat{\mathbf{x}}(k)$. The resulting vector $\mathbf{u}_{\text{plant}}(k)$ contains a switching probability for each bin being switched out of. Each TCL receives the vector, determines which switching probability to act upon based on its current state, then draws a random number between zero and one and switches state if its random number is less than

its switching probability.

2.3.2 State Estimator

In [52], a Kalman filter was used for state estimation and was designed for the system:

$$\mathbf{x}(k+1) = \mathbf{A}\mathbf{x}(k) + \mathbf{B}\mathbf{u}_{\text{model}}(k) + \mathbf{w}(k) \tag{2.13}$$

$$y(k) = \mathbf{C}\mathbf{x}(k) + v(k) \tag{2.14}$$

where $\mathbf{w}(k)$ is process noise, $v(k)$ is measurement noise, and the pair $(\mathbf{A}, \mathbf{C})$ is observable [52]. The process noise is due to plant-model mismatch as well as individual TCL noise. Measurement noise arises because of our assumption that the output measurement, $y(k)$ (in Fig. 2.3 labeled $P_{\text{total,meas}}$), is obtained by subtracting a forecast of the power consumption of uncontrolled loads from the total substation power. The measurement noise is thus equal to the error in the forecast of uncontrolled loads. We assume this forecast error is Gaussian with mean zero. The process noise is inherent to the model-plant mismatch and is not zero-mean or Gaussian. Thus, our use of a Kalman filter is suboptimal. We recognize that in practice the measurement noise is also unlikely to be zero-mean and Gaussian, and the application of a Kalman filter would thus be even less optimal.

In our implementation of the Kalman filter, we assume the output measurement can be used immediately for estimation as in [52]. Thus if the current time step is k, the algorithm updates the state estimate to the current time step by calculating the *a priori* state estimate:

$$\hat{\mathbf{x}}(k|k-1) = \mathbf{A}\hat{\mathbf{x}}(k-1|k-1) + \mathbf{B}\mathbf{u}_{\text{model}}(k-1) \tag{2.15}$$

This estimate is then corrected using $\mathbf{M}$, the Kalman filter's innovation gain, which weights how much the output measurement influences the state estimate correction. The algorithm calculates the *a posteriori* state estimate using:

$$\hat{\mathbf{x}}(k|k) = \hat{\mathbf{x}}(k|k-1) + \mathbf{M}(P_{\text{total,meas}}(k) - \mathbf{C}\hat{\mathbf{x}}(k|k-1)) \tag{2.16}$$

Once calculated, $\hat{\mathbf{x}}(k|k)$ is substituted for $\hat{\mathbf{x}}(k)$ in (2.11).

To calculate $\mathbf{M}$ we use the *kalman* function in MATLAB and provide as inputs the covariance matrix $\mathbf{W}$ of $\mathbf{w}$ and the variance V of v. $\mathbf{W}$ is calculated by first characterizing $\mathbf{w}$ through a simulation of the controlled plant. In this simulation, time series data is collected

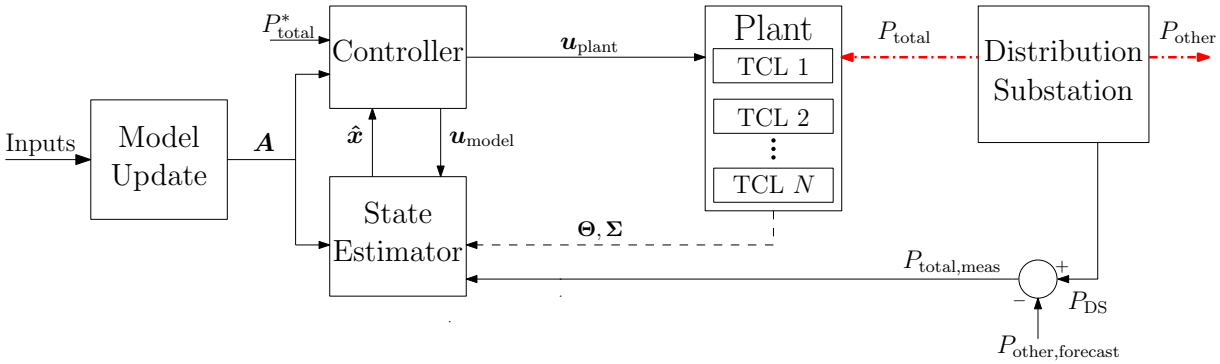


Figure 2.4: Block diagram of the original control system with model updates.

on the error between the actual states and the modeled states; this error can be thought of as samples of the random vector $\mathbf{w}$. V is calculated simply by squaring the standard deviation of v, which is predefined. In contrast to [52], we treat the magnitude of $\mathbf{M}$ as a tunable parameter; we scale it by K_{M} when searching for the optimal combination of tuning parameters for the control systems considered in the case studies.

2.4 Methods for Bias Rejection

We propose two methods for rejecting biased disturbances. Both methods update the LTI model to improve controller and estimator accuracy. The model updates are added to the control system as shown in Fig. 2.4. We also present two benchmark methods that will be used for comparison in the case study.

2.4.1 Proposed Methods

State-Based Update

In this method, we propose updating $\mathbf{A}$ using state histories $\boldsymbol{\Theta}$ and $\boldsymbol{\Sigma}$. When a state history arrives, we compute TCLs' current transition rates between bins. Then the updated entries of $\mathbf{A}$ are computed as the weighted sum of the previous $\mathbf{A}$ and the current bin transition rates. The weighting parameter K_{wt} takes a value between 0 and 1. The optimal value for K_{wt} is found through iterative tuning.

Output-Based Update

In this method, we update $\mathbf{A}$ as a function of estimated output error: $(P^*_{\text{total}} - \hat{P}_{\text{total}})$. The idea for the output-based update originated from a simple observation about the original control system: if output error has a consistent bias, then the control system is consistently under- or over-predicting the number of TCLs that will be *on* in the next time step. This bias can come from several sources of inaccuracy: 1) the $\mathbf{A}$-matrix does not capture the current rates of state transitions, 2) the state estimate has deviated substantially from the actual state, and 3) the TCL aggregation is unable to follow the target trajectory because all TCLs are operating at the limits of their temperature ranges. One way to address the first issue is to update $\mathbf{A}$ as a function of the integrated output error.

For instance, if the integrated output error is negative, this indicates that too many TCLs have been switched *on*, which means the control inputs have been too large. The control input can be too large if the model-based prediction, $P_{\text{total,pred}}(k) = \mathbf{CAx}(k)$, is too small. To compensate for this under-prediction, we should increase the number of TCLs predicted to be in *on*-states by decreasing the probability of transitioning from *on*-state to *on*-state in the $\mathbf{A}$-matrix. (A change of the opposite sign should be applied to the transition probabilities corresponding to *off*-states.) This update slows the modeled progression of the TCLs in the *on*-states and speeds up the progression in the *off*-states, resulting in the desired increase to *on*-state entries. In addition, these changes to the $\mathbf{A}$-matrix will result in an increase to the *on*-states for the state estimate, $\hat{\mathbf{x}}$, because the Kalman filter's *a priori* estimate depends on the $\mathbf{A}$-matrix as well.

We propose updating transition probabilities with an additive term that is proportional to the integrated output error. For a four-bin LTI model, the update is given by

$$\mathbf{A}(k) = \begin{bmatrix} p_{1,1} + \alpha(k) & p_{2,1} & p_{3,1} & p_{4,1} + \alpha(k) \\ p_{1,2} - \alpha(k) & p_{2,2} + \alpha(k) & p_{3,2} & p_{4,2} \\ p_{1,3} & p_{2,3} - \alpha(k) & p_{3,3} - \alpha(k) & p_{4,3} \\ p_{1,4} & p_{2,4} & p_{3,4} + \alpha(k) & p_{4,4} - \alpha(k) \end{bmatrix}. \tag{2.17}$$

The additive term $\alpha(k) = K_\alpha e_{\text{I}}(k)$, where K_α is a tunable gain and the integrated error is

$$e_{\text{I}}(k) = \sum_{j=1}^{k} (P^*_{\text{total}}(j) - \hat{P}_{\text{total}}(j)). \tag{2.18}$$

In (2.17), we update only the dominant transition probabilities because all other entries are

typically orders of magnitude smaller and have a negligible effect on model dynamics.

Analyzing the stability of the control system when using the output-based update is future work. It should be noted that, by substituting (2.17) into (2.12), the control law can be reformulated to have an integrator term. This suggests that output-based update may be equivalent to an integrator.

2.4.2 Benchmark Methods

Temperature-Based Update

The first benchmark method updates the aggregate model to account for changing temperatures. We update $\mathbf{A}$ and $\bar{P}_{\text{on}}$ using lookup tables for the parameter values at different temperatures, similar to the approach of [51]. Specifically, we find the two values in the lookup table that are associated with the closest temperatures that bound θ_{amb}, and linearly interpolate between them to generate updated values. The updated parameters replace the originals in both the estimation and control algorithms. Note that this approach differs slightly from that of [51], which does not use linear interpolation.

Integrator

The second benchmark method is the addition of an integrator to the original system's control law. Specifically, we subtract the term $K_{\text{I}}e_{\text{I}}$ from the numerator of (2.12), where K_{I} is the integral gain and e_{I} is calculated as in (2.18). When state feedback is available, we use measured output error within (2.18) rather than estimated output error.

2.5 Case Study

2.5.1 Disturbance Models

If ambient temperature has been forecasted poorly, or if ambient temperature is assumed constant over the course of an hour, a rapid change in ambient temperature will act as a disturbance to the original control system. We model this type of disturbance as a linear change in ambient temperature over the course of an hour. The disturbance affects all house heating/cooling TCLs located in the same geographical area similarly. In the case study, we produce a temperature disturbance by varying θ_{amb} in (2.2) from 32°C to 35°C for each TCL

in the plant. When this disturbance is not active, ambient temperature is held constant at 32°C.

Another type of disturbance that can cause biased error is user activities that occur at similar times for TCLs across the aggregation. As an example, consider the hour after the workday ends: families return home, warm air enters cooled spaces when doors open, and internal heat gains increase due to occupant activity. We model this type of disturbance by adding a Bernoulli random variable θ_{d}^{i} to (2.2) for each TCL model in the plant. For the case study's simulation hour, the disturbance θ_{d}^{i} has distribution:

$$\mathrm{pr}\Big(\theta_{\mathrm{d}}^{i}(k) = 0.5\delta^{i}\Big) = \frac{1}{2000} \tag{2.19}$$

$$\mathrm{pr}\Big(\theta_{\mathrm{d}}^{i}(k) = 0\Big) = 1 - \frac{1}{2000}, \tag{2.20}$$

where δ^{i} is the TCL's temperature range. Note the probability of the disturbance occurring (1/2000) is the same for all TCLs. Over the course of a day, the probability of user disturbances would vary, but, for the case study's one hour time period, we assume the probability is constant. Given that there are 1800 time steps in the simulation hour, the expected number of disturbances a TCL will experience during the hour is $1800/2000 = 0.9$.

2.5.2 Setup

We compare the proposed methods to the benchmark methods in six cases. Cases are defined in Table 2.2 and are distinguished by which disturbance(s) are active and what feedback is available to the control system. In the table, output feedback indicates that $P_{\mathrm{total,meas}}$ is available in all time steps. State feedback indicates that $\boldsymbol{\Theta}$, $\boldsymbol{\Sigma}$ are available every 5 minutes. The output feedback $P_{\mathrm{total,meas}}$ is noisy due to the error in the forecast of the 'other' loads. We assume that this error has a zero-mean Gaussian distribution with standard deviation of 5% of the total substation power. As in [52] we assume a substation load of 17 MW.

The simulated plant parameters are chosen to be reflective of air conditioners. To produce a heterogeneous population of 1,000 air conditioners, we randomly draw instances of parameters for each load from the uniform distributions described by the ranges in Table 2.1, except for the setpoint range for which we use 15-25°C. The same parameter instances are used in all simulations. We set the simulation time step, h, equal to 2 seconds, which is sufficiently small given that a TCL's temperature cycle is on the order of 20 minutes. Finally, a noise term, ϵ^{i}, is added to the plant's individual TCL equations (2.2) to make the

simulated plant more realistic. We use the same Gaussian distribution for ϵ^i as [52] with zero mean and standard deviation of 5×10^{-4} °C.

In the LTI model, we use 40 bins. The **A**-matrix is identified by simulating the unforced plant over one hour and computing transition probabilities based on the state information. For the temperature-based update method, we calculate one **A**-matrix for each integer value from 32 to 35 °C. For the output and state based update methods, the **A**-matrix in the first time step corresponds to $\theta_{\text{amb}} = 32$ °C. For the **C**-matrix, we identify $\bar{P}_{\text{on}}$ by simulating the unforced plant and computing $\overline{P}_{\text{on}} = \sum_{k=1}^{1800} P_{\text{total}}(k) / \sum_{k=1}^{1800} \sum_{i=1}^{N} \mathbb{1}_{\text{on}}(\sigma^i(k))$.

We use the reference signal from [52], which was designed to represent a 5-min market signal from the California Independent System Operator. The signal P^*_{total} is composed of piecewise linear deviations (of up to $\pm25\%$ of the signal's offset). The signal's offset is the aggregation's baseline power consumption $P_{\text{total,bl}}$. We identify a constant value of $P_{\text{total,bl}}$ for each of the four ambient temperatures by simulating the unforced plant at each temperature and taking the average of P_{total}.

We evaluate control system performance by measuring the error between P^*_{total} and P_{total} at every time step, normalizing the values by $P_{\text{total,bl}}$, and taking the root mean square (RMS). Results are reported as RMS errors averaged over 25 simulation runs. All parameters and gains are held fixed across the 25 simulations. The simulation runs differ from each other in that the random elements in the simulation (ϵ, θ_{d}, and v) have different realizations.

For each case-method combination, we search iteratively for a set of optimal tuning parameters by simulating the forced plant and searching for parameters that minimized the system's tracking error. The range of optimal values found for the tuning parameters were $K_{\text{M}} \in [0.8, 2.2]$; $K_{\text{P}} \in [0.4, 1.4]$; $K_{\text{I}} \in [0.001, 0.5]$; $K_{\text{wt}} \in [0.7, 1]$; and $K_\alpha \in [3 \times 10^{-6}, 8 \times 10^{-6}]$.

2.5.3 Results

Table 2.2 shows the tracking error results for each of the methods across the six cases. As expected, percent RMS error rises as disturbances are added. For every case that has state feedback available, the state-based update has the lowest RMS error. However, when state feedback is not available, the state-based update cannot be implemented. Among the cases that have only output feedback, in Cases 2 and 6 the output-based update results in the lowest RMS error, and in Case 4 the temperature-based update results in the lowest RMS error.

Cases 1 and 2 serve as base cases, since they include no disturbances, and can be directly

Table 2.2: Tracking Performance: % RMS Error

Case			Method			
			Basis for Model Update			
Case	Disturbance	Feedback	State	Output	θ_{amb}	Integrator
1	–	State, Output	3.34	3.42	3.42	3.42
2	–	Output	–	4.15	4.65	4.53
3	A	State, Output	3.66	3.98	3.70	4.72
4	A	Output	–	4.59	4.33	7.55
5	A, U	State, Output	6.36	7.23	7.52	7.82
6	A, U	Output	–	7.19	9.44	10.55

A= Ambient temperature change, U = User disturbance

compared to the results of other research work. In Case 1, three of the methods have the same RMS error because none of them are able to improve upon the result of the original control system. In contrast, the state-based update does offer a small amount of improvement. This result suggests that the state-based updates may be beneficial even when there are only small changes in the plant's transition probabilities, such as those due to the changing characteristics of TCL clusters that move together through the temperature space. In Case 2, the integrator and output-based update outperform the temperature-based update (which, again, for this case is equivalent to the original control system because ambient temperature is constant).

In Cases 3 and 4, the ambient temperature disturbance is active. Setting aside the state-based update, in both cases the temperature-based update performs better than the output-based update and the integrator. We attribute this result to the fact that the temperature-based update is designed specifically for ambient temperature disturbances.

In Cases 5 and 6, both ambient temperature and user disturbances are active, and the state and output based updates outperform the temperature based update, which in turn outperforms the integrator. Fig. 2.5 shows the improved tracking of the control system with the output-based update in comparison to the temperature-based update in Case 6.

One counter-intuitive result is that the output-based update results in lower RMS error when it has no state feedback (Case 6) compared to when it has state feedback (Case 5). This may be due to the sudden corrections to $\hat{x}$ and $\hat{P}_{\text{total}}$ that occur when state histories arrive periodically in Case 5. These corrections induce a step change in the control input that has the potential to cause an over or undershoot in the plant output.

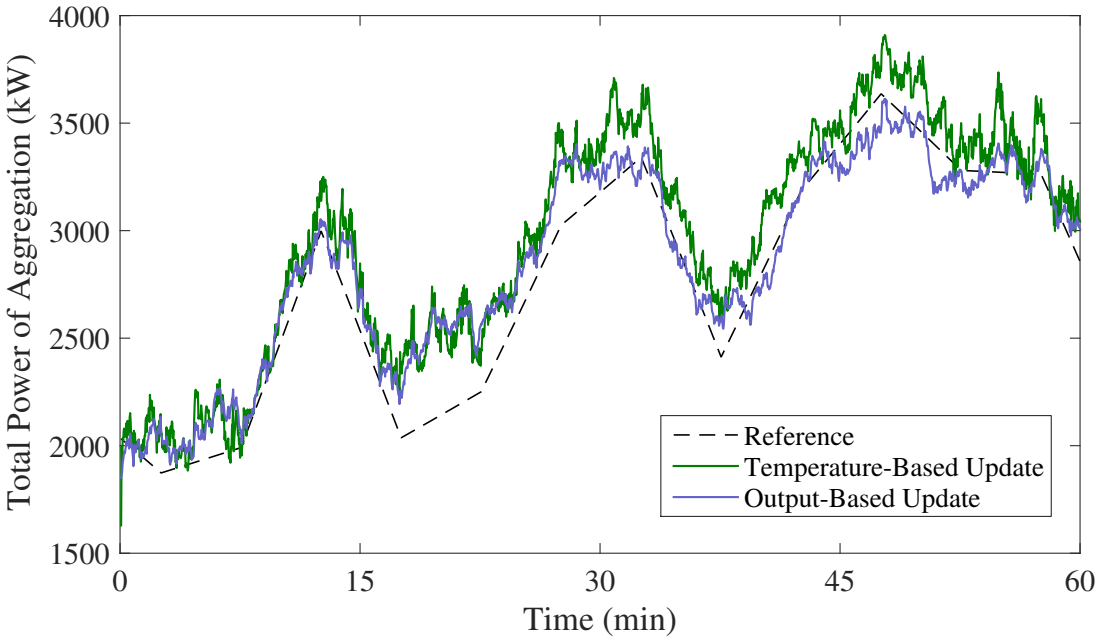


Figure 2.5: Tracking performance of the control system with two different model updates: temperature-based and output-based. Data is from Case 6. Percent RMS error is 9.44% for the temperature-based update and 7.19% for the output-based update.

2.6 Chapter Conclusion

We have developed control strategies that improve the tracking performance of large aggregations of TCLs when subject to biased, population-wide disturbances. We proposed two online methods of updating the state-bin model that is used within the model-based control system. These updates enable the control systems to partially reject biased disturbances and changing ambient temperature. In a case study, we found that the state-based model update outperformed all other methods, but this method is only an option when state measurements are available. When state feedback was unavailable, we found that the output-based update outperformed the benchmark methods in 2 of 3 cases. Regardless of the type of feedback available, the benefit of the proposed methods is clear when TCLs are subject to changes in ambient temperature and user-driven disturbances simultaneously.

In future work, we plan to develop disturbance models based on measured disturbance data and to conduct further testing using these more realistic disturbance models.

Chapter 3

Effects of Load-Based Energy Balancing on Distribution Network Operation

This chapter is largely based on the published work [73][2] and [69][3].

3.1 Chapter Introduction

Utilizing aggregations of loads for frequency regulation requires local distribution networks to transmit the service to the regional power system. As a result, distribution power flows will change, and load actions could cause local constraints, such as thermal limits of components or voltage limits of nodes, to become violated. Thus, although the load aggregator's objective is to provide a transmission-level ancillary service, the aggregator must also avoid violating distribution network constraints. This control problem is distinct from much of the work on control of distributed energy resources (DERs), which focuses on providing distribution-level services (e.g., [32, 98]).

In this chapter, we assess the impacts of load-based regulation on distribution network operation. Similar studies on network impacts have been conducted for DERs such as residential photovoltaic (PV) systems [15, 36], and electric vehicles (EVs) [13, 22, 31, 81], but

[2]S.C. Ross, G. Vuylsteke, and J.L Mathieu. "Effects of load-based frequency regulation on distribution network operation". In: *IEEE Transactions on Power Systems* 34.2 (2019), pp. 1569-1578.

[3]S.C. Ross, G. Vuylsteke, and J.L Mathieu. "Effects of load control for real-time energy balancing on distribution network constraints". In: *Proceedings of the IEEE Power & Energy Society PowerTech*. June 2017.

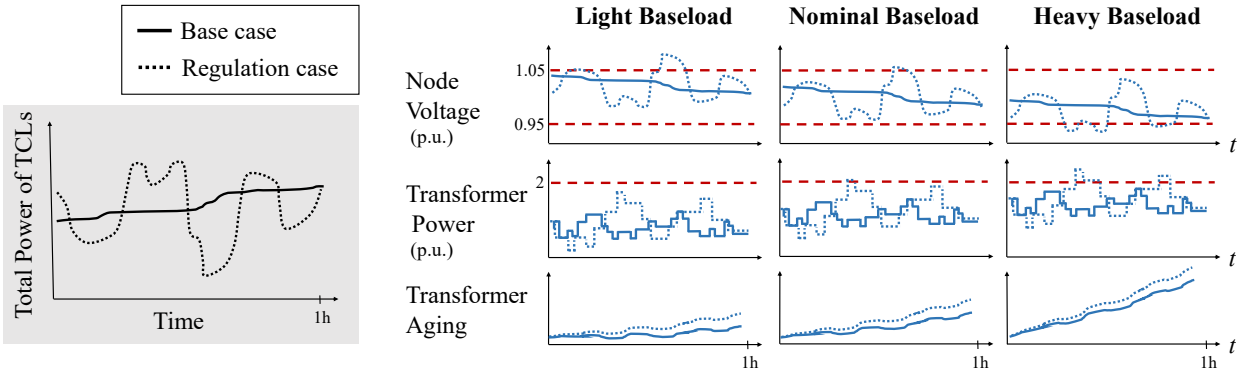


Figure 3.1: Illustration of possible effects of load-based regulation on distribution network operation at different baseload levels. On left: Aggregate TCL power is relatively flat in the base case and follows a regulation signal in the regulation case. On right: Increased variation in power during regulation could cause violations in voltage limits and transformer power limits; as baseload levels increase, under-voltage violations may become more likely than over-voltage violations, and transformer power violations may also become more likely. Operational limits are indicated with red dashed lines.

these studies did not investigate the network impact of DERs providing frequency regulation. For example, [15, 36] studied the impacts of high penetrations of grid-feeding PV systems, and [13, 22, 31, 81] studied the impacts of high penetrations of EVs with uncoordinated charging. Because load-based regulation has a unique effect on distribution network power flows—variation in power flows increase while mean power flow stays the same—its effects cannot be approximated by the aforementioned impact studies. The one impact study that does address load-based regulation, [97], found that regulation-provision by TCLs affects distribution transformers' maximum temperatures. However, the scope of this study was limited to distribution transformers; a model of the network and analysis of other network components was not included.

In this chapter, we study five different distribution networks and, for each network, identify the subset of network constraints that are at increased risk of violation when TCLs provide regulation. We identify these sets by simulating the distribution networks in a "base case" in which TCLs operate normally, and a "regulation case" in which TCLs are controlled to track a regulation signal. The illustration in Fig. 3.1 shows how network variables might change in the regulation case, as compared to the base case, at three different baseload levels. For instance, voltages may vary more when TCLs provide regulation, which may increase the prevalence of over-voltages in light baseload conditions and under-voltages in heavy baseload conditions. The full set of constraints that we assess includes line current, transformer power,

transformer aging rate, voltage magnitude, and voltage unbalance.

The primary contribution of this chapter is the comprehensive study of the effects of load-based regulation on distribution network operation. Additional contributions include: i) an examination of load-based regulation's effect on voltages at different baseload levels, ii) an explanation of why some transformers are more likely than others to age faster due to load-based regulation, given our chosen control strategy, and iii) identification of the common trends across the five feeders studied and preliminary ideas on network-protecting control techniques.

3.2 Feeder, HVAC, and Transformer Models

3.2.1 Simulation Software and Prototypical Feeder Models

We use GridLAB-D [63] to run power flow simulations of distribution networks with time-varying loads. GridLAB-D performs quasi-steady state analysis using a Newton-Raphson algorithm at each time step to solve for a network's three-phase, unbalanced power flow solution. We also use GridLAB-D's dynamic, physics-based models of heating, ventilation, and air conditioning (HVAC) systems.

We use network models from the Pacific Northwest National Lab's (PNNL's) prototypical feeder database [63]. These models are of actual networks and include fuses, voltage-regulators, capacitor banks, and distribution transformers. The feeders are prototypical in that their characteristics are representative of other networks, as determined by the rigorous statistical analysis of [75]. Each prototypical feeder represents a certain class of networks in one of five climatic regions in the U.S. For example, feeder R1-12.47-1 is estimated to represent 20.56% of networks on the west coast of the U.S. Thus, the effects of load-based regulation on a prototypical feeder will be indicative of, though not identical to, those that will occur in a much larger set of networks.

The original feeder models have only one constant-power load per distribution transformer; we require higher resolution, time-varying load models. Therefore, we disaggregate the transformer-level loads using a method provided by PNNL [64] that estimates the number of houses represented by each load and then constructs a time-varying model for each house. The house model comprises individual load models (HVAC, water heaters, and pool pumps) and ZIP models that aggregate all other loads in the house. The HVAC model is primarily driven by changes in hourly weather, with region-specific data sourced from the typical

meteorological year (TMY) database; the ZIP loads' base power values vary according to heterogeneous hourly schedules provided by [64].

We make the following modifications to the feeder models to improve the realism of our study.

- Capacitor banks: By default, capacitor banks with more than one phase are configured so that the voltage-sensing phase is the only phase that is controlled. We modify these types of banks so that all phases are controllable.
- Power factors: The disaggregation method [64] sets houses' zip-loads to a power factor of 1.0. We adjust the power factors using Table A.2 of [15] to better represent common loads in residential and commercial buildings.
- Small static loads: In the original feeder models, there are a number of small static loads. By default, method [64] converts these loads to "street lights", which are only on in the evening. We revert these back to constant loads so that the loads are on during the peak hour.
- Transformer sizing: The disaggregation method [64] can result in some transformers being undersized. We increase the size of transformers whose average loading is both greater than the transformer's original rating and greater than its original planning load by selecting the next largest transformer.

3.2.2 Controlled HVAC Model

GridLAB-D's HVAC model includes a thermal model of a house, as well as a model of the space-conditioning devices. The disaggregation method [64] selects randomized parameter values (e.g., wall insulation, house footprint size, temperature setpoints), such that each HVAC model is unique. Throughout this chapter we use the term "AC" to refer to the thermal model of the house, as well as the air cooling device itself. GridLAB-D's AC model has three dynamic states: indoor air temperature, mass temperature, and on/off status (see [65] for details).

We non-disruptively [8] control ACs with on/off commands that maintain indoor temperature within the user-set deadband. We use the simplest form of the probabilistic dispatch method (see [9, 52, 96]) in which a single switching probability is broadcast to all ACs and, based on this probability, each AC individually decides whether to switch. An AC is available to switch if three conditions are satisfied: 1) its indoor air temperature is within the

temperature deadband; 2) the AC has not switched within the last two minutes; and 3) if the AC were switched, it would not reach its temperature limit in under two minutes, as predicted by a model. The first condition ensures the control will be non-disruptive to the user and the last two conditions protect the unit's compressor by preventing excessive switching.

We use the proportional control scheme from [52] to calculate the switching probability u. The sign of u indicates the direction to switch: positive to switch on, negative to switch off. Let P^*_{total} be the desired power setpoint and P_{meas} the measured power of the AC population. If $(P^*_{\text{total}}(k{+}1){-}P_{\text{meas}}(k)) \geq 0$, then TCLs need to be switch on and the switching probability is calculated as

$$u(k+1) = K_{\text{P}} \frac{P^*_{\text{total}}(k+1) - P_{\text{meas}}(k)}{\bar{P}_{\text{on}} N_{\text{off}}(k)}. \tag{3.1}$$

If $(P^*_{\text{total}}(k+1) - P_{\text{meas}}(k)) < 0$, then TCLs need to be switched off and

$$u(k+1) = K_{\text{P}} \frac{P^*_{\text{total}}(k+1) - P_{\text{meas}}(k)}{\bar{P}_{\text{on}} N_{\text{on}}(k)}, \tag{3.2}$$

where the parameter K_{P} is a proportional gain, $\bar{P}_{\text{on}}$ is the average power of ACs that are on in steady-state, N_{on} is the number of *on* ACs that are available to switch *off*, and N_{off} is the number of *off* ACs that are available to switch *on*. We assume N_{on}, N_{off}, and P_{meas} are measured perfectly.

3.2.3 Aging Model of Distribution Transformers

The primary aging mechanism for distribution transformers is the deterioration of coil insulation due to heat from resistive losses [33]. To estimate transformer aging, we use GridLAB-D's built-in model, which is based on sections 5-7 of IEEE Standard C57.91-1995 [33]. Figure 3.2 shows three key variables within the model: the transformer's load, winding temperature, and estimated minutes aged.

The model has two dynamic states: $\tilde{\theta}_{\text{oil}}$ the difference between the transformer's top-oil temperature and ambient temperature (i.e., $\theta_{\text{oil}} - \theta_{\text{amb}}$); and $\tilde{\theta}_{\text{w}}$ the difference between the hot-spot winding temperature and the top-oil temperature (i.e., $\theta_{\text{w}}{-}\theta_{\text{oil}}$). Given these states,

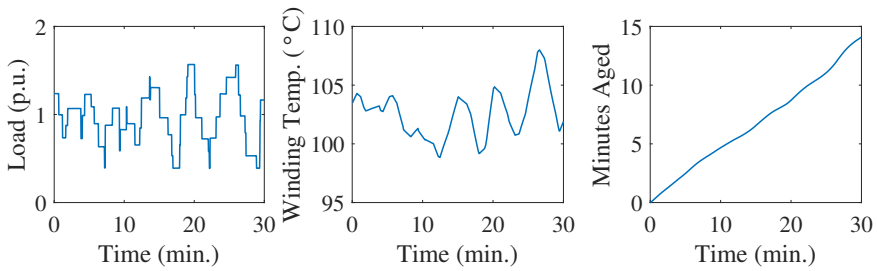


Figure 3.2: Variables within the transformer aging model: (left) load served by transformer; (middle) temperature of the winding's hot-spot; (right) minutes aged by the transformer. Note that minutes aged are less than the number of simulated minutes because the winding temperature is always <110°C.

Table 3.1: Single-Phase Transformer Thermal Parameters

Parameter	**Value**[a]	**Unit**	**Source**
Winding time-constant (τ_{w})	5	min	[33]
Winding hot-spot rise ($\tilde{\theta}_{\mathrm{w,R}}$)	80	°C	[33]
Top-oil rise ($\tilde{\theta}_{\mathrm{oil,R}}$)	60	°C	[20]
Full-load loss (l_{fl})	0.0232-0.0112	per unit	[66]
No-load loss (l_{nl})	0.0065-0.0042	per unit	[66]
Oil volume	5.7-62.7	gal	[91]
Core-plus-coil weight	56.6-484.9	lb	[91]
Tank-fittings weight	67.9-581.9	lb	[91]

[a]Values expressed as a range are parameters that depend on the rating of the transformer (5 kVA-175 kVA).

a transformer's thermal dynamics are described by

$$\begin{aligned} \frac{d\tilde{\theta}_{\mathrm{oil}}}{dt} &= \frac{1}{\tau_{\mathrm{oil}}}(\tilde{\theta}_{\mathrm{oil,u}} - \tilde{\theta}_{\mathrm{oil}}) \\ \frac{d\tilde{\theta}_{\mathrm{w}}}{dt} &= \frac{1}{\tau_{\mathrm{w}}}(\tilde{\theta}_{\mathrm{w,u}} - \tilde{\theta}_{\mathrm{w}}), \end{aligned} \tag{3.3}$$

where $\tilde{\theta}_{\mathrm{oil,u}}$ and $\tilde{\theta}_{\mathrm{w,u}}$ represent the ultimate temperatures that would be reached if the present load were sustained indefinitely [33]. The ultimate temperatures are computed as $\tilde{\theta}_{\mathrm{w,u}} = \tilde{\theta}_{\mathrm{w,R}} L^{1.6}$ and $\tilde{\theta}_{\mathrm{oil,u}} = \tilde{\theta}_{\mathrm{oil,R}}((L^2 l_{\mathrm{fl}}/l_{\mathrm{nl}} + 1)/(l_{\mathrm{fl}}/l_{\mathrm{nl}} + 1))^{0.8}$, where L is the time-varying load (per unit). The oil time constant τ_{oil} is computed according to equations (14) and (15) in [33]. Values for the thermal parameters in the above equations are derived from data found in [20, 33, 66, 91] (see Table 3.1).

Table 3.2: Network Constraints

Component	Variable	Lower limit	Upper limit
Service node	Continuous voltage	0.95 p.u.	1.05 p.u.
	Emergency voltage	0.9 p.u.	1.083 p.u.
3-Phase node	Voltage unbalance		3%
Transformer	Apparent power		200% of rating
	Average aging rate		1x nominal
Line	Current		100% of rating

The final stage of the model calculates a transformer's aging rate F_{AA} according to the empirically derived formula [33]

$$F_{\mathrm{AA}} = \exp\left(\frac{1500}{383} - \frac{1500}{\theta_{\mathrm{w}} + 273}\right), \tag{3.4}$$

where θ_{w} is the winding temperature and $\theta_{\mathrm{w}} = \theta_{\mathrm{amb}} + \tilde{\theta}_{\mathrm{oil}} + \tilde{\theta}_{\mathrm{w}}$. The nominal aging rate is equal to one and occurs when $\theta_{\mathrm{w}} = 110°\mathrm{C}$. Note that, in this model, the aging rate is not dependent on the transformer's age.

3.3 General Methodology for Studies

We test feeders with a scenario designed to reveal the negative effects of load-based regulation. In this scenario, we assume 100% of residential ACs are controllable, we test networks during the peak-load hour of the year, and we maximize the amplitude of the regulation signal (described in Section 3.3.2). We assess the effects of load-based regulation by comparing the operation of the distribution network through two 1-hour simulations: 1) the "base case" simulation, in which ACs operate normally, and 2) the "regulation case" simulation, in which ACs are controlled to track a regulation signal.

3.3.1 Assessment Criteria

We assess the effects of regulation by identifying changes in network variables between the base and regulation cases. Variables of interest and their corresponding constraints are listed in Table 3.2. First in the table are constraints on the voltage magnitude at service nodes, which are where service lines connect to distribution transformers. The continuous voltage

limits are based off of ANSI standard C84.1 [59] and are only violated if surpassed for over two minutes. The emergency limits are from [26] and are similar to those in [59] and [76]. Second in the table is voltage unbalance of 3-phase nodes: unbalance should be kept below 3% to keep 3-phase motors from overheating [59].

Third in the table are constraints to prevent transformers from aging too rapidly. Many utilities use apparent power as a proxy for aging rate. We use a limit of 200% of a transformer's rating as in [76]. Unlike utilities, we have access to our simulated transformers' aging rates. We constrain a transformer's aging rate such that its average rate over the simulation hour must be less than one. If a transformer surpasses this limit, its lifetime will be shorter than nominal (20 years). Fourth in the table is current flow, which is monitored for overhead and underground lines on the primary side of distribution transformers. We set the over-current constraint to 100% of the line's rating. We also monitor the status of all fuses.

3.3.2 Testing Conditions

To create peak-loading conditions in the test simulations, we prepare each feeder model as follows. First, we populate the feeder model using the disaggregation method [64] described in Section 3.2.1. Next, we find the populated feeder's peak-load hour by running a simulation over the summer months and selecting the date and hour of maximum substation-load. We calibrate the populated feeder's peak-load by comparing it to the original "planning load", which was selected by the utility to represent the expected peak load conditions [24]. If the populated feeder's peak load is less than 90% of the planning load, we use [64] to iteratively repopulate the feeder with more houses until its peak is between 90% and 100% of the original planning load.

For the regulation signal, we select a segment of the PJM Reg-D signal [68] in which the energy consumed during the regulation hour is equal to that of the base case. We scale the signal to achieve maximum power deviations by the population of 40% of its average baseline consumption—the largest capacity the AC population could provide before performance began to deteriorate.

To ensure appropriate initialization of dynamic states, we simulate the feeder in the 24 hours that precede the test hour (using a 30 second time step for the first 23.5 hours and a 2 second time step for the last 0.5 hours). Once properly initialized, we run the base case and regulation case simulations using a 2 second time step. We have stored all data and code

Table 3.3: Feeder Characteristics and Peak-Hour Average Conditions

Feeder	Areas served	Temp. (°C)	Load (MVA)	AC Load	Regulation Capacity
R1-12.47-1	Suburb, rural	34.7	5.71	43%	17%
R2-12.47-2	Suburb	34.4	7.08	51%	20%
R3-12.47-3	Suburb	45.6	8.13	42%	17%
R4-12.47-1	Urban, rural	35.8	6.12	51%	20%
R5-25.00-1	Suburb, urban	33.1	8.25	55%	22%

Regulation capacity expressed as a percentage of total feeder load

needed to replicate the test simulations in the public repository [72].

3.4 Main Study

3.4.1 Setup

We conduct a survey across five networks with the goal of determining how sensitive the results are to different feeder parameters and topologies. Each feeder is from one of five climate regions in the U.S. and varies in terms of voltage level, topology, geographical density of buildings, etc. [75]. Table 3.3 lists the feeders' key characteristics and conditions during the peak-hour. The first two parts of a feeder's name indicates (climate region number)-(voltage level). For brevity, we will refer to the feeders only by region number. In this study, we expect the effects of regulation to be larger for feeders R2, R4, and R5 because they have higher percentages of AC load and thus higher regulation capacities as a percentage of feeder load (see Table 3.3). We also anticipate that regulation could cause voltage issues on feeders with long lines, which are typically in rural areas (i.e., R1 and R4).

3.4.2 Results

Service-Node Voltage Results

We find that load-based regulation causes voltages to vary more, but the increase is not large enough to cause constraint violations in the networks we study. As shown in Fig. 3.3, voltages deviate in the opposite direction of the change in power injections due to load-based regulation. Figure 3.4 shows that, for all feeders, load-based regulation causes an increase in

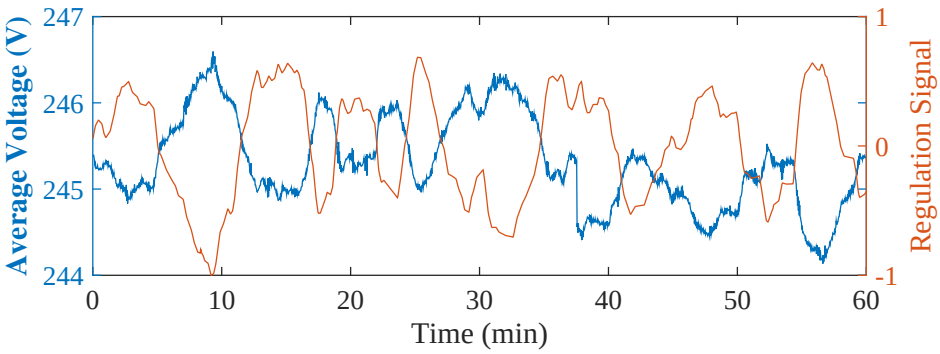


Figure 3.3: Negative correlation between the regulation signal and average voltage during a regulation case simulation. The voltage trajectory is the average across all residential service nodes. (Data from R1.)

the variation of voltage at service nodes. Despite the increase in voltage variation, no limit violations occur as a result of regulation for any of these feeders. Feeder R4 has voltages that surpass the continuous upper limit in the regulation case, but for less than the 2 minute duration required for continuous limit violations. For all feeders, there are no violations of emergency limits.

Figure 3.5 shows the voltage distributions of all service nodes for the five feeders. An increase in variation due to regulation is indicated by elongation in the distributions. We find that the voltage density outside of (or close to) voltage limits generally increases due to regulation: three of five distributions move closer to the upper limit, and five of five distributions move closer to the lower limit. There would likely be worse impacts on long, rural feeders with poor voltage regulation. In these "weak" feeders, we expect voltage distributions would be longer-tailed, and more density would shift outside of the limits in regulation cases. Unfortunately, we did not have access to a weak feeder model for testing.

3-Phase Node Voltage-Unbalance Results

Load-based regulation has little impact on voltage unbalance. Although feeders R1 and R4 have nodes that violate the upper limit of 3%, they do so in both the base case and the regulation case. The largest increase in unbalance due to load-based regulation is only +0.076%. Its minimal effect on voltage unbalance is likely due to the even distribution of ACs across the three phases of each network.

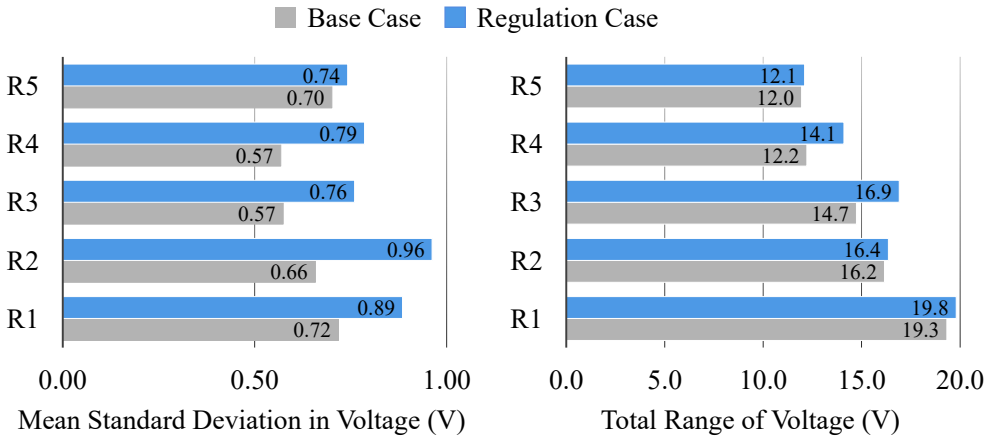


Figure 3.4: Variation of voltage at residential service nodes. Mean standard deviation of voltages is the mean across all nodes; total range of voltages is the range across all nodes. Both metrics increase in regulation cases for all five networks.

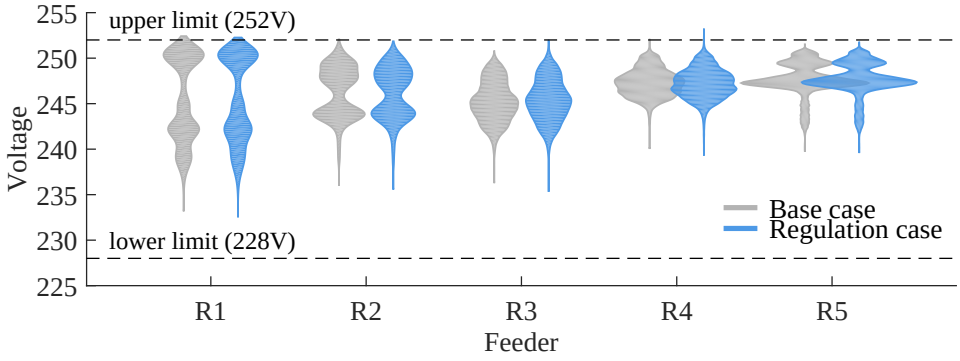
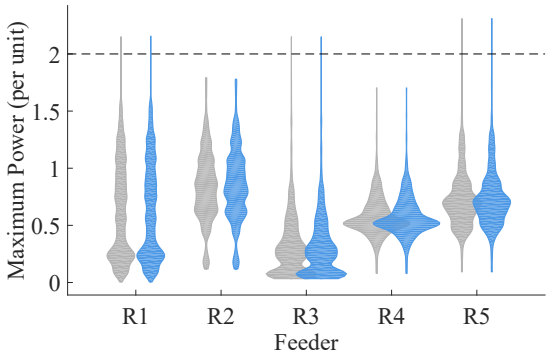


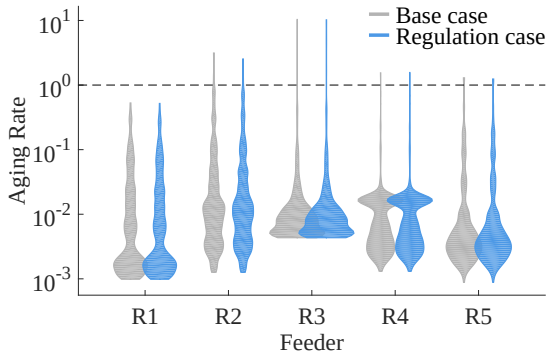
Figure 3.5: Voltage distributions of residential service nodes. Distributions for R1 and R4 cross the continuous upper limit in both cases. Regulation reduces over-limit voltages for R1 but increases over-limit voltages for R4. Distributions are visualized using kernel density plotting.

Table 3.4: Transformer Results: Apparent Power and Aging Rate

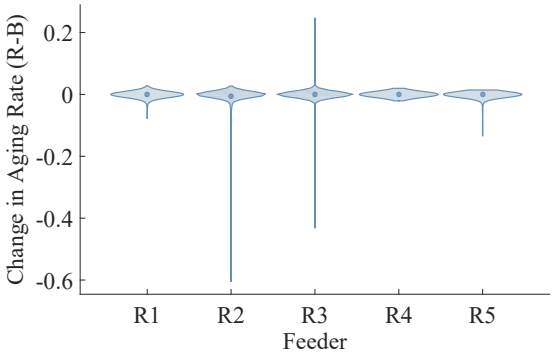
Feeder	Population mean			% of pop. w. violation	
	Power (p.u.)	Aging rate		Power	Aging
	base = *reg.*	*base*	*reg.*	*base* = *reg.*	*base* = *reg.*
R1	0.42	0.026	0.025	0.17	0
R2	0.55	0.076	0.070	0	1.04
R3	0.23	0.077	0.076	0.06	1.23
R4	0.43	0.015	0.015	0	0.21
R5	0.47	0.031	0.030	0.54	0.81



(a) Transformer apparent power distributions



(b) Transformer aging rate distributions



(c) Change in aging rate distributions

Figure 3.6: Transformer population distributions for (a) power, (b) aging rate, and (c) change in aging rate (regulation - base). Plots (a) and (b) show that the power and aging rate distributions are positively skewed and do not change substantially from base case to regulation case. Plot (c) shows that the change in aging rate is mostly symmetric about the $y = 0$ line, excluding a few large outliers.

Transformer Results

For all networks, we find that transformers experience very few apparent power or aging rate violations and that load-based regulation has little effect on these violations as shown in Table 3.4. For all five feeders, the per-unit apparent power of transformers, averaged across the population, is less than 0.6 p.u. (where the p.u. base is the transformer rating) and does not change from base case to regulation case. These low loading levels translate to low mean aging rates. In four of five feeders, the mean aging rate decreases slightly due to regulation, indicating that regulation may be beneficial for some transformers. The percentage of the population with violations (apparent power or aging rate) is small and unaffected by regulation.

Distributions of the transformer populations' loading and aging rates are shown in Fig. 3.6. In each feeder, a small number of transformers have much higher loading and aging rates than the rest of the population, as indicated by the positive skews in Fig. 3.6(a) and (b). Notably, the aging rate distributions require a log-scale for visualization. Figure 3.6(c) demonstrates that the changes in aging rate from base to regulation case are generally small. As indicated by the thin spikes, there are a few transformers that experience large changes in aging rate. Excluding these outliers, the distributions in Fig. 3.6(c) are approximately symmetric about the line $y = 0$, i.e., about 50% of transformers age faster due to regulation, and about 50% age slower.

Line Results

In all simulations, line-current stays well below the 100% conductor rating constraint, and all fuses remain closed.

3.5 Baseload Study

3.5.1 Setup

In this study, we examine the effects of load-based regulation as net baseload changes. In the future, large-scale adoption of DERs could result in large changes to net baseload. For example, EV charging would increase net baseload during popular charging hours, and PV in-feed would decrease net baseload during daylight hours. We conduct this study on feeder R1 and investigate a shift up in baseload in the "Heavy Baseload" trial, and a shift down

in baseload in the "Light Baseload" trial. We model these changes in baseload by adding a constant power ZIP load to 20% of R1's houses, with power demand of +3.3kW for an increase in load and -3.3kW for a decrease in load. (The power level of 3.3kW represents a reasonable charging rate for an EV.) In these trials, the aggregate capacity of the added ZIP loads is 25.6% of R1's nominal baseload of 5.71 MVA.

3.5.2 Results

We find that at different baseload levels, the prevalence of voltage violations changes, but not always in an intuitive direction. Voltage results are presented in Table 3.5. We restrict our attention to the 1.05 p.u. continuous limit because it is the only constraint violated. In the base case, the results for the Heavy Baseload and Nominal Baseload trials are similar: 5.69% and 5.35% of nodes, respectively, exceed the 1.05 p.u. limit for at least one time step. However, in the regulation case, over-limit nodes increase to 20.07% in the Heavy Baseload trial and decrease to 4.85% in the Nominal Baseload trial. A similar divergence occurs in the Heavy and Nominal Baseload trials when examining nodes with continuous voltage violations (i.e., over-limit for more than 2 minutes). In contrast to the other trials, the Light Baseload trial has only 1.34% of its nodes over-limit for any duration, and this percentage remains almost constant across cases and durations.

We use two metrics—voltage mean and voltage range—to explain the above results; these metrics are shown in Fig. 3.7 as distributions across all primary-side nodes. In general, we observe that mean voltages change primarily as a function of baseload level (see Fig. 3.7(a)), and voltage ranges change primarily as function of case, i.e., base case or regulation case (see Fig. 3.7(b)). Setting aside phase A, we observe two trends: 1) as baseload increases, mean voltages decrease; and 2) in all trials, voltage ranges substantially increase from the base case to the regulation case. However, because the majority of nodes on phase B and C have mean voltage less than 1.03 p.u., all voltages remain below the 1.05 p.u. upper limit even in regulation cases.

All of the nodes that experience over-voltages are located on phase A. Counterintuitively, as baseload increases, more nodes on phase A experience over-voltages for at least one time step (see Table 3.5). However, these over-voltage results are consistent with the distributions for phase A in Fig. 3.7. Specifically, relative to the Nominal Baseload trial, mean voltages on phase A decrease in the Light Baseload trial (due to a capacitor bank switching off), resulting in fewer over-voltages. In the Heavy Baseload trial, mean voltages stay constant

Table 3.5: Baseload Study: % of Nodes above 1.05 p.u. Voltage Limit

	For at least one time step		**For more than 2 min**	
	base	*regulation*	*base*	*regulation*
Heavy Baseload	5.69	20.07	0.17	0.84
Nominal Baseload	5.35	4.85	0.17	0
Light Baseload	1.34	1.34	1.34	1.17

relative to the Nominal Baseload trial, but voltage ranges on phase A increase, resulting in more over-voltages. It is worth noting that this feeder is particularly unbalanced: the substation's average apparent power flow is 1.1 MVA, 2.2 MVA, and 2.4 MVA for phases A, B, and C, respectively; and so the 3-phase capacitor bank, sized evenly across the three phases, over compensates phase A.

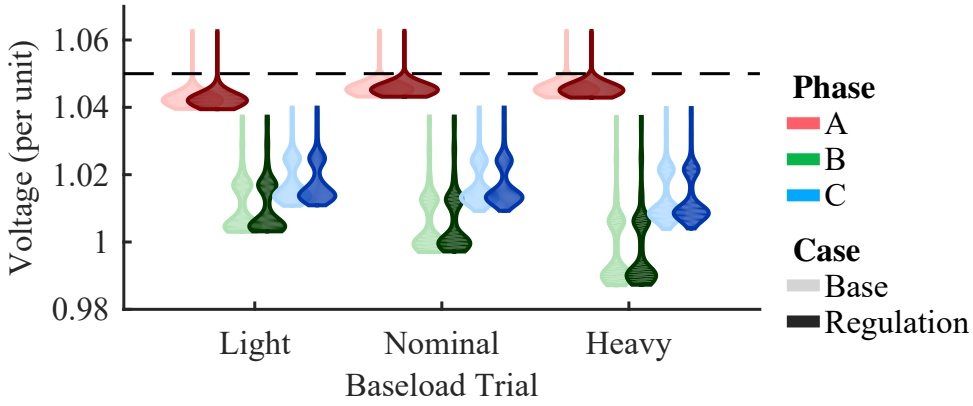
3.6 Randomization Study

3.6.1 Setup

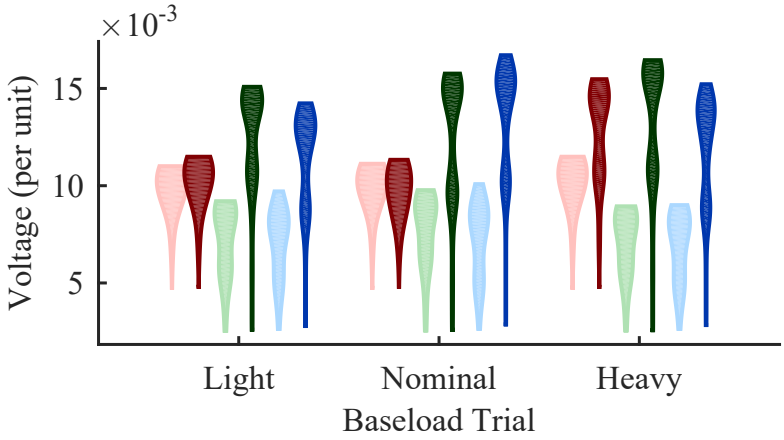
This study is designed to investigate one of the results from the main study—that some transformers have an increased aging rate due to load-based regulation. Our goal is to determine whether some transformers consistently have an increased aging rate across multiple trials. We run six randomized trials, all on feeder R1, with different random instantiations of the initial *on*/*off* status of the ACs as well as the ACs' probabilistic responses; all other parameters are kept constant.

3.6.2 Simulation Results

We find that some transformers are more likely to age faster due to load-based regulation and some are more likely to age slower. Figure 3.8 shows the observed distribution of the number of trials transformers experience an increased aging rate due to regulation. We compare this distribution to the "expected distribution", which represents the hypothesis that all transformers are equally likely to have an increased aging rate. Specifically, we model the transformers' aging outcomes as i.i.d. Bernoulli random variables with probability 0.49 of increased aging. (This value is the observed prevalence of increased aging in the randomized trials.) The expected distribution, then, is the number of increased aging outcomes that



(a) Mean voltage distributions



(b) Voltage range distributions

Figure 3.7: Voltage distributions across primary-side nodes for (a) mean voltages and (b) voltage ranges. Plot (a) indicates that mean voltages generally decrease as baseload increases and are relatively unaffected by regulation. Plot (b) indicates that voltage ranges increase in regulation cases, and the effect of baseload on voltage ranges is inconsistent.

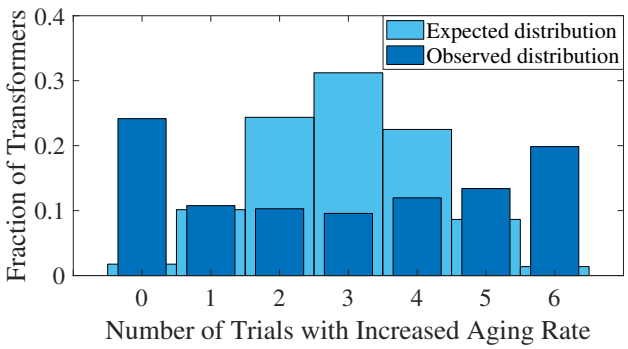


Figure 3.8: Observed and expected distributions for the number of randomized trials a transformer will experience an increased aging rate due to regulation. The difference between the two distributions indicates that some transformers are more likely than others to experience an increased aging rate.

occur over six Bernoulli trials, i.e., a Binomial distribution. The difference between the two distributions in Fig. 3.8 indicates that the aging outcomes cannot be modeled as i.i.d. random variables. Instead, the edges of the observed distribution indicate that a large portion of transformers is very likely to experience an increased aging rate, and another large portion is very likely to experience a decreased aging rate.

Load-based regulation does not have a consistent effect on transformers' aging rates, in part, because of an unanticipated effect of probabilistic dispatch. In the following discussion, we will refer to switching in response to probabilistic dispatch as "dispatch-switching". We will also refer to a cycle through an AC's deadband without dispatch-switching as a "natural cycle". When all ACs receive the same probabilistic command, an AC with a low natural duty-cycle is more likely to be dispatch-switched on than off, and an AC with a high natural duty-cycle is more likely to be dispatch-switched off than on. See Section 3.6.3 for a derivation of this result. If an AC is dispatch-switched on more than off, it will "cool-cycle", i.e., cycle in the cooler part of its deadband. This cycling behavior reduces the average temperature of the house, thus increasing its average power consumption and, consequently, the transformer's aging rate. A similar argument can be made for why warm-cycling reduces transformer aging rates. As shown in Fig. 3.9, there is a negative correlation between the percent change in aging rate of a transformer and the average natural duty cycle of the ACs the transformer supplies. Transformers supplying ACs with a lower than average duty cycle (< 0.503) frequently have an increased aging rate, and transformers supplying ACs with a higher than average duty cycle frequently have a decreased aging rate.

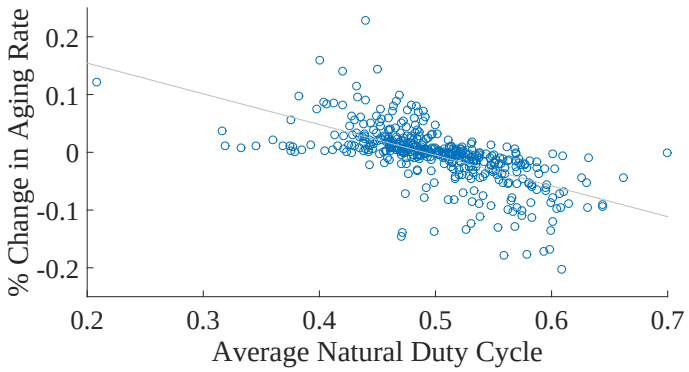


Figure 3.9: Scatter plot of each transformer's percent change in aging rate (averaged across the randomized trials) versus the average natural duty cycle of the ACs supplied by the transformer. The data is negatively correlated, with correlation coefficient = 0.628.

3.6.3 Analytical Result

We show that, in a heterogeneous population of ACs under probabilistic dispatch in which all ACs receive the same u, some ACs will be more likely to be dispatch-switched off than on, and others will be more likely to be dispatch-switched on than off. We then show that the natural duty cycle of an AC can be used to predict which behavior will be most likely.

When an AC is off, the probability of it dispatch-switching on is equal to one minus the probability of it not dispatch-switching on during the same time. Thus, the probability of an AC dispatch-switching on at least once during the time it would take to complete the off part of its natural cycle is

$$p_{\text{s,on}} = 1 - \prod_{k=1}^{\tau_{\text{off}}} (1 - u_{\text{on}}(k)), \tag{3.5}$$

and the probability of dispatch-switching off at least once during the on part of its natural cycle is

$$p_{\text{s,off}} = 1 - \prod_{k=1}^{\tau_{\text{on}}} (1 - u_{\text{off}}(k)). \tag{3.6}$$

Here u_{on} and u_{off} are the probabilistic dispatch commands for the off and on directions, and τ_{on} and τ_{off} are the number of time steps the AC is on and off during a natural cycle.

If an AC has $p_{\text{s,on}} > p_{\text{s,off}}$, then it is more likely to dispatch-switch on than off; if $p_{\text{s,on}} < p_{\text{s,off}}$, the AC is more likely to dispatch-switch off than on. To determine an AC's behavior *a priori*, we estimate $p_{\text{s,on}}$ and $p_{\text{s,off}}$ with a few simplifying assumptions. Let D_{c} be the average of the ACs' natural duty cycles and N be the total number of ACs. We

assume that the regulation signal's amplitude is relatively small such that the ratio of the number of ACs on to number of total ACs remains approximately equal to D_{c}. Thus, if we switch a fixed number of ACs, N_{S}, in each time step, we can approximate the switching commands as constant values: $u_{\mathrm{on}} \approx N_{\mathrm{S}}/((1-D_{\mathrm{c}})N)$ and $u_{\mathrm{off}} \approx N_{\mathrm{S}}/(D_{\mathrm{c}}N)$. Finally, let d_{c} be the natural duty cycle of the given AC and τ_{P} be the number of time steps in its period, then $\tau_{\mathrm{on}} = d_{\mathrm{c}}\tau_{\mathrm{P}}$ and $\tau_{\mathrm{off}} = (1-d_{\mathrm{c}})\tau_{\mathrm{P}}$. With these substitutions, the probabilities of dispatch-switching during a natural cycle are

$$p_{\mathrm{s,on}} \approx 1 - \left(1 - \frac{N_{\mathrm{S}}}{(1-D_{\mathrm{c}})N}\right)^{(1-d_{\mathrm{c}})\tau_{\mathrm{P}}} \text{ and} \tag{3.7}$$

$$p_{\mathrm{s,off}} \approx 1 - \left(1 - \frac{N_{\mathrm{S}}}{D_{\mathrm{c}}N}\right)^{d_{\mathrm{c}}\tau_{\mathrm{P}}}. \tag{3.8}$$

After manipulating (3.7) and (3.8), we find that an AC is more likely to dispatch-switch on than off (i.e., $p_{\mathrm{s,on}} > p_{\mathrm{s,off}}$) if

$$\frac{d_{\mathrm{c}}}{(1-d_{\mathrm{c}})} < \frac{\log\left(1 - \frac{N_{\mathrm{S}}}{(1-D_{\mathrm{c}})N}\right)}{\log\left(1 - \frac{N_{\mathrm{S}}}{D_{\mathrm{c}}N}\right)}. \tag{3.9}$$

Note that if $D_{\mathrm{c}} = 0.5$ then the right-hand side of (3.9) is equal to one. In this case, an individual AC will be more likely to dispatch-switch on than off if $d_{\mathrm{c}} < 0.5$, which is equivalent to $d_{\mathrm{c}} < D_{\mathrm{c}}$ and is what we observe empirically.

3.7 Control Recommendations

The simulation results have shown that, for the feeders studied, only a small subset of constraints are at risk of violation from load-based regulation. Because the feeders are prototypical, we expect this to be true in many other feeders. This is promising from the standpoint of developing a computationally efficient control algorithm, since it may be only necessary to enforce a subset, rather than the full set, of network constraints. The particular subset of at-risk constraints will be feeder-dependent. Networks with nodes whose voltage magnitudes are close to constraint limits are likely to be at risk because load-based regulation causes an increase in variation of voltage. Voltage unbalance constraints are unlikely to be a risk, unless the loads participating are unevenly distributed across the three phases.

Transformer constraint violations (power and aging) were rare in the networks studied. From a short-term reliability perspective, there may be a few transformers predisposed to overloading that are at risk of failure with any increase in loading—such transformers are at risk from load-based regulation. However, from a long-term reliability and cost perspective, a utility may benefit from load-based regulation since load-based regulation can decrease the transformer population's mean aging rate (and thus decrease the frequency of transformer replacement). Finally, it appears unlikely for line constraints to be at risk, unless load-based regulation capacity is significantly greater than the levels studied here, or a network has lines that already operate close to their limits.

We propose a few strategies for providing load-based regulation while protecting distribution networks. First, to prevent AC cool-cycling and thereby the likelihood of increased aging rates for transformers, ACs closer to naturally switching should be dispatched first (i.e., priority-stack control [28]). Second, to prevent constraint violations more generally, the amount of participation by loads in at-risk areas could be reduced, or additional controls could be used (e.g., smart inverters for voltage support). Reduction in participation could be implemented in one of two ways: 1) by allowing only a portion of loads in the area to participate in regulation, or 2) by dispatching loads in that area less frequently. Both of these strategies require some coordination between DER aggregators and the distribution operator.

3.8 Chapter Conclusion

In this chapter, we studied five prototypical feeders and, for each feeder, identified the subset of constraints that are at risk of increased violation due to the provision of regulation by TCLs. Ultimately, this subset is network dependent; however, in the five networks studied here, the subset of at-risk constraints is substantially smaller than the set of all network constraints, and we expect this is true for many other networks. This finding suggests that targeting at-risk constraints is a promising approach for computationally-efficient load control strategies.

It is possible that the impacts of regulation were not severe in the networks studied here because of their typical nature. Impacts would likely be larger for non-typical feeders, particularly ones with voltage-weak areas. Additionally, future distribution networks may be much more active and complex than those modeled here. For instance, a network could simultaneously have high penetrations of photovoltaics, EVs, and regulation-providing loads, as well

as multiple aggregators controlling portions of each DER population. On such a network, fluctuations in power injections due to transient clouds or intermittent EV charging could exacerbate the increased voltage variation caused by load-based regulation. The remaining chapters address the roles of the distribution operator and the aggregator in these types of networks—in particular, how to constrain aggregators' actions to ensure the reliability of the distribution network.

Chapter 4

Strategy for Network-Safe Load Control (Given Simplified Plant)

This chapter is largely based on the published work [71][4].

4.1 Chapter Introduction

Large-scale participation of load aggregations in energy balancing services could result in local operational issues on distribution networks, particularly as other distributed energy resources (DERs) such as electric vehicles and photovoltaic systems become more active on distribution networks [73]. Transmission operators, aggregators, and distribution operators will need to coordinate to prevent negative impacts on distribution networks [25, 57]. How this coordination should be structured is an open question. The U.S. Federal Energy Regulatory Commission has recently requested comments on the following questions: What, if any, real-time information do distribution operators need about aggregations or individual resources within an aggregation? Should distribution operators be able to override real-time dispatch of aggregations to resolve local reliability issues? [21].

Traditionally, load-control algorithms for frequency regulation services have not taken distribution network constraints into account. Instead, it has been assumed that the effect on network operation will be negligible if the percentage of loads participating in the aggregation is relatively small. To the best of our knowledge, only a few papers [17, 93] have proposed

[4] S.C. Ross, N. Ozay, and J.L. Mathieu. "Coordination between an aggregator and distribution operator to achieve network-aware load control". In: *Proceedings of the IEEE Power & Energy Society PowerTech Conference*. Milan, Italy, June 2019.

real-time algorithms for network-aware load control at the time scale required by frequency regulation services (i.e., seconds). However, in both [93] and [17], the distribution operator also acts as the load aggregator—an additional role that some operators may not want. Moreover, in some locations, operators may be required to allow independent aggregators to participate on their networks.

The purpose of this chapter is to achieve network-aware load control through aggregator-operator coordination. In Section 4.2, we propose two different frameworks for real-time coordination. In Section 4.3, we propose a specific coordination scheme in which an operator blocks an aggregator's commands if they will cause network issues, and we describe the resulting control problem for an aggregator of thermostatically controlled loads (TCLs). In Section 4.4, we propose a control strategy for the aggregator. Finally, in Section 4.5, we test the proposed controller against a benchmark controller in a simulation study.

The contributions of this chapter are: 1) development of two frameworks for real-time coordination between an independent aggregator and a distribution operator; 2) development of a controller for an aggregator whose commands can be blocked by an operator (specifically, design of a Kalman Filter that estimates the portion of TCLs not receiving commands and compensates within the controller accordingly); 3) a comprehensive simulation study that compares the proposed controller to a benchmark controller across multiple scenarios.

4.2 Operator-Aggregator Coordination

4.2.1 Objectives of Aggregator and Operator

The aggregator's objective is to non-disruptively control hundreds to thousands of loads such that their total power consumption accurately tracks a frequency regulation signal. Load control is considered non-disruptive if the end-use service delivered by loads (e.g., refrigeration) is not disrupted [8].

The distribution operator's objective is to provide sufficient quality power to consumers, while also ensuring safe operation of network components. Voltage magnitudes should be maintained between 0.95 and 1.05 p.u. and unbalance should be less than 3% [59]. Network components, such as lines and transformers, should not be loaded beyond their ratings.

4.2.2 Time Scales of Coordination

There are multiple time scales at which the aggregator and operator could coordinate: 1) upon formation of the aggregation, 2) prior to market-biding by the aggregator, and 3) during real-time operation. Upon formation of the aggregation, the distribution operator could run a power flow analysis of the network under worst case load-control scenarios as in [73]. If the operator finds operational issues, then coordination at the other two time scales will likely be necessary. Before submitting market bids, an aggregator would benefit from a prediction of network conditions from the operator. For example, if the network is likely to be constrained, the aggregator might reduce the power capacity of its bid. Lastly, the operator and aggregator could coordinate in real-time in order to adjust the aggregator's control if the operator judges that it is adversely affecting network operation. Different frameworks for real-time coordination are described in the next section.

4.2.3 Real-Time Coordination Frameworks

We propose two frameworks for real-time coordination between the aggregator and operator, as shown in Fig. 4.1. In the aggregator-centric framework, the operator sends the aggregator constraints on load actions (e.g., maximum and minimum power consumption of groups of loads), and the aggregator computes a control input that will satisfy these constraints. In the operator-centric framework, the operator receives the aggregator's desired control input (e.g., on/off commands to each load) and modifies the input if necessary to satisfy network constraints. The aggregator uses feedback from the loads to compute its control inputs, and the operator uses feedback from the network to estimate three-phase power flows and determine constraints on load actions.

Next we compare the frameworks in terms of their implications for tracking accuracy, network reliability, and information privacy,

Tracking Accuracy: In the aggregator-centric framework, the aggregator knows the constraints on its load actions and can adjust its control to mitigate the effect on tracking accuracy. In the operator-centric framework, the aggregator does not have explicit knowledge of the constraints and will have a harder time compensating for them.

Network reliability: In the operator-centric framework, the operator has full control over its network and can ensure its reliability; however, this reliability could come at the expense of the aggregator's objective. In the aggregator-centric framework, the operator may only be able to ensure network reliability by providing overly conservative constraints on load

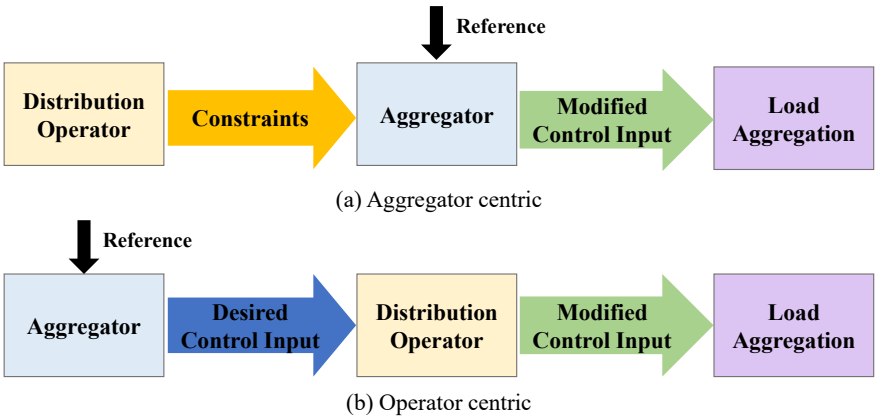


Figure 4.1: Proposed coordination frameworks. The distribution operator provides constraints on load actions, the aggregator provides its desired control input, and one entity ((a) aggregator or (b) operator) combines this information into a modified control input that is sent to the loads.

actions, which could result in poor tracking accuracy for the aggregator.

Information privacy: In the aggreagtor-centric framework, the aggregator's information is kept private, but the aggregator gains information about the operator and may be able to learn the network's parameters and configuration. In the operator-centric framework, the operator's information is kept private, but the operator gains information about the aggregator and may be able to learn the aggregator's control algorithm.

The proposed frameworks are only two of the many options for operator-aggregator coordination. Other options include a blend of these frameworks and a parallel framework in which the operator and aggregator independently send commands to the load aggregation.

4.3 Problem Description

This section describes a specific coordination strategy within the operator-centric framework and the resulting control problem for the aggregator.

4.3.1 Specific Coordination Strategy

A block diagram of the coordination strategy is shown in Fig. 4.2. As indicated by the figure, the aggregator's loads may be located on more than one of the operator's feeders, and each feeder has a safety controller to ensure the feeder's constraints are satisfied. When a safety controller receives the aggregator's desired control input, it uses its current estimate

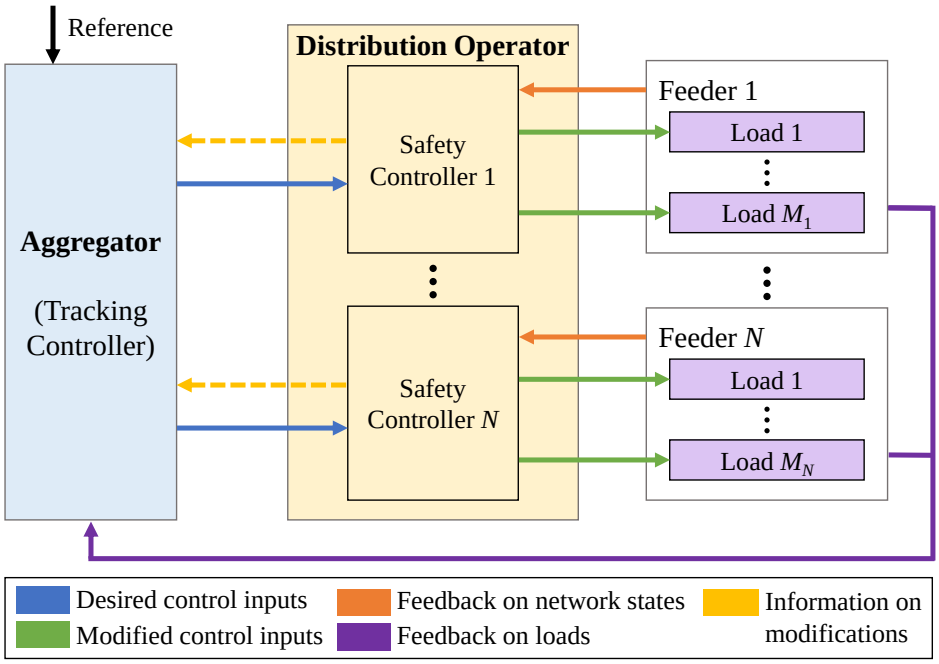


Figure 4.2: Specific coordination strategy within the operator-centric framework. The aggregator calculates its desired control input based on the reference signal and feedback from loads. The operator estimates the feeders' states based on measurements and modifies the control inputs if necessary. Information on modifications may or may not be shared with the aggregator.

of the feeder's states to determine if the input needs to be modified to ensure constraint satisfaction.

Although the modification of the aggregator's input could take many forms, in this problem we assume the operator "blocks" control inputs to certain loads. Unblocked inputs are sent unmodified to the appropriate loads; blocked inputs are simply not sent. Note, we refer to a load as "being blocked" if its control input is blocked. Finally, we define $\beta(k)$ to be the percentage of loads that are blocked at time step k.

We consider three scenarios in which the information available to the aggregator varies (see Table 4.1). In Scenario F, the aggregator has full information on which loads are blocked, as well as the on/off state of each load. In Scenario P, the aggregator has partial information on blocking. In Scenario N, the aggregator has no information on blocking.

Table 4.1: Information Scenarios

	Scenario F	Scenario P	Scenario N
Measurement of P_{total}	✓	✓	✓
Percent of loads blocked	✓	✓	
On/*Off* state of each load	✓		
Blocked state of each load	✓		

F = Full information on blocking, P= Partial information on blocking, N = No information on blocking

4.3.2 Aggregator's Control Problem

The aggregator's objective is to minimize the error between the reference power P^*_{total} and the load aggregation's actual power consumption P_{total} in each time step. In this problem, the load aggregation consists of TCLs, such as water heaters, air conditioners, and refrigerators. We use the term "TCL" to refer to both the conditioning device (e.g., heat pump) and the system being conditioned (e.g., water tank or house).

The aggregator's control problem is to track the reference power by switching TCLs on or off, in a manner that is non-disruptive to the end-user. Thus, the control method must respect the TCLs' user-set temperature constraints:

$$\theta^i_{\text{set}} - \delta^i/2 < \theta^i(k) < \theta^i_{\text{set}} + \delta^i/2, \tag{4.1}$$

where θ is the TCL's temperature, θ_{set} is the setpoint, and δ is the width of the allowed range. A TCL's temperature dynamics can be described by the following model (developed in [79] and frequently used in the literature [7, 19]):

$$\theta^i(k+1) = \begin{cases} a^i\theta^i(k) + (1-a^i)(\theta_{\text{amb}}(k) + r^i p^i_\theta), & \text{if } \sigma^i(k) = 1 \\ a^i\theta^i(k) + (1-a^i)\theta_{\text{amb}}(k), & \text{if } \sigma^i(k) = 0, \end{cases} \tag{4.2}$$

where $a^i = \exp(-h/(c^i\, r^i))$ and h is the duration of the discrete model's time step. Variable σ represents the TCL's power status (1 for *on* and 0 for *off*). All other TCL parameters are defined in Table 2.1.

We assume that the TCLs' temperature constraints and blocking conditions take precedence over the aggregator's switching commands. Let b^i indicate whether the ith TCL is blocked ($b^i = 1$) or not blocked ($b^i = 0$). Let s^i be the aggregator's switching command

to the ith TCL with $s^i = 1$ indicating "switch *on*", $s^i = -1$ indicating "switch *off*", and $s^i = 0$ otherwise. Thus, when the ith TCL is under aggregator control, its *on*/*off* state is determined by

$$\sigma^i(k+1) = \begin{cases} 1 & \text{if } \theta^i(k) \geq \theta^i_{\text{set}} + \delta^i/2, \\ 0 & \text{if } \theta^i(k) \leq \theta^i_{\text{set}} - \delta^i/2, \\ 1 & \text{if } s^i(k) = 1,\ b^i(k) = 0, \text{ and (4.1)}, \\ 0 & \text{if } s^i(k) = -1,\ b^i(k) = 0, \text{ and (4.1)}, \\ \sigma^i(k) & \text{otherwise.} \end{cases} \tag{4.3}$$

Lastly, if there are N TCLs in the aggregation, their total power consumption is given by $P_{\text{total}}(k) = \sum_{i=1}^{N} \sigma^i(k) P^i_{\text{R}}$, where $P^i_{\text{R}} = p^i_\theta/\zeta^i$ is the rated electrical power consumption of the ith TCL and ζ^i is the coefficient of performance.

4.4 Control System Design

4.4.1 Probabilistic Control

We use probabilistic commands to switch TCLs *on*/*off*, as in [52], because of the light communication requirements. If the operator does not modify the aggregator's desired control inputs, then only two numbers need to be broadcast to all TCLs: u_{off} the probability with which *on* TCLs should switch *off* and u_{on} the probability with which *off* TCLs should switch *on*. Each TCL determines its individual switching command s_i by drawing a random number $z_i(k)$ from the uniform distribution between 0 and 1, and

$$s^i(k) = \begin{cases} 1 & \text{if } \sigma^i(k) = 0 \text{ and } z^i(k) < u_{\text{on}}(k), \\ -1 & \text{if } \sigma^i(k) = 1 \text{ and } z^i(k) < u_{\text{off}}(k), \\ 0 & \text{otherwise.} \end{cases}$$

Once s^i is calculated, a TCL switches according to (4.3).

Probabilities to switch are calculated based on the predicted error between $P^*_{\text{total}}(k+1)$ the desired total power in the next time step and $P_{\text{total}}(k+1|k)$ the predicted total power

in the next time step. If $P^*_{\text{total}}(k+1) \geq P_{\text{total}}(k+1|k)$, then $u_{\text{off}}(k) = 0$ and

$$u_{\text{on}}(k) = K\frac{|P^*_{\text{total}}(k+1) - P_{\text{total}}(k+1|k)|}{\hat{P}_{\text{off}}(k)}. \tag{4.4}$$

If $P^*_{\text{total}}(k+1) < P_{\text{total}}(k+1|k)$, then $u_{\text{on}}(k) = 0$ and

$$u_{\text{off}}(k) = K\frac{|P^*_{\text{total}}(k+1) - P_{\text{total}}(k+1|k)|}{\hat{P}_{\text{on}}(k)}. \tag{4.5}$$

In (4.4) and (4.5), K is a proportional control gain, and $\hat{P}_{\text{on}}$ and $\hat{P}_{\text{off}}$ are estimates of the total capacity of all *on* and unblocked TCLs and all *off* and unblocked TCLs, respectively. Lastly, we constrain switching probabilities to within the range $[0, 1]$.

4.4.2 Estimator

To implement the controller in (4.4)-(4.5), the aggregator needs to estimate the total capacities available to switch, $\hat{P}_{\text{on}}$ and $\hat{P}_{\text{off}}$. We assume the aggregator has little to no information about the operator's blocking actions. Thus, our approach is to estimate the number of TCLs that are *on* and not blocked N_{on} and the number that are *off* and not blocked N_{off}, and then estimate the capacities as

$$\hat{P}_{\text{off}}(k) = \overline{p}_{\text{off}}\hat{N}_{\text{off}}(k) \quad \text{and} \quad \hat{P}_{\text{on}}(k) = \overline{p}_{\text{on}}\hat{N}_{\text{on}}(k).$$

Here $\hat{}$ indicates an estimate, and $\overline{p}_{\text{off}}$ and $\overline{p}_{\text{on}}$ are the average power ratings of an *off* TCL and an *on* TCL, respectively. For simplicity, we model $\overline{p}_{\text{off}}$ and $\overline{p}_{\text{on}}$ as constant parameters; in practice, they are time-varying for a heterogeneous aggregation.

We formulate a linear time varying model that represents aggregate TCL dynamics and for which a Kalman Filter can be designed. The aggregate model is

$$\begin{aligned} \mathbf{x}(k+1) &= \mathbf{A}(k)\mathbf{x}(k) + \mathbf{w}(k) \\ \mathbf{y}(k) &= \mathbf{C}\mathbf{x}(k) + \mathbf{v}(k), \end{aligned} \tag{4.6}$$

where $\mathbf{w}$ and $\mathbf{v}$ are process noise and measurement noise, respectively. The state $\mathbf{x}$ is defined as

$$\mathbf{x}(k) = \begin{bmatrix} N_{\text{on}}(k) & N_{\text{off}}(k) & N_{[\text{on}]}(k) & N_{[\text{off}]}(k) \end{bmatrix}^{\text{T}},$$

where $N_{[\text{on}]}$ the number *on* and blocked, and $N_{[\text{off}]}$ the number *off* and blocked.

We model the state dynamics as a Markov chain, in a similar manner to [52]. However, in this formulation the Markov chain is time-varying because we include the time-varying commands u_{on} and u_{off} within the transition probabilities, as in [11]. The Markov chain can be represented by the transition matrix

$$\mathbf{A}(k) = \begin{bmatrix} 1-f-u_{\text{off}}(k) & g+u_{\text{on}}(k) & 0 & 0 \\ f+u_{\text{off}}(k) & 1-g-u_{\text{on}}(k) & 0 & 0 \\ 0 & 0 & 1-f & g \\ 0 & 0 & f & 1-g \end{bmatrix}.$$

The upper block in $\mathbf{A}$ consists of the transition probabilities for TCLs that are not blocked; the lower block consists of transition probabilities for TCLs that are blocked. The model does not include transition probabilities between blocked and unblocked states because the operator determines these transitions with its blocking actions; thus, from the perspective of the aggregator, these transitions are an unknown disturbance to the state. Scalars f and g represent the probability of transitions caused by TCLs' internal thermostat control, i.e., from "internally switching". Scalar f is the probability that an *on* TCL is switched *off* internally, and scalar g is the probability that an *off* TCL is switched *on* internally. In the upper block, the probability of transitioning to a new state is equal to the sum of the probabilities of switching internally and switching externally.

The output $\mathbf{y}$ and corresponding $\mathbf{C}$ matrix are defined as

$$\mathbf{y}(k) = \begin{bmatrix} P_{\text{total,meas}}(k) \\ N \end{bmatrix} \quad \mathbf{C} = \begin{bmatrix} \overline{p}_{\text{on}} & 0 & \overline{p}_{\text{on}} & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}.$$

Here the first output equation relates the number of TCLs that are *on* to the measured power consumption of the aggregation. The second output equation is an equality constraint on the states: the sum of the states must be equal to the number of TCLs in the aggregation. To incorporate this constraint into the Kalman Filter, we assume the second output is a perfect measurement with zero measurement noise [77].

The system (4.6) is observable in the time interval $\tau = [k, (k+3)]$ if the observability matrix $\mathbf{O}(\tau) = [\mathbf{C}; \mathbf{CA}(k); \mathbf{CA}(k+1)\mathbf{A}(k); \mathbf{CA}(k+2)\mathbf{A}(k+1)\mathbf{A}(k)]$ has rank 4 [4]. We find sufficient conditions for observability by finding conditions under which the matrix has four linearly independent rows. Let the matrix $\mathbf{O}_{\text{sub}}$ be made up of rows $\{1, 2, 3, 5\}$ of $\mathbf{O}$. Then $\mathbf{O}$ has four linearly independent rows if $\mathbf{O}_{\text{sub}}$'s determinant is non-zero. The

determinant is given by

$$\det \mathbf{O}_{\text{sub}}(\tau) = \overline{p}_{\text{on}}^{3}\Big(u_{\text{on}}(k)(u_{\text{off}}(k)(1-f) - f u_{\text{on}}(k+1)) + u_{\text{off}}(k)(g u_{\text{off}}(k+1) + u_{\text{on}}(k+1)(g-1))\Big).$$

Recall that in each time step either $u_{\text{off}} = 0$ or $u_{\text{on}} = 0$. Thus, there are four conditions for a non-zero determinant:

1. $u_{\text{off}}(k+1) \neq 0$ & $u_{\text{on}}(k) \neq 0$, if $u_{\text{on}}(k+1) = 0$ & $u_{\text{off}}(k) = 0$;
2. $u_{\text{on}}(k+1) \neq 0$ & $u_{\text{off}}(k) \neq 0$, if $u_{\text{off}}(k+1) = 0$ & $u_{\text{on}}(k) = 0$;
3. $u_{\text{off}}(k+1) \neq 0$ & $u_{\text{off}}(k) \neq 0$, if $u_{\text{on}}(k+1) = 0$ & $u_{\text{on}}(k) = 0$;
4. $u_{\text{on}}(k+1) \neq 0$ & $u_{\text{on}}(k) \neq 0$, if $u_{\text{off}}(k+1) = 0$ & $u_{\text{off}}(k) = 0$.

In summary, a sufficient condition for system observability is for either u_{on} or u_{off} to be non-zero in each time step. Note, the internal switching probabilities f and g cannot cause the determinant to be zero because we assume TCLs are cycling through their temperature range, which implies f and g are not equal to 0 or 1 by definition.

We use a time-varying Kalman Filter [53] to estimate the states of the stochastic model described by (4.6). Process noise $\mathbf{w}$ and measurement noise $\mathbf{v}$ represent plant-model mismatch and unknown disturbances. We assume that $\mathbf{w}$ and $\mathbf{v}$ are zero-mean, Gaussian, white noise processes with covariance $\mathbf{Q}$ and $\mathbf{R}$, respectively.

4.4.3 Implementation

In the case of full feedback, the proposed controller does not use all of the information available to it in every time step. Instead, the Kalman Filter is used to estimate the state, except for time steps in which the percentage blocked changes, i.e., $\beta(k) \neq \beta(k-1)$. At these time steps, the Kalman Filter is bypassed and the state estimate is updated according to $\hat{\mathbf{x}}(k) = [\sum \sigma^i(k)(1-b^i(k)), \sum(1-\sigma^i(k))(1-b^i(k)), \sum \sigma^i(k)b^i(k), \sum(1-\sigma^i(k))b^i(k)]$, where all summations are from $i = 1$ to N.

In the case of moderate feedback, the proposed controller uses information on the percentage blocked to improve its state estimate. At time steps when $\beta(k) \neq \beta(k-1)$, the Kalman filter produces the state estimate $\hat{\mathbf{x}}(k)$ as usual, and then we update the estimate such that the percentage of TCLs estimated to be blocked equals $\beta(k)$. The update is given by $\hat{\mathbf{x}}(k) = [(1-\beta(k))(\hat{x}_1(k) + \hat{x}_3(k)), (1-\beta(k))(\hat{x}_2(k) + \hat{x}_4(k)), \beta(k)(\hat{x}_1(k) + \hat{x}_3(k)), \beta(k)(\hat{x}_2(k) + \hat{x}_4(k))]$.

4.4.4 Benchmark Controller

The benchmark controller computes switching probabilities with (4.4)-(4.5) but does not use a Kalman Filter to estimate $\hat{P}_{\text{on}}$ and $\hat{P}_{\text{off}}$. Instead, with full feedback, the benchmark controller estimates are given by $\hat{P}_{\text{on}}(k) = \sum_{i=1}^{N} P_{\text{R}}^{i}\sigma^{i}(k)(1-b^{i}(k))$ and $\hat{P}_{\text{off}}(k) = \sum_{i=1}^{N} P_{\text{R}}^{i}(1-\sigma^{i}(k))(1-b^{i}(k))$. With moderate feedback, the benchmark controller estimates are given by $\hat{P}_{\text{on}}(k) = (1-\beta(k))P_{\text{total}}(k)$ and $\hat{P}_{\text{off}}(k) = (1-\beta(k))(\sum_{i=1}^{N} P_{\text{R}}^{i} - P_{\text{total}}(k))$. Finally, with minimal feedback, the controller uses the same equations as with moderate feedback, but $\beta(k)$ is set to 0 because the aggregator does not have any information about blocking.

4.5 Simulation Study

4.5.1 Setup

We test controller performance using 1-hour simulations of 1000 TCLs controlled to tracking a frequency regulation signal. A constant set of TCLs is blocked between minutes 20 and 40. We test each controller in 12 different scenarios, where a scenario is defined by a combination of blocking level (0%, 20%, 40%, or 60% of the population) and feedback level (see Table 4.1).

We run 24 randomized trials for each scenario. Each trial has: a different random instance of TCL parameters, different random numbers generated during probabilistic dispatch, and a different frequency regulation signal. TCL parameters are drawn from the uniform distributions described in Table 2.1. For the frequency regulation signal, Trial 1 uses the first 1-hour segment of PJM's RegD signal [68] from June 3, 2018, Trial 2 uses the second segment, and so on. In all trials, the ambient temperature is 32°C, and the TCL aggregation's regulation capacity is set to +/-20% of baseline power consumption.

For these simulations, we use a persistence model to predict the total power of the aggregation in the next time step (i.e., $P_{\text{total}}(k+1|k) = P_{\text{total,meas}}(k)$ in (4.4)-(4.5)) because using the aggregate model (4.6) for prediction would result in poor accuracy. In addition, we do not add noise to the measurement of P_{total}, i.e., $P_{\text{total,meas}}(k) = P_{\text{total}}(k)$.

Values for tuning parameters are listed in Table 4.2. The parameters for each combination of controller and feedback level are tuned separately, although sometimes are equal. For the proposed controller, we iteratively tune $\mathbf{Q}$ and $\mathbf{R}$ with $K = 1$. Then, with $\mathbf{Q}$ and $\mathbf{R}$ fixed for the proposed controller, we sweep through values for K from 0.95 to 1.05 for both controllers. We select the parameter values that yield the lowest sum of root-mean-square errors (RMSE)

Table 4.2: Tuning Parameters

Controller	Scenario	K_{P}	$\mathbf{W}_{\mathrm{b}}$	$\mathbf{V}_{\mathrm{b}}$
Benchmark	F, P	0.96	–	–
Benchmark	N	1.03	–	–
Proposed	F, P	0.96	2.5diag(3, 3, 1, 1)	diag(1, 0)
Proposed	N	1	9.3diag(1, 7, 1, 7)	diag(1, 0)

diag$(\cdot)$ maps an n-tuple to the corresponding nxn diagonal matrix.

in tracking, across Trials 1-4 and all blocking levels.

Prior to the test hour, we run the simulation for an hour to ensure steady-state conditions at the start of the test. We also use this pre-test hour to calculate the aggregation's baseline power consumption, the $\mathbf{A}$ matrix's transition probabilities, and the parameter $\overline{p}_{\mathrm{on}}$. For simplicity, we set $\overline{p}_{\mathrm{off}} = \overline{p}_{\mathrm{on}}$.

4.5.2 Results

Estimation Performance

The proposed controller's Kalman Filter is able to estimate the number of TCLs that are blocked, even when no explicit blocking information is available. Performance of the Kalman Filter during Trial 4 is shown in Fig. 4.3. State estimates are more accurate when feedback on percent blocked is available (left column) than when no feedback on blocking is available (right column), as would be expected. Step changes in the percentage blocked occur at minutes 20 and 40; in the case of minimal feedback, these step changes cause a lag in the state estimates.

Tracking Performance

Fig. 4.4 shows the proposed and benchmark controllers tracking the frequency regulation signal while 60% of TCLs are blocked in Trial 2. We evaluate performance in terms of percent RMSE, which is the error in total power consumption as a percentage of the aggregation's average baseline power consumption. In this trial, the proposed controller improves upon the benchmark controller at every level of feedback. The decrease in RMSE is largest in the case of minimal feedback.

Table 4.3 reports the average RMSE in tracking for each scenario, with the average taken

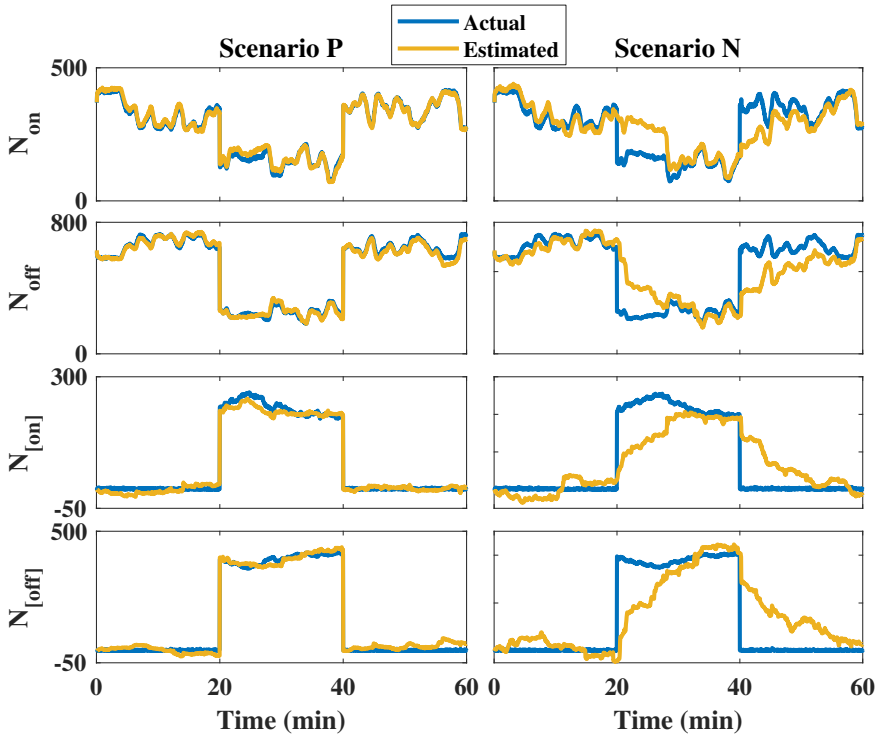


Figure 4.3: Estimated states and actual states with the proposed controller when partial information on blocking is known (left) and when no information on blocking is known (right). (Data is from Trial 4.)

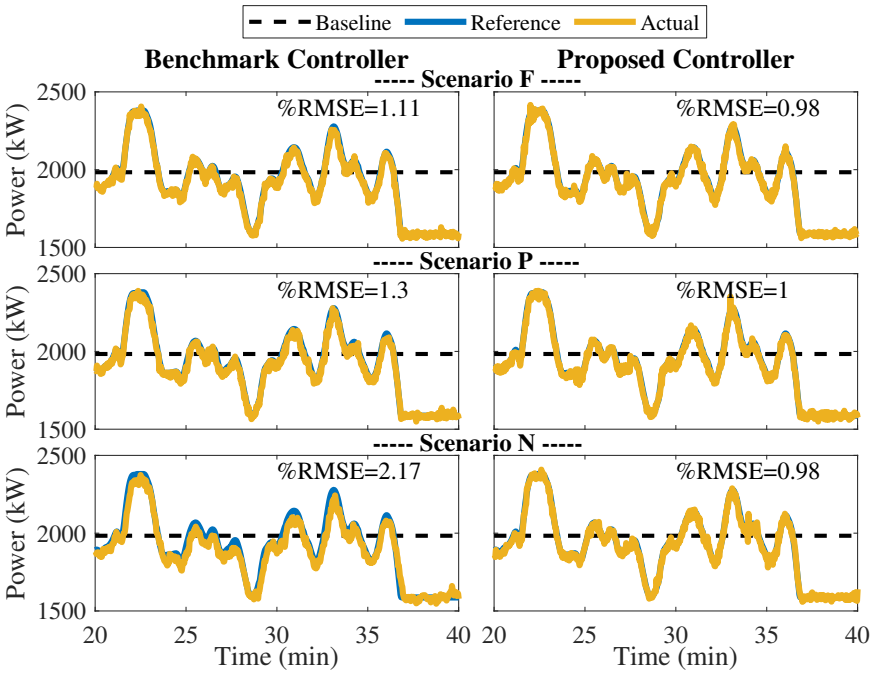


Figure 4.4: Benchmark and proposed controllers tracking the frequency regulation signal during the blocking period of Trial 2 with 60% blocked. The proposed controller improves upon the benchmark controller in all three scenarios.

across the 24 trials. In every scenario, the proposed controller's average RMSE is less than or equal to that of the benchmark controller. On average, error increases as the percentage blocked increases and as the level of feedback decreases. The proposed controller's improvement over the benchmark is not very large in most scenarios, with a decrease in average percent RMSE of $\leq 0.03\%$ in 10 of 12 scenarios. The largest improvements occur in the remaining two scenarios when there is minimal feedback and blocking is at 40% and 60%. In these scenarios, average percent RMSE decreases by 0.09% and 0.29%, respectively.

Box plots of the 24 trials across all scenarios are shown in Fig. 4.5. The proposed controller generally reduces the maximum and median RMSE across the 24 trials. In all scenarios, the maximum error decreases with the proposed controller; in all but two scenarios, the median error also decreases. As with the mean error results, the box plots show that the proposed controller yields the biggest improvements when 40% or 60% of TCLs are blocked and minimal feedback is available.

It is somewhat surprising that the proposed controller outperforms the benchmark controller when full feedback is available. This result may be because the benchmark controller lacks information on how many TCLs are outside of their set temperature range and thus

Table 4.3: Controller Performance: Average % RMSE in Tracking

Percent Blocked	Scenario F		Scenario P		Scenario N	
	BM	PP	BM	PP	BM	PP
0%	0.76	0.75	0.76	0.75	0.77	0.75
20%	0.77	0.75	0.77	0.76	0.78	0.75
40%	0.78	0.76	0.79	0.77	0.86	0.77
60%	0.86	0.86	0.89	0.86	1.22	0.93

BM = Benchmark controller and PP = Proposed controller

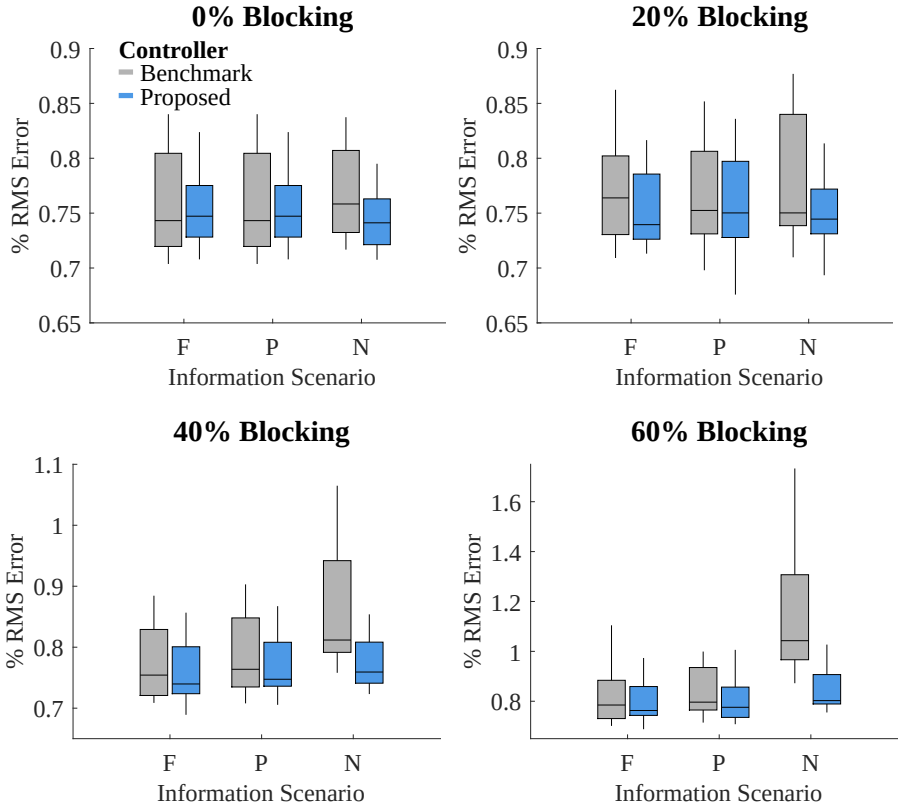


Figure 4.5: Box plots of tracking errors in 24 trials for each blocking level and information scenario. Box diagram indicates median (horizontal line), interquartile range (box), and maximum and minimum points (whiskers). (Note: outliers are not plotted, and not all y-axis scales are equal.)

unavailable to switch. In contrast, the proposed controller has some information of this nature because its Kalman Filter estimates how many TCLs are generally unavailable, whether it is due to blocking or out-of-range temperature.

4.6 Chapter Conclusion

We have proposed two coordination frameworks—"operator centric" and "aggregator centric"—that enable network-aware load control while maintaining the independence of the aggregator and the distribution operator. Working within the operator-centric framework, we proposed a specific scheme in which the operator has the authority to block aggregator commands in order to protect network reliability. We then designed a controller for the aggregator that successfully estimates and compensates for the number of loads that are blocked. In future work, we plan to develop a strategy for the operator that it is maximally permissive of the aggregator's control and includes a real-time assessment of network reliability.

Chapter 5

Mode-Count Control

This chapter is largely based on the published work [70][5].

5.1 Chapter Introduction

In this chapter, we propose a control strategy that restricts the range in power consumption of a group of TCLs in order to reduce the TCLs' impact on at-risk distribution constraints. For a particular group of TCLs, the strategy limits the number of TLCs that are on at a particular time, referred to as the group's on-count. Specifically, the strategy switches individual TCLs such that the group's on-count always satisfies an upper bound and a lower bound. The strategy is based on the "mode-count control" policy developed in [61, 62] for a group of one-dimensional two-mode switched systems. In our application, the switched system is a TCL, which has one continuous dimension (temperature) and two discrete modes (on and off).

We propose using mode-count control to reduce the aggregate variability of a TCL group's power consumption, and to thereby relieve at-risk constraints on the distribution network. A group's variability can be reduced by setting the upper and lower bounds on the group's on-count as close together as possible. There has been related work, in terms of reducing on/off synchronization of TCLs, for the prevention of power oscillations or rebound after a period of load reduction [78, 92]. However, these strategies control TCLs by adjusting temperature setpoints or ranges; in contrast, we control TCLs by switching them on/off,

[5]S.C. Ross, P. Nilsson, N. Ozay, and J.L. Mathieu. "Managing voltage excursions on the distribution network by limiting the aggregate variability of thermostatic loads". In: *Proceedings of the American Control Conference (ACC)*. Philadelphia, PA, July 2019, (Invited).

and do so within the user-set temperature range so that the control is non-disruptive to the end-user [8].

Prior work on non-disruptive control of TCLs has focused on aggregating hundreds to thousands of TCLs to provide transmission-level services (e.g., [3, 52, 96]), rather than providing distribution-level services. Distribution voltage constraints are considered within [12, 17, 93], but their centralized optimization-based algorithms are computationally intensive, considering all network constraints rather than targeting at-risk constraints as proposed here.

The contributions of this chapter are as follows:

- We propose a control strategy, using the theory developed in [61, 62], to reduce the impacts of TCLs' variable power consumption on at-risk distribution network constraints.
- We examine the effectiveness of the control strategy for different sizes of TCL groups.
- We apply the proposed control strategy to the problem of distribution voltage management for a long line with substantial PV generation at its end bus.
- Finally, we extend the control strategy to the case in which TCLs have cycling constraints.

5.2 Background

This section summarizes the theory developed in [61, 62] as it applies to TCLs. We refer the reader to [61, 62] for full details.

5.2.1 Counting Problem for TCLs

The control objective of [61, 62] is to maintain a TCL group's on-count between an upper bound and lower bound, while also satisfying each TCL's individual temperature constraints $\underline{\theta}^i \leq \theta^i \leq \overline{\theta}^i$. The underlying temperature dynamics are the same as (2.1), with the ambient temperature assumed constant. The notation of equation (2.1) is simplified to

$$\frac{\mathrm{d}}{\mathrm{d}t}\theta^i(t) = \begin{cases} f^i_{\mathrm{on}}(\theta^i(t)) & \text{if } \sigma^i(t) = on, \\ f^i_{\mathrm{off}}(\theta^i(t)) & \text{if } \sigma^i(t) = off. \end{cases} \tag{5.1}$$

Formally, the on-count of a group of N TCLs is

$$H(t) = \sum_{i=1}^{N} \mathbb{1}_{\{\text{on}\}}(\sigma^i(t)), \tag{5.2}$$

and is controlled between the lower bound $\underline{H}$ and the upper bound $\overline{H}$:

$$\underline{H} \leq H(t) \leq \overline{H}. \tag{5.3}$$

5.2.2 Switching Strategy

The proposed control strategy determines when a TCL's power mode should be switched. The switching policy uses time rather than temperature to determine which TCL to switch next. In the definitions that follow, we refer to cooling TCLs (e.g., ACs) simply as TCLs. The policy ranks TCLs in terms of their "time-to-upper-limit", defined as the time it takes for the a TCL to travel in the off-mode from its current temperature $\theta^i(t)$ to its upper limit $\overline{\theta}^i$, and their "time-to-lower-limit", defined as the time to travel in the on-mode from $\theta^i(t)$ to the lower limit $\underline{\theta}^i$. Let the output of the function $T^i_{\text{on}}(\theta_1, \theta_2)$ be the time it takes the ith TCL to progress from θ_1 to θ_2 in the on-mode, and let the function $T^i_{\text{off}}(\theta_1, \theta_2)$ be similarly defined but for the off-mode. The control strategy proposed in [62] is below.

Original Strategy for a Cooling TCL [62]

- If $T^i_{\text{off}}(\theta^i(t), \overline{\theta}^i) = 0$ and TCL i is *off*, switch it *on*.
- If $T^i_{\text{on}}(\theta^i(t), \underline{\theta}^i) = 0$ and TCL i is *on*, switch it *off*.
- If $H(t^+) < \underline{H}$ for $t^+ > t$, switch *on* the TCL in off-mode with the largest time-to-lower-limit. Repeat this step if the constraint continues to be violated.
- If $H(t^+) > \overline{H}$ for $t^+ > t$, switch *off* the TCL in on-mode with the largest time-to-upper-limit. Repeat this step if the constraint continues to be violated.

5.2.3 Conditions on Feasible Count Bounds

In [62], the authors derive conditions for values of $\underline{H}$ and $\overline{H}$ that can be satisfied indefinitely by the proposed switching strategy. For a group of N TCLs, the condition for lower bound

values that can be satisfied indefinitely is

$$\underline{H} \;<\; \sum_{i=1}^{N} \frac{-f^i_{\text{off}}(\underline{\theta}^i)/f^i_{\text{on}}(\underline{\theta}^i)}{1 - f^i_{\text{off}}(\underline{\theta}^i)/f^i_{\text{on}}(\underline{\theta}^i)}, \tag{5.4}$$

and the condition for upper bound values is

$$\overline{H} \;>\; N - \sum_{i=1}^{N} \frac{-f^i_{\text{on}}(\overline{\theta}^i)/f^i_{\text{off}}(\overline{\theta}^i)}{1 - f^i_{\text{on}}(\overline{\theta}^i)/f^i_{\text{off}}(\overline{\theta}^i)}. \tag{5.5}$$

Moreover, these constraints are shown to be tight. That is, no switching strategy can achieve a greater lower bound or a smaller upper bound and guarantee that the bound will be satisfied indefinitely.

5.3 Methods Part 1: Reducing the Variability of a TCL Group's Power Consumption

5.3.1 Reducing the Variability of a TCL Group's Power Consumption

We reduce the variability of a TCL group's power consumption by minimizing the variability of the group's on-count. We use mode-count control, and select on-count bounds that maximally constrain the on-count. Specifically, the bounds are selected to minimize the difference $(\overline{H} - \underline{H})$ and to ensure feasibility (i.e., $(\overline{H} - \underline{H}) \geq 0$). For large enough group sizes, we can set $\underline{H} = \overline{H}$ because the greatest lower bound (GLB) that satisfies (5.4) is greater than the least upper bound (LUB) that satisfies (5.5). Note the GLB is equal to the right hand sides of (5.4), rounded down to the nearest integer, and the LUB is equal to the right hand side of (5.5), rounded up to the nearest integer. Figure 5.1 shows the (unrounded) GLB and LUB for TCL groups of different sizes; each group consists of heterogeneous ACs with parameters drawn from Table 2.1.

For a homogeneous group of TCLs, controlling the on-count between $\underline{H}$ and $\overline{H}$ also constrains the group's aggregate power consumption P_{grp} such that

$$\underline{H} P_{\text{R}} \leq P_{\text{grp}}(t) \leq \overline{H} P_{\text{R}}, \tag{5.6}$$

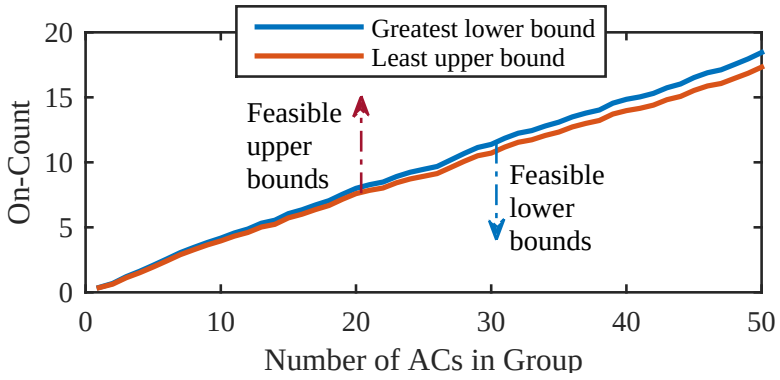


Figure 5.1: The GLBs and LUBs that can be satisfied indefinitely for different AC-group sizes. The set of feasible upper bounds lie above the LUB line, and the set of feasible lower bounds lie below the GLB line. Bounds were calculated for a particular set of heterogeneous ACs.

where P_{R} is the TCLs' individual power rating. For a group of TCLs with heterogeneous power ratings, a count bound can be satisfied by different sets of TCLs within the group, and thus can result in different aggregate power levels. Let $\underline{P}_{\mathrm{grp}}(\underline{H})$ be the group's minimum power consumption while satisfying the lower count bound, and let $\overline{P}_{\mathrm{grp}}(\overline{H})$ be the group's maximum power consumption while satisfying the upper count bound. We find an expression for these power bounds by first ordering the TCLs in the group by power rating from least to greatest with index j, i.e., $P_{\mathrm{R}}^{j} \leq P_{\mathrm{R}}^{j+1}$. Given a group of size N and count bounds $\overline{H}$ and $\underline{H}$, the group's power consumption will be bounded between that of the $\underline{H}$ lowest-power-rating TCLs and the $\overline{H}$ highest-power-rating TCLs, i.e.,

$$\underline{P}_{\mathrm{grp}} = \sum_{j=1}^{\underline{H}} P_{\mathrm{R}}^{j} \tag{5.7}$$

$$\overline{P}_{\mathrm{grp}} = \sum_{j=N-\overline{H}+1}^{N} P_{\mathrm{R}}^{j}. \tag{5.8}$$

5.3.2 Numerical Results

We use Monte Carlo methods to assess the effectiveness of using mode-count control to reduce the variability of a TCL group's power consumption. We compare the variability of TCL groups in two cases: 1) the group is not controlled, and 2) the group is controlled with the mode-count policy, with upper and lower bounds set as close together as possible. We run 100 simulations for each of eight group sizes, which range from 5 to 1000 ACs. For

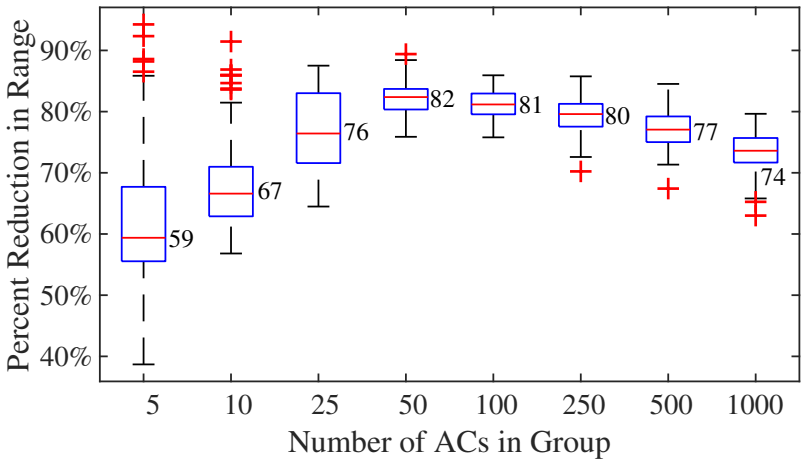


Figure 5.2: Reduction in range of power when the Original Strategy is used, across different size groups of ACs. Each box shows the statistics (25th percentile, median, 75th percentile) for 100 Monte Carlo runs at each group size. Medians are also shown numerically. Red + signs are outliers.

each simulation, AC parameters and initial conditions are re-sampled from their respective random distributions. AC parameters are drawn from Table 2.1. All simulations have the following settings: the simulation duration is 12 hours; the outdoor temperature is 32°C; and one third of ACs are initialized in the on-mode. We simulate the ACs for an hour prior to the test to ensure steady-state conditions have been reached.

Figure 5.2 shows that the control strategy reduces the range of aggregate power for each group size. The reduction in range is relative to that of the uncontrolled TCL groups; each box plot represents the distribution across the group size's 100 Monte Carlo simulations. For group sizes under 50, the median percent reduction in range decreases with decreasing group size because setting $\underline{H} = \overline{H}$ is infeasible for some of the groups. Instead, for these smaller groups, the on-count is maintained between two distinct bounds, for example, $\overline{H} = 4$ and $\underline{H} = 3$ for a group of 10. For group sizes of 50 or more, the range in the on-count decreases by 100% because setting $\underline{H} = \overline{H}$ is feasible for all groups of this size. However, the range in power cannot decrease by 100% because of TCLs' heterogeneous power ratings (see (5.7)-(5.8)). For group sizes greater than 50, the controlled range in power grows with increasing group size, which results in a slow decline in the percent reduction in range. Thus, we find that the control strategy is most effective at reducing the range in power of medium sized groups, but, for all group sizes tested, one can expect at least a 40% reduction in range.

5.3.3 Application to Scenario with High Penetration of PV

As distribution-level photovoltaic (PV) generation continues to grow, so too does the possibility of large, fast changes in net load. Aggregate PV power generation can drop by 60% in less than 30 seconds when a cloud passes over a neighborhood [38]. There is also substantial variation in the aggregate power consumption of small groups of thermostatically controlled loads (TCLs) because of random periods of synchronization in their on/off power cycles. In an area of a network with high penetration of PV and TCLs, a simultaneous drop in PV generation and increase in TCL consumption could result in under-voltages, particularly if the systems are located at the end of a long distribution line.

Conventionally, distribution voltages are regulated with on-load tap-changing transformers, voltage regulators, and capacitor banks. However, these techniques have not been designed to mitigate fast, repetitive excursions in voltage; for example, tap changers typically have built-in delays to avoid responding to transient conditions [39]. New forms of voltage management are needed in network areas with highly variable power injections and high voltage sensitivity to these variations (e.g., at nodes far from the substation with high TCL and PV penetration).

We propose mode-count control of TCLs as a new tool for voltage management. The idea is simple: The variation of voltage at a bus can be reduced by reducing the variation of net load at the bus, which can be reduced by reducing the variation of TCL-load at the bus. In this subsection, first we show the voltage problem that can arise when highly variable PV generation and a group of ACs are co-located at the end of a long distribution line. Then we use mode-count control to reduce the TCL group's variability and show a corresponding reduction in voltage variability.

Variation in PV Power Generation

If multiple PV systems are located in close proximity, changes in power generation due to transient clouds will occur nearly simultaneously and could cause voltages to fluctuate in that area of the distribution network. We use the following model and simulation to demonstrate this effect.

The PV model is based on that of [18] and estimates PV power output as a function of solar irradiance. Neglecting temperature effects, PV power output can be calculated as

$$P_{\mathrm{pv}}(t) = \eta_{\mathrm{inv}} D P_{\mathrm{dc0}} S(t)/S_{\mathrm{ref}}. \tag{5.9}$$

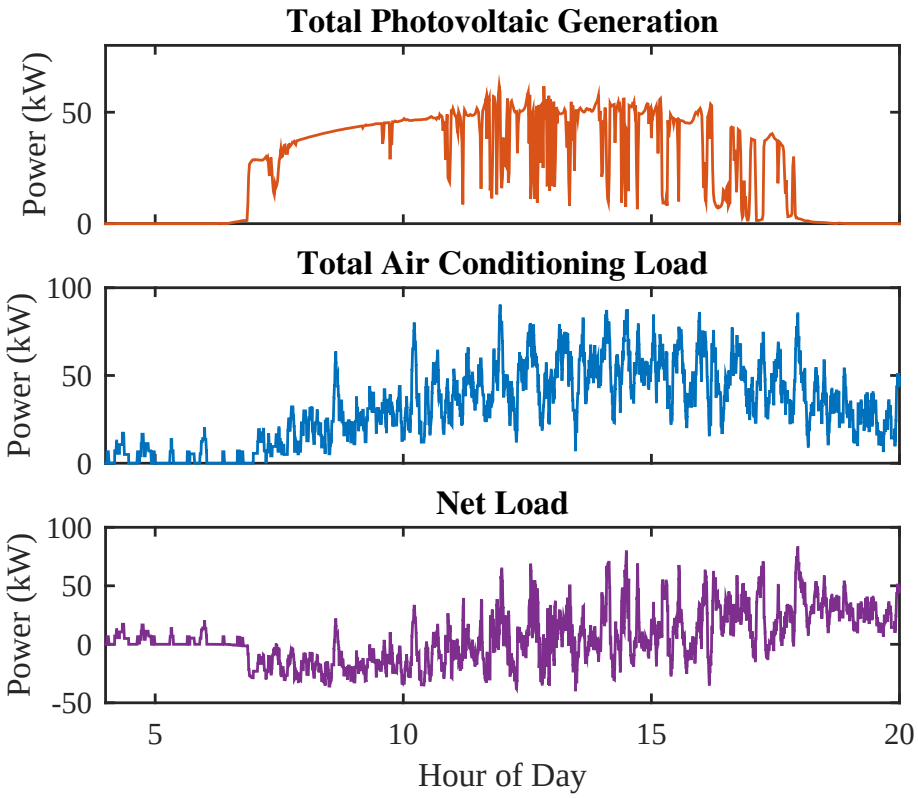


Figure 5.3: Simulated power profiles of PV generation, AC consumption, and net load. Top: The PV generation profile drops sharply due to partial clouding. Middle: Variation in the ACs' power consumption is due to random periods in which ACs are synchronized. Bottom: The variation of the net load is larger than that of either resource alone.

Parameters are defined in Table 5.1 with values sourced from [18, 23]. The variable S is the incident irradiance on the PV surface; we approximate the incident irradiance as the sum of direct normal irradiance and diffuse horizontal irradiance, which should overestimate PV production [18]. Lastly, we assume PV systems are controlled to inject power at a unity power factor.

Using solar irradiance data from Oak Ridge National Lab [54], we simulate the total output power of 12 residential PV systems on a partially cloudy day (see Fig. 5.3 (top)). The irradiance data has 1 minute resolution and was measured in Oak Ridge, TN on Sept. 15, 2018 [54]. As shown in Fig. 5.3, partial clouds can cause PV generation to drop quickly; for example, at 12:42 PM the power output of the PV systems decreases by 82% in less than 1 minute.

Table 5.1: Photvoltaic Parameters

Parameter	Values	Unit
Inverter efficiency (η_{inv})	96%	–
Derating factor (D)	0.86	–
Rated DC power capacity (P_{dc0})	5.7	kW
Reference irradiance for rated power (S_{ref})	1000	W/m^2

Variation in a TCL Group's Power Consumption

A group of TCLs co-located at a distribution node can also cause large variations in power injections due to random periods of synchronization. We use the following model and simulation to demonstrate this effect.

We model a group of heterogeneous ACs with equations (2.2)-(2.4) and with parameters drawn randomly from Table 2.1. We assume that, when an AC is on, it is a constant power load with a power factor of 0.97. (Power factor values for residential loads are from Table A.2 of [15].)

We simulate a group of 25 heterogeneous ACs with outdoor temperature varying over the course of the day (see Fig.5.3 (middle)). The outdoor temperature data has 1 minute resolution and is from the same day and location as the solar data [54]. Variations in power are of a similar magnitude and frequency as those of the PV systems; for example, at 1:22 PM power consumption by ACs decreases by 89% in 6 minutes.

Net Load Variation

When groups of ACs and PVs are co-located at the same node, the variation of the net power injection can be greater than that of either resource independently. The bottom plot of Fig. 5.3 shows the net load profile, which is just the difference of the PV profile (top) and the AC profile (middle). The total range in power for the three profiles in Fig. 5.3 is 62 kW, 90 kW, and 122 kW for the PV, AC, and net load profiles, respectively. Figure 5.4 shows the range of the net load in each hour of the simulated day; the range is largest in the mid-afternoon when large variations in PV generation and AC consumption coincide. Note, in Fig. 5.4, range is calculated as the difference between the 5th and 95th percentiles.

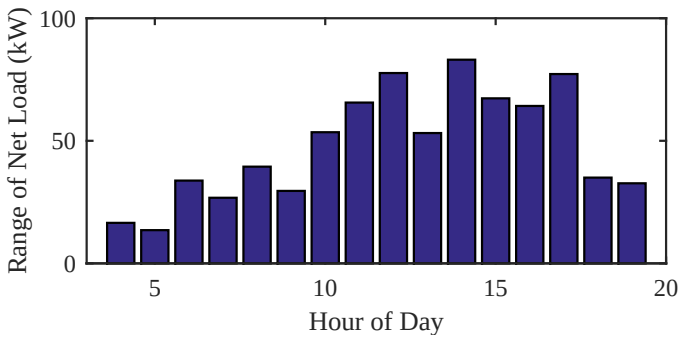


Figure 5.4: Range of net load over each hour of the simulated day. The range is largest in the afternoon when PV generation and AC consumption both have large variations.

Voltage Excursions

Large and fast variations in net load can cause voltage excursions at the end of a long distribution line. We use the following model and simulation to demonstrate this effect.

As shown in Fig. 5.5, we model a distribution line that connects two nodes: node 1, an infinite bus with voltage of 1.0 p.u., and node 2, a PQ bus that supplies 25 households. We assume twelve of these households also have a PV system. Each household has an AC and a constant power load that aggregates all other loads. The constant power load draws 1.5 kW of real power and has a power factor of 0.95. The PV systems and ACs are modeled as described in Sections 5.3.3 and 5.3.3.

We use an intentionally simple distribution system model so that we can fully interpret the results. The line is single phase, and we omit distribution transformers between node 2 and the loads. Node 1 models a substation with voltage regulation. Given these assumptions, we derive the following power flow equations:

$$P_{\text{net}} = V_1 V_2 (g_{12} \cos \phi_{21} + b_{12} \sin \phi_{21}) - g_{12} V_2^2 \tag{5.10}$$

$$Q_{\text{net}} = V_1 V_2 (g_{12} \sin \phi_{21} - b_{12} \cos \phi_{21}) + b_{12} V_2^2, \tag{5.11}$$

where P_{net} and Q_{net} are the net power injections at node 2, V_1 and V_2 are the nodes' voltage magnitudes, ϕ_{21} is the difference in voltage angle between the two nodes ($\phi_2 - \phi_1$), and parameters g_{12} and b_{12} are the line's conductance and susceptance, respectively. The line is 10 miles long and consists of a single phase and neutral, both of which have a 180-ampere conductor with parameters from a feeder model provided by [63]. We calculate the line's per mile impedance to be $1.86 + 1.41j$ using the "Modified Carson's Equations" (see Chapter 4

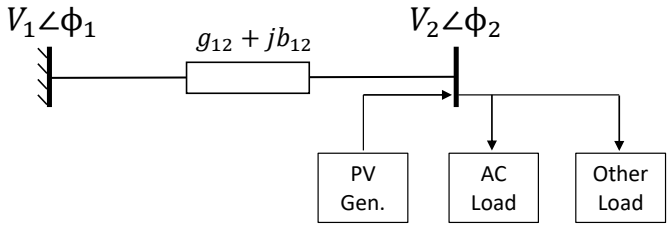


Figure 5.5: Diagram of distribution line between infinite bus (node 1) and load bus (node 2). The line is single phase, 10 miles long, and supplies a group of 25 households.

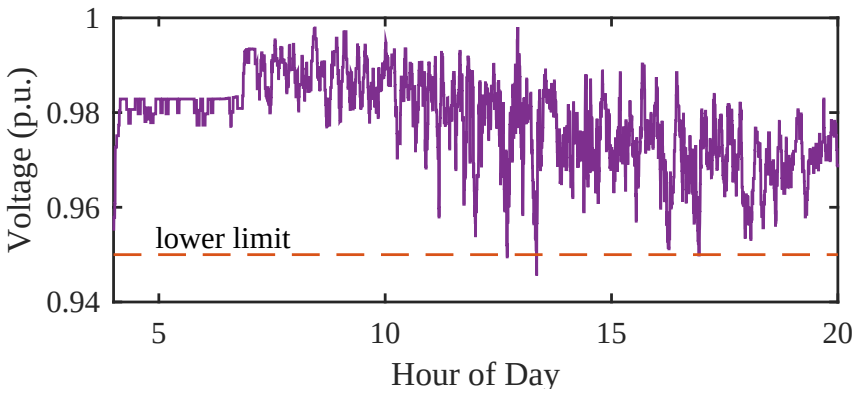


Figure 5.6: Voltage profile at node 2 on the simulated distribution line. Voltage excursions reach below the 0.95 p.u. lower limit during large positive variations in net load.

of [39]).

We simulate the power flow along the modeled distribution line using the same daily weather data as in previous sections. We solve for node 2's voltage in each time step by applying the Newton Raphson method to (5.10)-(5.11). The voltage profile of node 2 is shown in Fig. 5.6. In the middle of the day, large variations in the net power injection at node 2 cause voltage excursions below 0.95 p.u. (the national standard for service voltage is 0.95-1.05 p.u. [59]).

In a real system, the under-voltages in Fig. 5.6 could be prevented by increasing the setpoint of the voltage regulator at node 1 above 1 p.u. However, a higher setpoint could result in over-voltages on a sunny day with minimal load. Because of this trade-off, we expect that some systems' voltage regulation schemes will not be able to prevent all possible voltage excursions, and excursions similar to those in Fig. 5.6 will sometimes occur.

Voltage Management with Mode-Count Control

Through simulation, we demonstrate the control strategy's ability to constrain the aggregate power of an AC group and thereby reduce voltage excursions on a network. We consider two scenarios: the "Uncontrolled Scenario" in which there are no external switching commands, and the "Controlled Scenario" in which mode-count control is used. We set the upper and lower count bounds as close together as possible: in this case, $\overline{H} = \underline{H} = 10$. The simulations use the same models, parameters, and weather data as previously described. The 1-hour simulations start at 14:00, which is the hour with the largest range in net load (see Fig. 5.4). Initial temperatures are drawn from the uniform distribution between an AC's upper and lower temperature limit, and the first 10 ACs are initialized to the on mode.

Fig. 5.7 compares the results for the two scenarios. The total demand of the AC group is much less variable in the Controlled Scenario than the Uncontrolled Scenario (top plot), which results in less variability in net load (second plot), which in turn results in less variability in voltage (third plot). The bottom plot shows that variation in the AC group's average temperature is very small, both with and without control.

5.4 Methods Part 2: Mode-Count Control Considering Cycling Constraints

5.4.1 Cycling Constraints

A drawback of the strategy proposed in [62] is the possibility of cycling a TCL too frequently, which could irritate the end-user or could damage the TCL's compressor [67]. In this section, we assume TCLs have cycling constraints and propose a new switching strategy that can satisfy these new constraints. Cycling constraints are a current thrust of research on TCL control, and our proposed control strategy adds to a growing body of work including [10, 82, 99]. Note cycling constraints are often referred to as lockout constraints in the literature.

Cycling constraints can be defined in terms of a minimum time that must elapse before a TCL can be switched again, referred to as a lockout period. Given a homogeneous lockout

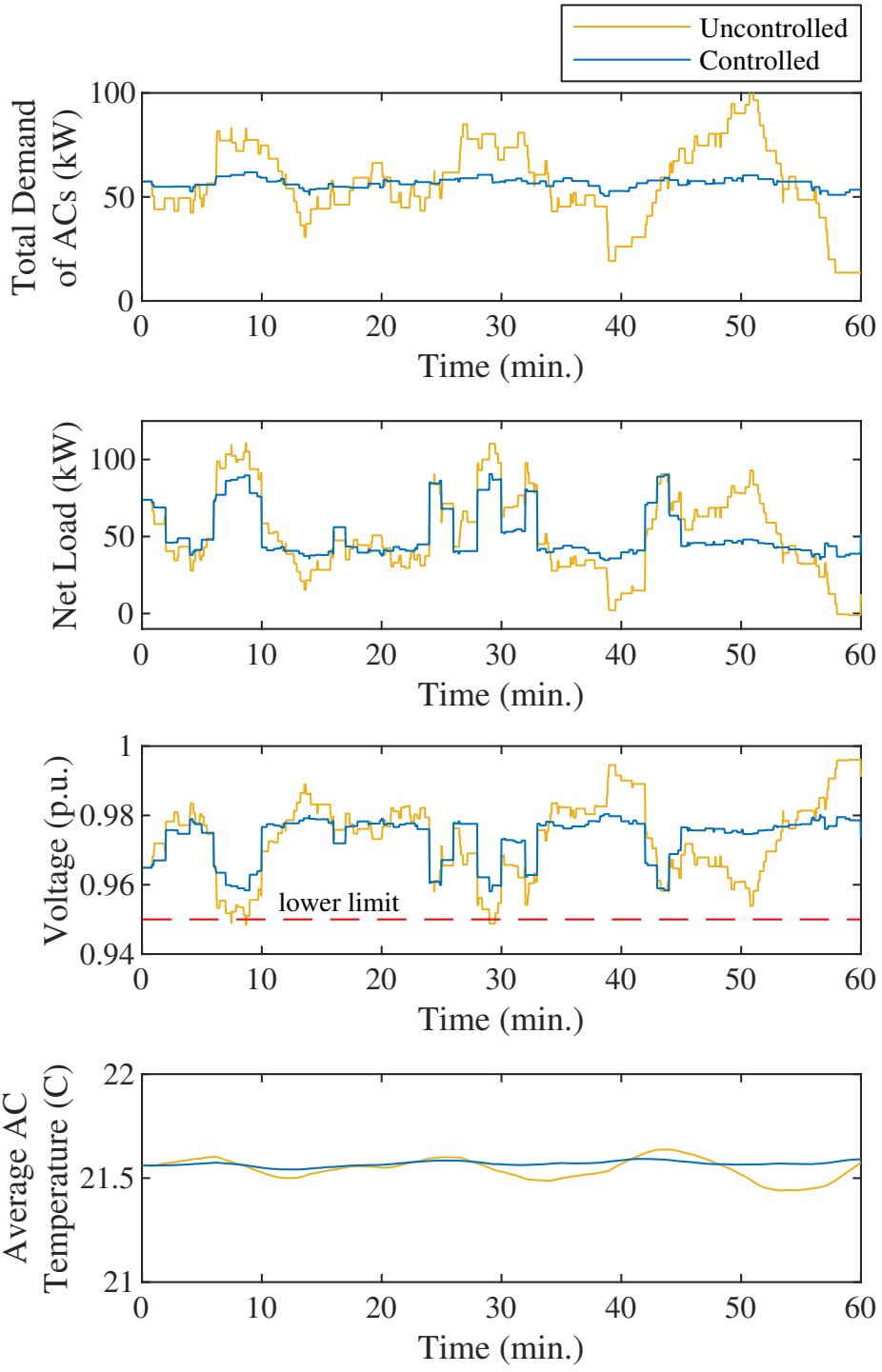


Figure 5.7: Comparison of loads and voltage with and without control. The control strategy reduces the variation in the aggregate power consumption of ACs, which reduces net load variation and results in a reduction in voltage excursions. The average temperature of the AC group stays relatively constant in both cases.

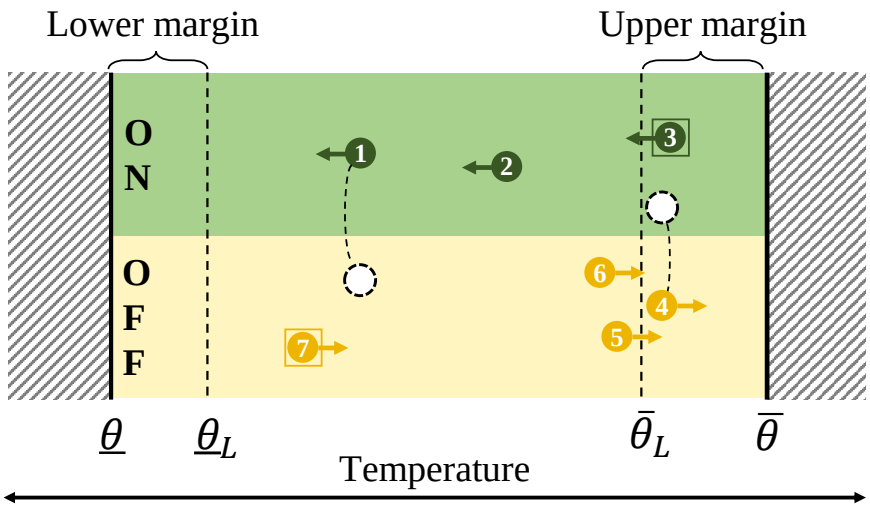


Figure 5.8: Illustration of switching strategy for TCLs with cycling constraints. TCLs are represented by numbered circles; outer boxes indicate the TCL is locked.

period of t_{L} for all TCLs, the cycling constraints for the ith TCL are

$$\lambda^i(t) = \begin{cases} \textit{locked} & \text{if } \int_{t-t_{\mathrm{L}}}^{t} \mathbb{1}_{\{\sigma^i(t)\}}(\sigma^i(\tau))d\tau < t_{\mathrm{L}}, \\ \textit{unlocked} & \text{otherwise.} \end{cases} \tag{5.12}$$

If a TCL is locked, then its power mode cannot be switched. This condition is enforced by

$$\sigma^i(t) = \sigma^i(t^-) \quad \text{if} \quad \lambda^i(t) = \textit{locked}. \tag{5.13}$$

5.4.2 Switching Strategy

Given a group of TCLs with cycling constraints, we propose a switching strategy that satisfies the cycling constraints (5.12)-(5.13), the on-count bounds (5.3), and the individual TCL temperature constraints. The strategy hinges on switching TCLs before they reach their upper and lower temperature limits. The strategy begins to try to switch a TCL once it has entered the "upper temperature margin" or "lower temperature margin" (see Fig. 5.8). A switch from *off* to *on* within the upper margin can occur if the on-count is less than the upper bound or if another TCL is available to switch in the opposite direction (*off*). Similarly, a switch from *on* to *off* within the lower margin can occur if the on-count is greater than the lower bound or if another TCL is available to switch in the opposite direction (*on*).

For the ith TCL, the inner temperature of the lower margin $\underline{\theta}^i_{\mathrm{L}}$ is calculated as $\underline{\theta}^i_{\mathrm{L}} = \max(\theta^i_{\mathrm{L1}}, \theta^i_{\mathrm{L2}})$. The temperature θ^i_{L1} is the temperature reached after the TCL travels in the off-mode from $\underline{\theta}^i$ for a t_{L} length of time. The temperature θ^i_{L2} is the starting temperature from which it takes a t_{L} length of time for the TCL to travel in the on-mode to $\underline{\theta}^i$. Temperatures θ^i_{L3} and θ^i_{L4} are defined similarly but with respect to the upper limit $\overline{\theta}^i$. The inner temperature of the upper margin $\overline{\theta}^i_{\mathrm{L}}$ is calculated as $\overline{\theta}^i_{\mathrm{L}} = \min(\theta^i_{\mathrm{L3}}, \theta^i_{\mathrm{L4}})$.

To formalize the switching strategy, we must first define the following sets. Let $\overline{\mathbb{M}}(t)$ be the set of *off* TCLs in the upper margin, and let $\underline{\mathbb{M}}(t)$ be the set of *on* TCLs in the lower margin. Let $\mathbb{A}_{\mathrm{on}\to\mathrm{off}}(t)$ be the set of TCLs that are available to switch off at time t, i.e., $\mathbb{A}_{\mathrm{on}\to\mathrm{off}}(t) := \{i \mid \theta^i(t) < \overline{\theta}^i_{\mathrm{L}}, \sigma(t)^i = \mathit{on}, \lambda(t)^i = \mathit{unlocked}\}$. Similarly, let $\mathbb{A}_{\mathrm{off}\to\mathrm{on}}(t) := \{i \mid \theta^i(t) > \underline{\theta}^i_{\mathrm{L}}, \sigma(t)^i = \mathit{off}, \lambda(t)^i = \mathit{unlocked}\}$.

Proposed Strategy: Mode-Count Control for TCLs with Cycling Constraints

- If $|\overline{\mathbb{M}}(t)| > 0$, order the TCLs in $\overline{\mathbb{M}}$ by time-to-upper-limit in ascending order.
  - Set m equal to the smallest index in $\overline{\mathbb{M}}$ for which the TCL is also *unlocked*.
  - If $m \leq ((\overline{H} - H(t)) + |\mathbb{A}_{\mathrm{on}\to\mathrm{off}}(t)|)$,
    - switch TCL m *on* if $(\overline{H} - H(t)) > 0$,
    - otherwise switch TCL m *on* and simultaneously switch *off* the TCL in $\mathbb{A}_{\mathrm{on}\to\mathrm{off}}(t)$ with the largest time-to-upper-limit.
- If $|\underline{\mathbb{M}}(t)| > 0$, order the TCLs in $\underline{\mathbb{M}}$ by time-to-lower-limit in ascending order.
  - Set m equal to the smallest index in $\underline{\mathbb{M}}$ for which the TCL is also *unlocked*.
  - If $m \leq ((H(t) - \underline{H}) + |\mathbb{A}_{\mathrm{off}\to\mathrm{on}}(t)|)$,
    - switch TCL m *off* if $(H(t) - \underline{H}) > 0$,
    - otherwise switch TCL m *off* and simultaneously switch *on* the TCL in $\mathbb{A}_{\mathrm{off}\to\mathrm{on}}(t)$ with the largest time-to-lower-limit.
- If $T^i_{\mathrm{off}}(\theta^i(t), \overline{\theta}^i) = 0$ and TCL i is *off*, switch it *on*.
- If $T^i_{\mathrm{on}}(\theta^i(t), \underline{\theta}^i) = 0$ and TCL i is *on*, switch it *off*.

The strategy fails if a TCL travels all the way through its margin without being switched. When it reaches its upper or lower limit, it will switch and a count-bound will be violated.

Prior to the start of the strategy, we assume that TCLs have been operating "naturally", i.e., without external control. This implies that an *off* TCL will only be locked if $\theta^i(t_0) < \theta^i_{\mathrm{L1}}$ and an *on* TCL will only be locked if $\theta^i(t_0) > \theta^i_{\mathrm{L3}}$, where t_0 is the time at which the strategy starts.

For the switching strategy to succeed, TCLs' initial states must satisfy the following conditions. First is a trivial condition: TCLs' initial temperatures and the group's on-count must satisfy their constraints. Second, to prevent more than $\overline{H}$ from being locked *on* at the same time, no more than $\overline{H}$ TCLs should be in the *off* mode with $T^i_{\mathrm{off}}(\theta^i(t), \overline{\theta}^i) < t_{\mathrm{L}}$. Third, to prevent fewer than $\underline{H}$ from being *on*, no more than $(N - \underline{H})$ TCLs should be in the *on* mode with $T^i_{\mathrm{on}}(\theta^i(t), \underline{\theta}^i) < t_{\mathrm{L}}$.

5.4.3 Conditions on Feasible Count Bounds

We hypothesize that, under the proposed strategy for cycling constraints, the count bounds that can be satisfied indefinitely will have bound values that satisfy conditions similar to those in (5.4)-(5.5). Instead of depending on the upper and lower temperature limits, we expect the conditions will depend on the margin temperatures $\underline{\theta}_{\mathrm{L}}$ and $\overline{\theta}_{\mathrm{L}}$, as formulated below.

$$\underline{H} \;<\; \sum_{i=1}^{N} \frac{-f^i_{\mathrm{off}}(\underline{\theta}^i_{\mathrm{L}})/f^i_{\mathrm{on}}(\underline{\theta}^i_{\mathrm{L}})}{1 - f^i_{\mathrm{off}}(\underline{\theta}^i_{\mathrm{L}})/f^i_{\mathrm{on}}(\underline{\theta}^i_{\mathrm{L}})} \tag{5.14}$$

$$\overline{H} \;>\; N - \sum_{i=1}^{N} \frac{-f^i_{\mathrm{on}}(\overline{\theta}^i_{\mathrm{L}})/f^i_{\mathrm{off}}(\overline{\theta}^i_{\mathrm{L}})}{1 - f^i_{\mathrm{on}}(\overline{\theta}^i_{\mathrm{L}})/f^i_{\mathrm{off}}(\overline{\theta}^i_{\mathrm{L}})} \tag{5.15}$$

5.4.4 Simulation Results

We demonstrate the proposed strategy with simulations of a group of 1000 ACs with cycling constraints. The group's heterogeneous parameters are generated by sampling the ranges listed in Table 2.1. We use a lockout period of 1 minute for all ACs. We run three simulations: in the first, $\overline{H}$ is set to the "adjusted" least upper bound (LUB) determined by (5.15); in the second, $\overline{H}$ is set to the original LUB determined by (5.5); and in the third, the original strategy from [62] is tested with the adjusted LUB. In all simulations, the AC group is initialized such that each AC has an initial temperature of $\overline{\theta}^i_{\mathrm{L}}$, and $\overline{H}$ number of ACs are set to the on-mode. The lower bound $\underline{H}$ is set to zero for all simulations.

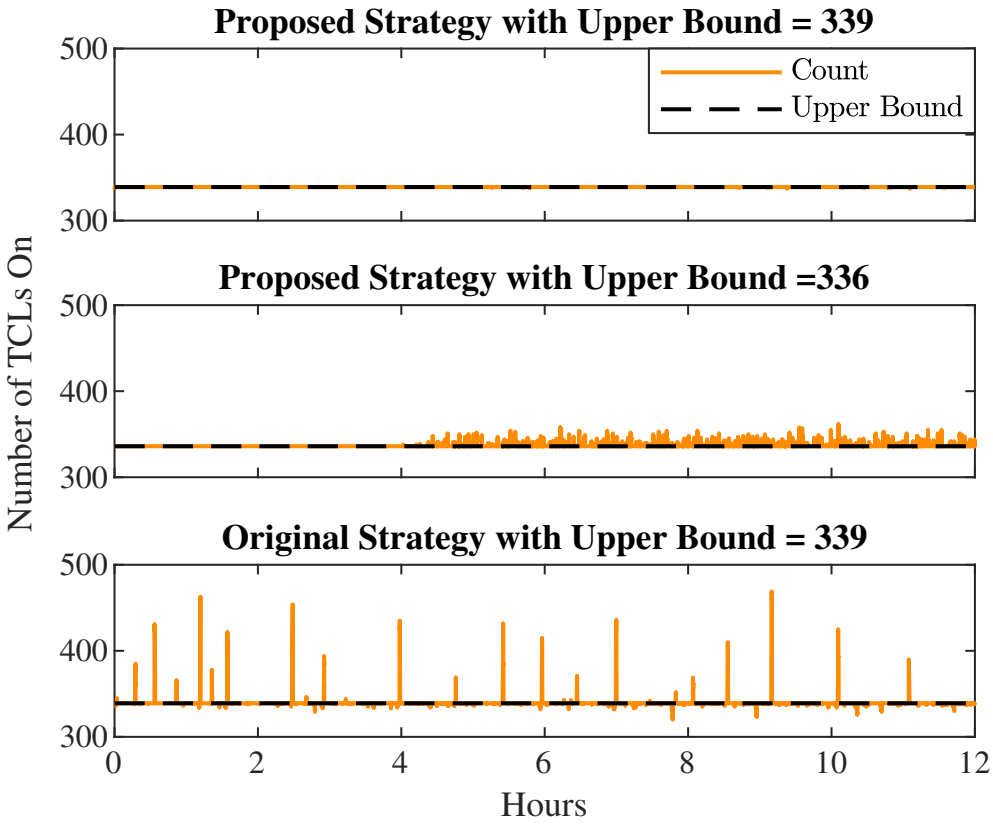


Figure 5.9: Group of 1000 ACs with cycling constraints controlled to satisfy an upper count-bound. The proposed strategy is able to satisfy the upper bound of 339 computed with (5.15) (top) but unable to satisfy the upper bound of 336 computed with (5.5) (middle). When we use the original strategy from [62] and the upper bound computed with (5.15), the strategy is unable to satisfy the upper bound (bottom).

As shown in the top plot of Fig. 5.9, the proposed strategy is able to satisfy the adjusted LUB for the simulated 12 hours. However, when the upper bound decreases to the original LUB, the strategy is unable to satisfy the bound (see middle plot). Finally, the original strategy is unable to satisfy the adjusted LUB and causes large spikes in a group's on-count (see bottom plot).

5.5 Chapter Conclusion

We have demonstrated a novel TCL control strategy that minimizes variation in the number of TCLs that are on by constraining the on-count between close lower and upper bounds. We have shown that the strategy is able to reduce the range of a TCL group's aggregate power and have found that the strategy is most effective at reducing the variability of medium-size groups (i.e., 50-100 TCLs).

We have also proposed a new control strategy for TCLs with cycling constraints and have demonstrated the strategy's ability to satisfy on-count bounds over 12 consecutive simulation hours. In future work, we will investigate heterogeneous lockout periods and the effect of longer lockout periods on the set of feasible bounds. In addition, we plan to incorporate the proposed strategy into a broader control architecture to track a balancing signal with aggregated TCLs while still ensuring that local constraints, such as voltage, are satisfied.

Chapter 6

Strategies for Network-Safe Load Control

This chapter is largely based on a manuscript[6] that is under review at a journal.

6.1 Chapter Introduction

High penetrations of aggregator-controlled distributed energy resources (DERs) pose a challenge to safe operation of distribution networks. In particular, DER aggregations participating in wholesale markets may cause unexpected power flows at the distribution level that result in reliability issues, as was noted by the U.S. Federal Energy Regulatory Commission (FERC) in a recent notice [21]. This is of particular concern with third-party aggregators, which are separate from distribution operators and do not necessarily have awareness of the local impacts of their control actions. In the U.S., third-party load aggregators are active in ancillary service markets as a result of FERC Order No. 719 [87]. As an example of the problem, consider a third-party load aggregator participating in regulation (i.e., secondary frequency control). To achieve a maximum setpoint, the aggregator may need 80% of its loads to draw power at the same time—effectively turning a distribution network's diversified load into coincident load—which could cause violations of lower limits on voltage magnitudes [73]. In this chapter, we propose new strategies for controlling a particular type of aggregation—an aggregation of residential, thermostatically controlled loads (TCLs).

The objective of this chapter is to develop TCL control strategies that provide regula-

[6] S.C. Ross and J.L. Mathieu. "Strategies for network-safe load control with a third-party aggregator and a distribution operator". (Under review).

tion without causing network constraint violations, all while preserving the privacy of the operator's and third-party aggregator's proprietary information. We propose a control architecture in which the aggregator controls a TCL aggregation to provide regulation, but, when necessary for network safety, the operator overrides the aggregator's commands to specific TCLs in the aggregation.

Our approach contrasts with that of prior work, which has not focused on maintaining privacy and separation between the operator and aggregator. In [93], the aggregator and operator are treated as the same entity, and a centralized AC-OPF is solved to achieve regulation while satisfying network constraints. In [17, 30], distributed strategies are proposed which could be implemented by a separate operator and aggregator; however, the two entities would be highly dependent on one another as they executed prescribed (i.e., not private) algorithms coupled through Lagrange multipliers.

We develop and compare two control strategies with different attributes. Strategy I is easier to implement but lower accuracy than Strategy II. In Strategy I, the aggregator uses an aggregate-model based control strategy to provide regulation, while the operator blocks particular TCLs from receiving the aggregator's commands when the network is at risk. In contrast, in Strategy II both the aggregator's and operator's control algorithms rely on individual models of each TCL; to ensure network safety, the operator not only blocks TCLs but also controls them with its own commands. In a case study, we evaluate the strategies in terms of regulation-signal tracking accuracy, and we test two versions of each control strategy. In the first, there is no coordination (i.e., no direct communication between operator and aggregator); in the second, the operator sends the aggregator partial information about its control actions.

The main contributions of this chapter are as follows. First, we propose two new control strategies for network-safe TCL control that preserve the privacy of the third-party aggregator and the distribution operator. The strategies differ in terms of cost, ease of implementation, and accuracy, and thus provide options for operators and aggregators with different capabilities and preferences. Second, we evaluate the performance of the proposed strategies, with and without coordination, in simulations of a realistic distribution network model with a high penetration of aggregated TCLs.

This chapter builds off of our work in Chapters 4 and 5. Strategy I is similar to the strategy in Chapter 4 in that they both use blocking. However, we develop new control methods for the aggregator to satisfy TCLs' cycling constraints—something that was not considered in Chapter 4. In Strategy II, we apply Chapter 5's mode-count control algorithm

to protect network constraints.

6.2 Problem Description

The *aggregator's objective* is to control the total power consumption of its TCL aggregation such that it tracks a regulation signal with sufficient accuracy. We assume that the aggregator's control actions should be non-disruptive [8] to the TCL's end-user. Thus, we restrict an aggregator's control actions to switching TCLs on/off only when the TCLs are within their user-set temperature range. The aggregator's control is subject to the individual dynamics and constraints of each TCL.

A *distribution operator's objective* is to reliably deliver power of sufficient quality to consumers. In Table 6.1, we list the set of constraints for power quality and reliability that are considered in this chapter. For transformers, apparent power is averaged over an hour because a transformer's thermal mass enables short-term violations of its power rating without causing overheating [33]. Note we will use the term "network safety" to refer to a distribution network's power quality and reliability.

Operators may need new tools to ensure network safety when third-party aggregators are active on their networks. Traditional network management tools, such as network reconfiguration and voltage regulation with regulators, have not been designed for the short but recurring fluctuations in power that can occur when aggregated loads provide frequency regulation. In this chapter, we give the operator the ability to override the aggregator's control actions when necessary to avoid network constraint violations. As with the aggregator, we assume that the operator's TCL control actions are non-disruptive to the end-user.

The *control plant*, for both the aggregator and operator, is an aggregation of thousands of heterogeneous TCLs. Each TCL in the plant is modeled separately to capture its individual temperature dynamics and constraints. We use an individual TCL model that was developed in [79] and is commonly found in the literature (e.g.,[28, 52]). For an aggregation of cooling TCLs, the ith TCL's temperature at time step $t{+}1$ is modeled as

$$\theta^i_{t+1} = \begin{cases} a^i\theta^i_t + (1-a^i)(\theta_{\text{amb},t} - r^i p^i_\theta) & \text{if } \sigma^i_t = 1, \\ a^i\theta^i_t + (1-a^i)\theta_{\text{amb},t} & \text{if } \sigma^i_t = 0, \end{cases} \tag{6.1}$$

where $a^i = \exp(-h/(c^i\, r^i))$ and h is the duration of the discrete model's time step. Variable σ represents the TCL's power status (1 for *on* and 0 for *off*). All other TCL parameters are

Table 6.1: Distribution Network Constraints

Component	Variable	Lower Limit	Upper Limit
Line	Current magnitude	–	100% of rating
Transformer	Avg. apparent power	–	100% of rating
Service node	Voltage magnitude	0.95 p.u.	1.05 p.u.

defined in Table 2.1.

To ensure non-disruptive control, we give the TCL's internal thermostat controller priority over external controllers. A TCL's power status is switched on/off by its internal controller if it reaches its upper or lower temperature limit. This condition is given by

$$\sigma_{t+1}^i = \begin{cases} 1 & \text{if } \theta_t^i \geq \overline{\theta}^i, \\ 0 & \text{if } \theta_t^i \leq \underline{\theta}^i, \end{cases} \tag{6.2}$$

where $\underline{\theta}^i = \theta_{\text{set}}^i - \delta^i$ and $\overline{\theta}^i = \theta_{\text{set}}^i + \delta^i$ (see Table 2.1 for parameter definitions).

TCLs can also have cycling constraints that prevent them from cycling too frequently. Such constraints are particularly important for TCLs with compressors (e.g., air conditioners) which must be off for a specific length of time to protect the compressor from physical damage. One way to reduce cycling frequency is to enforce a lockout period after a TCL has switched. During the lockout period, a TCL is "locked" in that it cannot be switched. Given a lockout period of τ_{L} time steps, a TCL's power status is constrained by

$$\sigma_t^i = \sigma_{t-1}^i \quad \text{if} \quad \sum_{k=t-\tau_{\text{L}}}^{t-1} \mathbb{1}_{\{\sigma_{t-1}^i\}}(\sigma_k^i) < \tau_{\text{L}}, \tag{6.3}$$

where the indicator function $\mathbb{1}_{\{f\}}(g)$ equals one if $g = f$ and zero otherwise. In this chapter, we assume the cycling constraint is designed to prevent excessive switching by external controllers and that it does not apply to internal control.

Finally, the total power of an aggregation of N TCLs is given by $P_{\text{total},t} = \sum_{i=1}^N P_{\text{R}}^i \sigma_t^i$, where P_{R}^i is a TCL's electric power rating with $P_{\text{R}}^i = p_\theta^i / \zeta^i$.

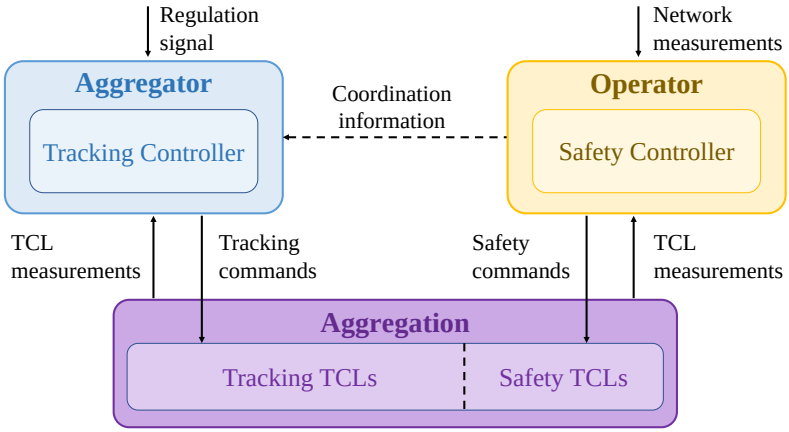


Figure 6.1: Control Architecture. The operator controls a portion of TCLs when necessary to prevent network violations. The aggregator controls all other TCLs to provide tracking. The dashed arrow indicates possible coordination.

6.3 Control Architecture

We propose a control architecture by which the aggregator and operator can both achieve their control objectives. Figure 6.1 shows the architecture. In this architecture, the operator determines in real-time which (if any) TCLs need to be temporarily removed from the aggregator's control in order to prevent network violations. Because the operator's interventions occur in real-time, we assume the aggregator is unable to reduce its regulation capacity, remove TCLs from its aggregation, or adjust its contract with the wholesale market operator in any other way. This creates a challenge for the aggregator: the aggregator must provide the same regulation service but with fewer TCLs responding to its control than expected.

In this architecture, the aggregator and operator act in parallel. In each time step, the aggregator and operator both receive measurements from, and send commands to, the TCL aggregation. On the aggregator side, the tracking controller receives a regulation setpoint from the bulk-system operator and computes switching commands such that aggregation's power consumption will match the setpoint in the next time step. On the operator side, the safety controller receives network measurements (e.g., currents and voltages) and determines which TCLs to block from receiving the aggregator's control and what commands to send to them. Finally, the operator may coordinate with the aggregator by sending the aggregator information about its control actions.

We propose two control strategies—Strategy I and II—both of which use the proposed architecture. Table 6.2 provides details on the components of each strategy and enables a

comparison of their measurement and communication (M&C) requirements.

The aggregator's M&C requirements are more substantial in Strategy II than Strategy I. In Strategy I, the aggregator's controller requires only an aggregate TCL power measurement and sends back a small number of probabilistic commands to all TCLs. In contrast, in Strategy II, the aggregator's controller requires power and temperature measurements from each TCL and sends back an individual on/off command to each TCL.

The operator's M&C requirements are also more substantial in Strategy II than Strategy I. In Strategy I, the operator's controller requires only power measurements from each TCL and sends back a "block" or "unblock" command to each unit, as needed. In contrast, in Strategy II, the operator's controller requires temperature and power measurements from each TCL and sends back individual on/off commands to blocked units in addition to block/unblock commands.

6.4 Methods: Strategy I

6.4.1 Blocking Control

In Strategy I, we propose that operators use blocking control to ensure safe network operation. When a TCL is blocked, it is unresponsive to aggregator commands, and it returns to its regular on/off cycles governed by its internal thermostat. To protect a particular network constraint, we block a group of TCLs whose demand would otherwise contribute to the violation of the constraint. To protect the full network, we must identify which constraints are at risk of violation due to regulation and select an appropriate set of TCLs to block. The ideal control solution would identify the at-risk constraints online, so as to adjust for changing operating conditions, and would select the minimum set of TCLs to block, so as to minimize impact on the aggregator's control. Since this is not the focus of this chapter, we identify at-risk constraints using offline simulations and use approximate methods to select which TCLs to block. Developing an online control solution is future work.

The method that we use for selecting TCLs depends on the type of constraint that is at risk. If a line or transformer are overloaded, then we incrementally block TCLs that are downstream from the component until the overload is relieved. Specifically, we block TCLs furthest from the component first, where distance is in terms of line length. This is an approximate method for blocking TCLs according to how much they contribute to the overload, where a TCL's contribution is the sum of its demand and associated line losses. If a

Table 6.2: Control Strategies

	Entity	Controller	Measurements	Commands to TCLs
Strategy I	Aggregator	Aggregate-Model Based Control	Aggregate power	Probabilistic, one value for each bin
	Operator	Blocking Control	Power of each TCL	Block or unblock to each TCL
Strategy II	Aggregator	Individual-Model Based Control	Power & temp. of each TCL	On/off to each TCL
	Operator	Mode-Count Control	Power & temp. of each TCL	On/off to each TCL

service node has a voltage violation, then we incrementally block TCLs that are downstream of the node. If the violation persists, then we incrementally block TCLs that are directly or indirectly on the "path" of lines that connect the node to the substation. (A TCL is indirectly on the path if located on a line that branches off of the path.) We block TCLs according to their distance along the path: TCLs closer to the node are blocked first.

6.4.2 Aggregate-Model Based Tracking Control

Strategy I's proposed tracking controller is based on an aggregate model of TCL state progression. In Section 6.6, we benchmark the tracking performance of the proposed, aggregate-model based controller against that of a model-less, PI controller. We expect the proposed controller to have better performance because it uses model-based prediction to compensate for the internal control actions of TCLs and uses state estimation to prioritize switching TCLs that are about to internally switch.

Aggregate TCL Model

The proposed model takes the form of a population transition model, which has been commonly used in the literature on TCL control (e.g., [44, 52]). In this type of model, the state space of all TCLs is discretized into the same number of "bins" and the progression of TCLs from bin to bin is modeled with transition probabilities.

In this chapter, we extend the population transition model to include TCL lockout dynamics. Figure 6.2 depicts the proposed model. Each bin is represented by a circle and is defined by a locked/unlocked status, on/off status, and one of N_{I} temperature intervals. In

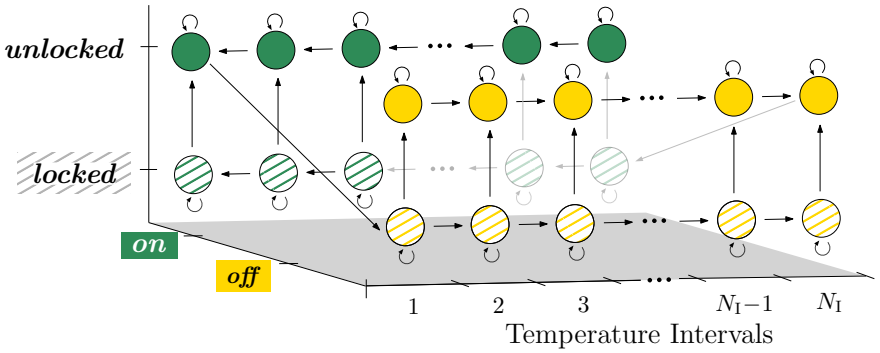


Figure 6.2: State Transition Diagram with Locked States. Arrows indicate the most likely state transitions for an uncontrolled TCL with a lockout period.

Fig. 6.2, horizontal arrows represent transitions from one temperature interval to the next due to TCLs' temperature dynamics. Diagonal arrows represent transitions due to thermostat actions. Vertical arrows represent TCLs becoming unlocked after their lockout times have elapsed. Transitions due to aggregator-control are not shown but would be downward diagonals in each temperature interval.

We formulate the aggregate model as a linear time-varying system:

$$\begin{aligned} \mathbf{x}_{t+1} &= \mathbf{A}_t \mathbf{x}_t \\ \mathbf{y}_t &= \mathbf{C}\mathbf{x}_t, \end{aligned} \tag{6.4}$$

where $\mathbf{x}$ is the state vector, $\mathbf{A}$ the state transition matrix, $\mathbf{y}$ the output, and $\mathbf{C}$ the output matrix. The state $\mathbf{x}$ is the distribution of the TCL population across the bins. Given the different types of bins, the state vector can be divided into four categories: $\mathbf{x} = [\underbrace{x^1 \dots x^{N_I}}_{\text{off, unlocked}} \quad \underbrace{x^{N_I+1} \dots x^{2N_I}}_{\text{on, unlocked}} \underbrace{x^{2N_I+1} \dots x^{3N_I}}_{\text{off, locked}} \quad \underbrace{x^{3N_I+1} \dots x^{4N_I}}_{\text{on, locked}}]^T$. The entries of the state transition matrix $\mathbf{A}_t$ are the probability that a TCL will transition from bin to bin. For example, the entry in the ith row and jth column is the probability that a TCL in bin j will transition to bin i in the next time step.

We model transitions due to the aggregator's probabilistic commands within the $\mathbf{A}$ matrix. This is similar to the Markov Decision Process approach in [11]. We formulate $\mathbf{A}_t$ as the product of two transition matrices: $\mathbf{A}_t = \mathbf{A}_{u,t}\mathbf{A}_s$, where $\mathbf{A}_s$ models "internal" transitions due to temperature dynamics, lockout dynamics, and internal control, and $\mathbf{A}_{u,t}$ models external transitions due to aggregator control. For the purpose of modeling, we have assumed that, in a given time step, internal transitions occur before external transitions. The matrix $\mathbf{A}_s$ can be identified for a particular outdoor temperature by counting bin transitions when

aggregator-control is inactive. Here we assume a constant outdoor temperature, which makes $\mathbf{A}_{\mathrm{s}}$ time-invariant.

Transition probabilities in $\mathbf{A}_{\mathrm{u},t}$ depend on the aggregator's control command $\mathbf{u}_t$. The ith entry of $\mathbf{u}_t$ is the probability with which TCLs in bin i should switch at time step t. The entries only correspond to bins that are unlocked, so $\mathbf{u}_t$ has length $2N_{\mathrm{I}}$. Given this definition of $\mathbf{u}_t$, matrix $\mathbf{A}_{\mathrm{u},t}$ takes the form

$$\left[\begin{array}{cc|cc} \mathbf{I} - \mathrm{diag}(u_t^1, \ldots, u_t^{N_{\mathrm{I}}}) & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} - \mathrm{diag}(u_t^{N_{\mathrm{I}}+1}, \ldots, u_t^{2N_{\mathrm{I}}}) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathrm{adiag}(u_t^{2N_{\mathrm{I}}}, \ldots, u_t^{N_{\mathrm{I}}+1}) & \mathbf{I} & \mathbf{0} \\ \mathrm{adiag}(u_t^{N_{\mathrm{I}}}, \ldots, u_t^1) & \mathbf{0} & \mathbf{0} & \mathbf{I} \end{array}\right],$$

where $\mathbf{I}$ is the identity matrix, $\mathrm{diag}(\cdot)$ maps the input vector to a diagonal matrix, and $\mathrm{adiag}(\cdot)$ maps the input vector to an anti-diagonal matrix, where the diagonal runs from the upper right corner to the lower left corner. The left side of $\mathbf{A}_{u,t}$ models transitions due to aggregator control: the diagonal sub-matrices model transitions out of unlocked bins, and the anti-diagonal sub-matrices model transitions into locked bins.

The two outputs of the aggregate model are: 1) the total power consumption of the aggregation and 2) the fraction of TCLs in the aggregation, which should always equal one. The output matrix is given by $\mathbf{C} = [\mathbf{C}_{\mathrm{s}} \;\; \mathbf{C}_{\mathrm{s}}]$, where

$$\mathbf{C}_{\mathrm{s}} = \left[\begin{array}{cccccc} 0 & \ldots & 0 & \overline{p}_{\mathrm{on}} N & \ldots & \overline{p}_{\mathrm{on}} N \\ 1 & \ldots & 1 & 1 & \ldots & 1 \end{array}\right]$$

and $\overline{p}_{\mathrm{on}}$ is the average power consumption of an *on* TCL.

State Estimation

We use a time-varying Kalman filter [53] to estimate the aggregate state in (6.4). The output of the Kalman filter in each time step is $\hat{\mathbf{x}}_t$. We account for process noise and measurement noise by adding terms $\mathbf{w}$ and $\mathbf{v}$ to the state and output equations in (6.4), respectively. We treat the covariances of $\mathbf{w}$ and $\mathbf{v}$, denoted as $\mathbf{Q}$ and $\mathbf{R}$, as tuning parameters. For output measurements, we use $\mathbf{y}_{\mathrm{meas},t} = \left[\begin{array}{cc} P_{\mathrm{total},t} & 1 \end{array}\right]^{\mathrm{T}}$. We assume the second output measurement is perfect (i.e., has zero measurement noise) and set the corresponding entry in $\mathbf{R}$ to zero. In this way, the second output equation acts as an equality constraint within the Kalman Filter [77].

Control Policy

We propose a control policy based on one-step model prediction. The policy selects a control input such that the model's predicted output in the next time step matches the desired output. We note that the proposed policy is similar to that of [52]; differences emerge due to differences in aggregate models.

We derive the policy by solving the aggregate model (6.4) for $\mathbf{u}_t$ given the state estimate $\hat{\mathbf{x}}_t$. To achieve the desired output $\mathbf{y}^*_{t+1}$, we choose $\mathbf{u}_t$ such that $\mathbf{y}^*_{\mathrm{t+1}} = \mathbf{C}\mathbf{A}_{\mathrm{u},t}\mathbf{A}_{\mathrm{s}}\hat{\mathbf{x}}_t$. The matrix $\mathbf{A}_{\mathrm{u}}$ can be decomposed into two matrices: $\mathbf{A}_{\mathrm{u},t} = \tilde{\mathbf{A}}_{\mathrm{u},t} + \mathbf{I}$, and therefore $\mathbf{y}^*_{\mathrm{t+1}} = \mathbf{C}\tilde{\mathbf{A}}_{\mathrm{u},t}\mathbf{A}_{\mathrm{s}}\hat{\mathbf{x}}_t + \mathbf{C}\mathbf{A}_{\mathrm{s}}\hat{\mathbf{x}}_t$. Given the structure of $\mathbf{A}_{\mathrm{u},t}$, the second row of this equation is satisfied by any choice of $\mathbf{u}_t$; however, $\mathbf{u}_t$ must be chosen carefully to satisfy the first row. The first row of this equation is

$$P^*_{\mathrm{total},t+1} = \overline{p}_{\mathrm{on}} N \left(\sum_{n=1}^{N_{\mathrm{I}}} u_t^n x^n_{\mathrm{pred},t} + \sum_{n=N_{\mathrm{I}}+1}^{2N_{\mathrm{I}}} -u_t^n x^n_{\mathrm{pred},t} \right) + \mathbf{C}^1 \mathbf{A}_{\mathrm{s}} \hat{\mathbf{x}}_t, \tag{6.5}$$

where $P^*_{\mathrm{total},t+1}$ is the desired power in the next time step and is the first entry of $\mathbf{y}^*_{\mathrm{t+1}}$. The vector $\mathbf{x}_{\mathrm{pred},t}$ is the predicted state if only internal transitions were to occur in the next time step and is given by $\mathbf{x}_{\mathrm{pred},t} = \mathbf{A}_{\mathrm{s}}\hat{\mathbf{x}}_t$. The vector $\mathbf{C}^1$ is the first row of $\mathbf{C}$.

We design the control policy such that, in each time step, TCLs are switched in only one direction (*on* or *off*). The entries of $\mathbf{u}_t$ are probabilities and must be between 0 and 1. Given these restrictions, when a positive change in power is needed, (6.5) can be simplified and rearranged such that

$$\sum_{n=1}^{N_{\mathrm{I}}} u_t^n x^n_{\mathrm{pred},t} = K \frac{P^*_{\mathrm{total},t+1} - \mathbf{C}^1 \mathbf{A}_{\mathrm{s}} \hat{\mathbf{x}}_t}{\overline{p}_{\mathrm{on}} N}, \tag{6.6}$$

where K has been added as a proportional gain and all other entries of $\mathbf{u}_t$ are set to zero. Similarly, when a negative change in power is needed, the last N_{I} entries of $\mathbf{u}_t$ must satisfy

$$\sum_{n=N_{\mathrm{I}}+1}^{2N_{\mathrm{I}}} -u_t^n x^n_{\mathrm{pred},t} = K \frac{P^*_{\mathrm{total},t+1} - \mathbf{C}^1 \mathbf{A}_{\mathrm{s}} \hat{\mathbf{x}}_t}{\overline{p}_{\mathrm{on}} N}. \tag{6.7}$$

Because (6.6) is under-determined when $N_{\mathrm{I}} > 1$, we have the freedom to choose $\mathbf{u}$ as long as (6.6) is satisfied. We design a rule for selecting $\mathbf{u}$ that prioritizes switching TCLs that are closest to being switched by internal control. When switching TCLs *on*, bin N_{I} has first priority, bin $N_{\mathrm{I}}-1$ has second priority, and so on. Let b be the minimum index for which

$\sum_{k=0}^{b} x_{\text{pred},t}^{N_\text{I}-k}$ is greater than the right hand side of (6.6). We set the entries of $\mathbf{u}$ equal to 1 for all bins of higher priority than $N_\text{I}-b$ and equal to zero for all bins of lower priority. Finally, the value for bin $N_\text{I}-b$ is chosen such that (6.6) is satisfied.

The control policy is very similar when a decrease in power is needed. In this case, $\mathbf{u}$ must satisfy (6.7), and bins $2N_\text{I}$ to $N_\text{I}+1$ are switched out of, in that order.

Coordination

When there is coordination in Strategy I, the operator informs the aggregator of the number of TCLs it is blocking in the current time step. To incorporate this information into the aggregator's control policy, we use a lookup table of proportional gains for each of five blocking levels: 0%, 10% 20%, 30%, and 40%. When the number blocked changes, we update K in (6.6) and (6.7) by linearly interpolating between the nearest values in the lookup table.

6.5 Methods: Strategy II

6.5.1 Mode-Count Control

For the operator in Strategy II, we propose the use of Chapter 5's mode-count control algorithm to ensure network safety. We propose two applications of mode-count control: 1) prevention of under-voltages or over-currents by reducing a TCL group's maximum demand, and 2) prevention of over-voltages by increasing a TCL group's minimum demand. For brevity, we present only the first application; the methods for the second application follow directly.

For a group of co-located TCLs, the policy determines which TCLs to switch *on*/*off* to constrain the group's "on-count", and thereby the group's demand. A group's on-count is the number of TCLs in the group that are *on* at a particular time and is denoted H_t^j for group j at time t. We constrain a group's on-count such that $H_t^j \leq \overline{H}^j$, where $\overline{H}^j$ is a low (but feasible) upper bound.

Figure 6.3 provides an example of the control policy. Consider a group of 7 TCLs with upper bound $\overline{H} = 3$. The central idea of the policy is to switch TCLs *on* as soon as possible once they have entered the "upper margin" of their temperature range. At the time step shown in Fig. 6.3, the policy switches TCL #4 *on* because it has just entered its upper margin, and switches TCL #1 *off* to satisfy the counting constraint. A main goal of the

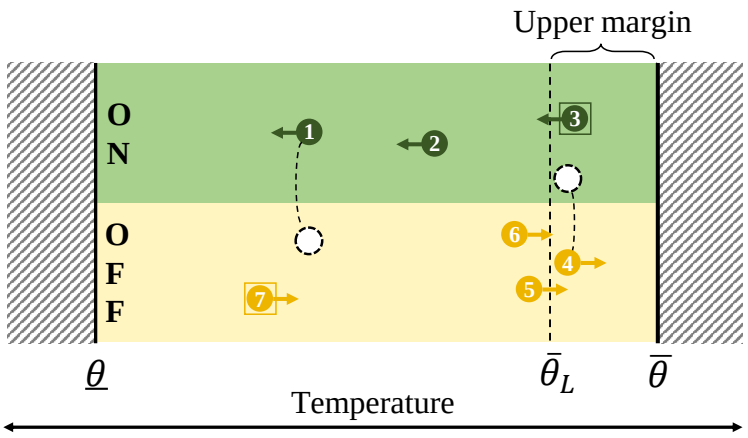


Figure 6.3: Demonstration of Mode-Count Control. The policy maintains the on-count ≤ 3 by switching *off* TCLs *on*, as soon as possible after entering the upper margin, and if needed, switching an *on* TCL *off* to compensate. TCLs are indicated by numbered circles; boxes indicate locked status.

policy is to avoid having too many TCLs locked *on* at the same time, which could happen if we waited to switch TCLs at their upper temperature limit.

We summarize the steps of the control policy in Algorithm 1 and define three terms used within the algorithm as follows. (For a more detailed treatment of the policy see Section 5.4.) First, a TCL's "time-to-upper-limit" is the time it would take a TCL to progress from its current temperature θ_t^i to its upper limit $\overline{\theta}^i$ in the *off* direction. This time can be explicitly calculated from the individual TCL model, and for the ith TCL is given by $t_{\mathrm{UL}}^i = r^i c^i \ln\left((\theta_{\mathrm{amb}} - \theta_t^i)/(\theta_{\mathrm{amb}} - \overline{\theta}^i)\right)$. Second, the "upper margin temperature" $\overline{\theta}_{\mathrm{L}}$ is the lesser of two temperatures: 1) the temperature from which it takes the TCL τ_{L} time steps to reach $\overline{\theta}$ when *off*, and 2) the temperature reached by the TCL τ_{L} time steps after leaving $\overline{\theta}$ when *on*. Third, a TCL "counter-switches" with another TCL by switching at the same time but in the opposite direction (i.e., to the opposite power status).

To prevent the violation of a particular network constraint, an operator must first identify the group of TCLs to control and then set a lower bound on the group's on-count. As with blocking, we wish to control the smallest set of TCLs that will prevent the network constraint violation. In this work, we use the same approximate methods described in Section 6.4.1 to determine this set, but instead of blocking the TCLs, we constrain them using Algorithm 1. After a group of TCLs has been selected, we set the group's upper bound $\overline{H}^j$ to the lowest feasible value; see Section 5.4 where we make a conjecture on this value.

ALGORITHM 1: MODE-COUNT CONTROL FOR UPPER BOUND

In each time step,

1. Initialize $\Delta H = \overline{H} - H_t$, and find the *off* unit in the upper margin with shortest time-to-upper-limit. Let this unit's index be g.
2. If unit g is *unlocked*:
   a) If $\Delta H > 0$, then switch unit g *on* and decrement ΔH.
   b) Otherwise, find a TCL that is available to counter-switch; this TCL must be *on* and *unlocked*, must have $\theta_t^i < \overline{\theta}_{\mathrm{L}}^i$, and must not be reserved to counter-switch with a different TCL. Let s be the index of the available TCL with the longest time-to-upper-limit. Switch unit g *on* and unit s *off*.
3. If unit g is *locked*:
   a) If possible, find a TCL that is available to counter-switch. Reserve this TCL for counter-switching with g in the future.
   b) Otherwise if $\Delta H > 0$, decrement ΔH so that unit g will be able to switch in the future, as soon as it is unlocked.
4. Find the *off* unit with the next shortest time-to-upper limit. Let this unit's index be g.
5. Repeat 2)-4) until either there are no more *off*, *unlocked* units in the upper margin, or $\Delta H = 0$ and there are no more *on* TCLs available for counter-switching.

6.5.2 Individual-Model Based Tracking Control

For the aggregator in Strategy II, we propose a control policy based on individual TCL models. The general principal of the control policy is to switch TCLs that are closest in terms of time to being switched internally. The proposed controller is similar to the priority stack method of [28], but we use time to determine priority instead of temperature. We summarize the steps of the control policy in Algorithm 2 and discuss each step in the remainder of the section.

ALGORITHM 2: INDIVIDUAL-MODEL BASED TRACKING CONTROL

In each time step,

1. Calculate $\Delta P_{\text{track},t}$, the change in power needed to track $P^*_{\text{total},t+1}$, taking into account changes in power due to internal switching and, if there is coordination, safety control.
2. Update priority stacks $\mathbb{S}_{\text{on},t}$ and $\mathbb{S}_{\text{off},t}$.
3. If $\Delta P_{\text{track},t} \geq P_{\text{small}}$
   a) Find index $j^* \in \mathbb{S}_{\text{off},t}$ that minimizes $|\sum_{i=1}^{j} P^i_{\text{R}} - \Delta P_{\text{track},t}|$
   b) Switch *on* unit j^* and all units of higher priority in $\mathbb{S}_{\text{off},t}$
4. Otherwise, if $\Delta P_{\text{track},t} \leq -P_{\text{small}}$
   a) Find index $j^* \in \mathbb{S}_{\text{on},t}$ that minimizes $|\sum_{i=1}^{j} P^i_{\text{R}} + \Delta P_{\text{track},t}|$
   b) Switch *off* unit j^* and all units of higher priority in $\mathbb{S}_{\text{on},t}$

In step 1, we calculate the change in power that is needed to achieve tracking in the next time step. When there is no coordination, this change in power is given by

$$\Delta P_{\text{track},t} = P^*_{\text{total},t+1} - (P_{\text{total},t} + \Delta P_{\text{internal}}); \tag{6.8}$$

see the end of this section for the calculation with coordination. Variable $\Delta P_{\text{internal}}$ is the change in power that will occur in the next time step due to internal switching. We determine $\Delta P_{\text{internal}}$ by predicting which TCLs will be switched by their internal controllers in the next time step. For each TCL, we calculate its time-to-upper-limit t_{UL} (as defined in Section 6.5.1) and its time-to-lower-limit t_{LL} (defined as the time it would take a TCL to progress from its current temperature to its lower limit in the *on* direction). Both metrics can be solved for using (6.1). A TCL is predicted to switch in the next time step if it is *off* and has $t_{\text{UL}} < h$, or if it is *on* and has $t_{\text{LL}} < h$.

In step 2, we update the priority stacks. Priority stack $\mathbb{S}_{\text{on}}$ is composed of *on* and *unlocked* TCLs and is sorted by time-to-lower-limit; priority stack $\mathbb{S}_{\text{off}}$ is composed of *off* and *unlocked* TCLs and is sorted by time-to-upper-limit. In each time step, we update the priority stacks according to the switching actions from the last time step and newly calculated values for t_{UL} and t_{LL}. The set of TCLs under control by the operator are excluded from the priority

stacks; we assume the aggregator is able to identify this set given its TCL measurements.

In step 3, we select which TCLs to switch. To prevent excessive switching, we only switch TCLs if the desired change in power has magnitude greater than a threshold, P_{small}. Parameter P_{small} can be thought of as a tuning parameter; here we set it to 25% of the smallest, individual power rating in the aggregation. If the threshold is reached, then we calculate how much of the stack should switch to minimize the tracking error (see step 3a). In this calculation, we take the sum of the TCLs' power ratings in order of priority. The selected TCLs are switched in step 3b.

Step 4 is similar to step 3, except ΔP_{track} is negative, so units need to be switched *off* rather than *on*.

When there is coordination, the operator sends the aggregator ΔP_{safety}, the change in power its safety control actions will cause in the next time step. The aggregator compensates for the operator's actions by subtracting ΔP_{safety} from the right-hand side of (6.8). All other aspects of the control are the same.

6.6 Case Study

6.6.1 Benchmark Strategy

In the case study, we compare the proposed control strategies to a benchmark strategy. The benchmark is identical to Strategy I, except the aggregator's tracking controller is a model-free proportional-integral (PI) controller. The PI controller computes a scalar switching command, whose magnitude is the probability with which TCLs should switch and whose sign is the direction with which TCLs should switch: positive (*on*) or negative (*off*). We use a discrete-time PI control algorithm with anti-windup (see pp. 311-312 of [2]). The tuning parameters are a proportional gain K_{P} and an integral gain $K_{\text{P}}/T_{\text{I}}$, where T_{I} is the integral time constant. We set the anti-windup time constant equal to that of the integral gain. When there is coordination, we use a look-up table for the gains, as in Strategy I.

6.6.2 Setup

We test the control strategies in simulations of a distribution network with a high-penetration of aggregator-controlled residential air conditioners (ACs). We assume 100% of houses on the network have ACs, and we use an outdoor temperature of 32°C in order to capture a peak-load hour. We compare the strategies' performances across 10 1-hour trials, where each

Table 6.3: Case Study: Aggregation and Network Details

# of Nodes	# of ACs	Baseline AC Load	Non-AC Load
613	2265	4.42 MW	2.01 MW

trial uses a different segment of the RegD regulation signal from the PJM Interconnection [68]. We scale the signal such that its amplitude is +/-33% of the ACs' baseline power consumption. Good control performance is marked by high accuracy in tracking and low prevalence of network violations.

The model has two main parts: 1) the network and 2) the AC aggregation and controllers. We use a realistic network model and GridLAB-D [63] to solve the network's three-phase, unbalanced power flow. The network model is of an actual system: # R1-12-47-1 in Pacific Northwest National Lab's (PNNL's) database [75]. The network model includes dynamic voltage regulation in the form of one voltage regulator and two capacitor banks. During simulations, we measure current flow through lines, apparent power flow through transformers, and voltages at residential meters.

We make the following five modifications to improve the accuracy of the network model: 1) we increase a distribution transformer's rating if its average load over an hour is higher than both its original planning load and its original rating; 2) we increase the size of triplex lines if their maximum current is larger than their rating; 3) we shift the regulation range of capacitor banks so that voltages stay within 0.95-1.05 p.u. during nominal operation; 4) we adjust the capacitor bank settings such that phases are individually controlled; 5) we set the initial conditions of capacitors and regulators so that they are at steady state condition when the test trials begin.

We model the AC aggregation and controllers in MATLAB. We determine the number of houses served by each residential distribution transformer using a disaggregation method developed by PNNL [64]. Each house is composed of one AC model and one constant-power baseload. We use (6.1)-(6.3) to model each AC and generate a heterogeneous population by randomly selecting parameter values from the uniform distributions described in Table 2.1. A power factor of 0.97 is used for ACs and 0.95 for baseload. Table 6.3 reports the AC aggregation's baseline power consumption, as well as other important metrics about the network and loads.

We tune the tracking controllers of Strategy I and the benchmark strategy for the best average tracking performance over a set of tuning trials. When there is no coordination,

Table 6.4: Tuning Parameters

Strategy I			Benchmark	
K	$\mathbf{Q}$	$\mathbf{R}$	K_{P}	T_{I}
1	$\mathrm{diag}(I, 10^2 I)$	$10^9 \mathrm{diag}(1,\ 0)$	3.4×10^{-4}	100

$\mathrm{diag}(\cdot)$ maps the input matrices to a block diagonal matrix

we use 50 tuning trials consisting of 25 different hours of the RegD signal and two different blocking levels: 0% and 20%. Table 6.4 lists the resulting tuned parameters. When there is coordination, we tune the controllers for each blocking level independently, again using 25 hours of the RegD signal. The tuned parameters have the following ranges of values: $K \in [1, 1.15]$, $K_{\mathrm{P}} \in [3 \times 10^4, 5 \times 10^5]$, and $T_{\mathrm{I}} \in [80, 800]$, where the lower and upper values correspond to 0% and 40% blocked, respectively. Finally, in Strategy I's aggregate load model, we use five temperature intervals, i.e., $N_{\mathrm{I}} = 5$.

For each strategy and trial, we determine which ACs should be under safety control through a series of preliminary simulations. Before each simulation, we select additional ACs for safety control according to the methods proposed in Section 6.4.1. We repeat this process until no violations occur.

6.6.3 Results

Table 6.5 shows the results of our case study, with values averaged over the 10 test trials. For all strategies, the safety controllers are successful—no network violations occur. Without safety control, there are an average of 2.2 over-voltage nodes and 1.2 overloaded lines, where the averages are taken over all trials and strategies. The number of ACs under safety control varies considerably, depending on whether blocking or mode-count control is used. In the benchmark strategy, 13.8% of ACs are blocked, on average, in order to avoid network violations; this number falls slightly to 10.3% when coordination is used. Similarly, in Strategy I, 15.3% of ACs are blocked, and 15.2% with coordination. In contrast, Strategy II's mode-count controller requires only 0.4% of the aggregation to prevent network violations, with and without coordination.

In terms of tracking error, Strategy II outperforms Strategy I, and both outperform the benchmark strategy. In Table 6.5, root-mean-square (RMS) error in tracking is reported as a percentage of the baseline power of the aggregation. Comparing against the benchmark, we

Table 6.5: Case Study Results: Averages over 10 Trials

Strategy	Coord-ination	ACs under Safety Control	Network Violations	Tracking Error (RMS)
Benchmark	No	13.8%	0	1.37%
	Yes	10.3%	0	1.28%
Strategy I	No	15.3%	0	0.72%
	Yes	15.2%	0	0.70%
Strategy II	No	0.4%	0	0.10%
	Yes	0.4%	0	0.10%

find that the tracking error of Strategy I (0.72%, and 0.70% with coordination) is almost 50% less than that of the benchmark (1.37%, and 1.28% with coordination), and the tracking error of Strategy II (0.10% with and without coordination) is more than an order of magnitude less than that of the benchmark. In all strategies, tracking performance improves with coordination; we note that for Strategy II this improvement is not shown by Table 6.5 due to rounding.

In Fig. 6.4, we demonstrate a few of Strategy I's key features using time-series plots from a selected trial. In the upper plot Strategy I's aggregate-model based controller closely tracks the regulation signal, despite a large percentage—in this trial, 38%—of its aggregation being blocked. However, the tracking is not perfect: there is noticeable error in the 6^{th} minute, when the aggregation has saturated (i.e., not enough ACs are available to switch *off* due to blocking or lockout). The middle plot of Fig. 6.4 demonstrates an attribute of Strategy I's state estimator: for bins that frequently receive external switching commands (i.e., bins N_{I} and $2N_{\text{I}}$), the state estimate corresponds to the number of ACs that are available to switch (i.e., unblocked ACs), rather than the total number of ACs in the bin. This helps the controller compensate for blocking. The lower plot of Fig. 6.4 demonstrates the effects of blocking. In this trial, phase C of overhead line 301 experiences an over-current in minutes 7 and 8 when no ACs are blocked; this over-current is effectively reduced by blocking a portion of the ACs on phase C.

Fig. 6.4 also highlights the main drawback of using blocking for safety control: a relatively large percentage of ACs must be blocked to protect just a few network constraints. This is because, unlike mode-count control, blocking does not actively constrain a group's maximum demand. Instead, blocking returns a group to its normal variations in demand and level of load-diversity. Because larger groups of ACs have more diversified load, blocking is more

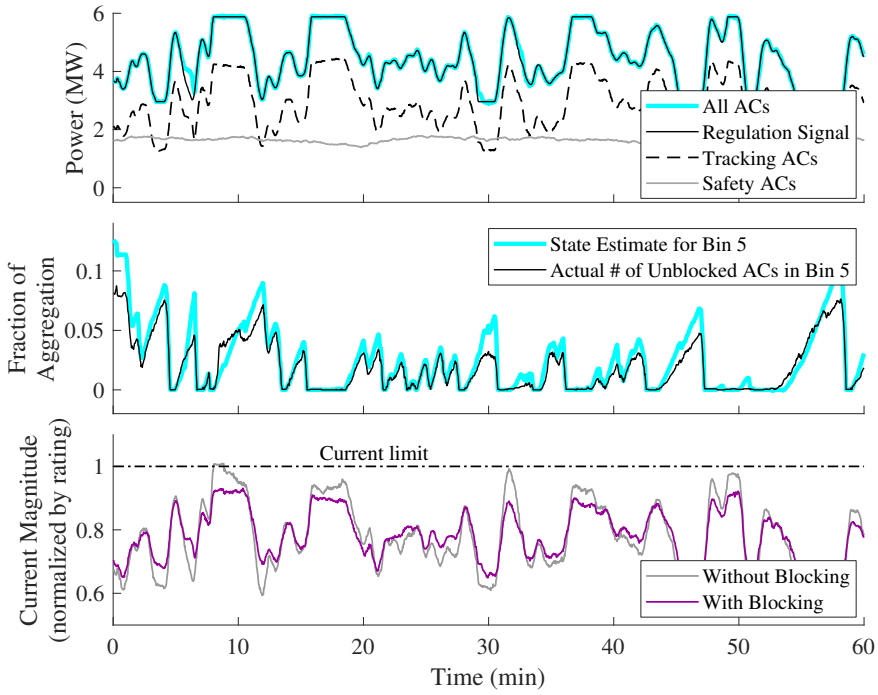


Figure 6.4: Strategy I during a selected trial. The plots shows the strategy's high tracking accuracy (top plot) and state-estimation performance for bin 5 (middle plot). The bottom plot shows that blocking a large portion of TCLs can prevent over-currents, here on overhead line #301.

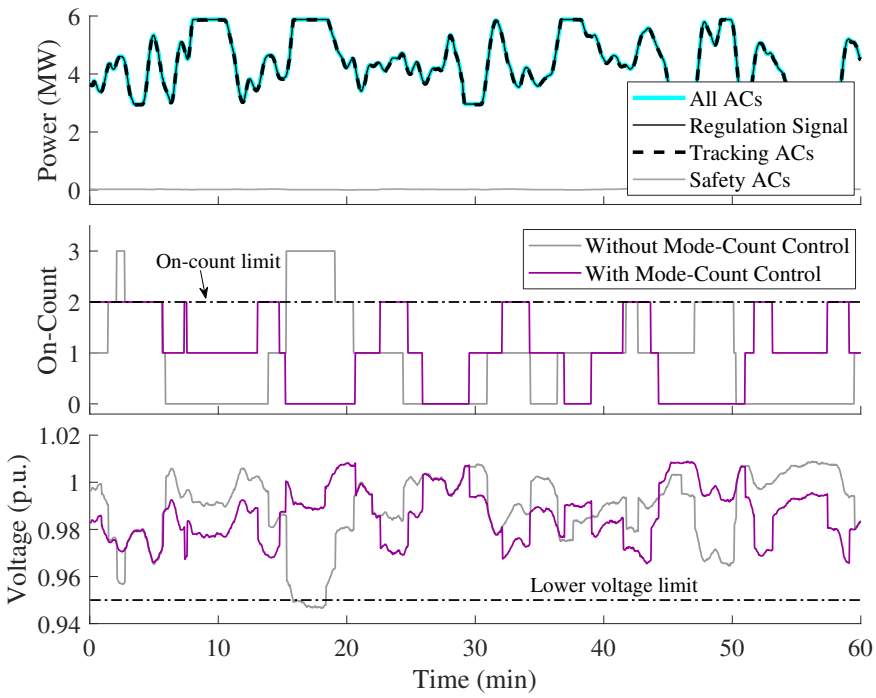


Figure 6.5: Strategy II during a selected trial. Plot (a) shows the strategy's very high tracking accuracy. The plots in (b) shows that mode-count control can maintain a group of 3 TCLs' on-count to ≤ 2 (top plot) and thereby prevent under-voltages (bottom plot), here for residential meter # 469.

effective with larger group sizes.

In Fig. 6.5, we demonstrate a few of Strategy II's key features using the same selected trial. The top plot shows the high-accuracy tracking of Strategy II's individual-model based controller. Compared to Strategy I, Strategy II has very few ACs under safety control—only 0.4% in this trial. Strategy II also results in fewer switching actions than Strategy I, and therefore, fewer locked ACs. Both of these features make Strategy II less prone to saturation, as can be seen by its accurate tracking in the 6^{th} minute. The middle and lower plots of Fig. 6.5 demonstrate Strategy II's mode-count controller. Without mode-count control, an under-voltage occurs at meter 469 between minutes 15-18. With mode-count control, the three AC's connected to meter 469 are constrained to have an on-count ≤ 2, and this prevents the under-voltage from occurring.

Overall, our results suggest that when a control strategy has more access to information it has better tracking performance. Notably, this increased information can come in several forms: estimation, coordination between aggregator and operator, communication with TCLs, or detailed modeling. Strategy I uses state estimation and modeling and thereby

has better tracking performance than the benchmark strategy. All three strategies perform better when there is coordination between the operator and aggregator. Finally, Strategy II utilizes more detailed models and more communication than the other two strategies, and has the best tracking performance.

6.7 Chapter Conclusion

We have proposed two strategies for network-safe load control; in the strategies, a third-party aggregator controls TCLs to provide frequency regulation, while the operator overrides the aggregator's control when necessary to ensure network safety. The two control strategies differ markedly in terms of performance and ease of implementation. Strategy II substantially outperforms Strategy I. Its safety controller is able to prevent network violations using less than 3% of the ACs used by Strategy I. This reduction in the number of safety ACs, in combination with a more information-rich tracking controller, translates into improved tracking performance for Strategy II. However, Strategy II would likely be more difficult to implement. Compared with Strategy I, it is more computationally complex—using individual AC models rather than an aggregate model—and requires more communication infrastructure. Choosing between these alternative control strategies will require balancing the costs of implementing the more intensive strategy against the benefits of its improved performance.

Chapter 7

Safety Constraints

This chapter is largely based on a manuscript[7] that has been submitted to a conference.

7.1 Chapter Introduction

The objective of this chapter is to develop a method that ensures a third-party load aggregator's control actions are safe for distribution networks without requiring the operator and aggregator to share private information. We assume that, to maintain a competitive advantage, third-party aggregators prefer to keep their control algorithms private. Distribution operators protect the privacy of their consumers by keeping network models and load measurements private.

Most prior research on aggregate load control for wholesale services either does not consider distribution network safety or does not consider the privacy needs of third-party aggregators. Many proposed strategies are "grid-agnostic" [17]; these strategies ignore network constraints and cannot ensure safe distribution operation (e.g., [3, 52]). A few "grid-safe" strategies have been developed that ensure safe distribution operation. In [93], an AC-OPF is solved to safely provide load frequency control with an aggregation of loads. In [17], a gradient-based algorithm is used to safely control distribution resources to provide frequency regulation. However, the strategies in [93] and [17] do not enable a third-party aggregator to use its own private control algorithm.

A recent strategy proposed in [56] is both grid-safe and suitable for a third-party aggregator. The proposed method certifies whether a distribution network will operate safely under

[7] S.C. Ross and J.L. Mathieu. "A method for ensuring a load aggregator's power deviations are safe for distribution networks". (Under review).

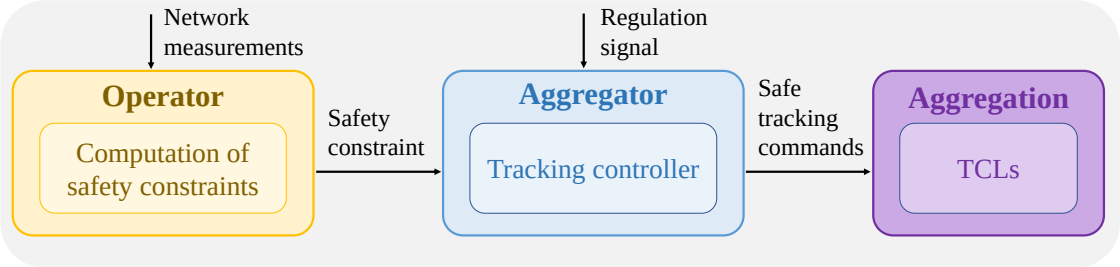


Figure 7.1: Control architecture in which the operator and aggregator coordinate using the safety constraint. (Feedback from the aggregation is not shown.)

any set of possible power injections. If a network is certified safe, then any control actions by a third-party aggregator will be safe. However, this approach leaves open the problem for networks that cannot be certified as safe.

In this chapter, we design and compute an explicit "safety constraint" that ensures safe distribution operation by limiting the size of the power deviations an aggregator can cause across the network, where deviations are with respect to a nominal operating point. The design of the safety constraint is conservative and ensures that the size of the aggregator's deviations are always smaller than the size of the smallest *unsafe* deviation for the network's operating point. We compute the safety constraint's limit by solving a set of optimization problems; each problem is for a particular bus in the network and finds the minimum sized vector of deviations that causes an unsafe voltage at that bus. The minimum size of these minimum sized vectors is the safety constraint's limit.

Fig. 7.1 shows a control architecture in which the operator and aggregator coordinate using the safety constraint. This control architecture protects the privacy of the operator and aggregator: The operator computes the safety constraint's limit given its private network information and measurements, and sends the limit—which reveals little about the network—to the aggregator; the aggregator incorporates the constraint into its private control algorithm, and ensures the resulting regulation-tracking-power-deviations satisfy the constraint.

The main contributions of this chapter are as follows. First, we propose a method of constraining a third-party aggregator—without prescribing a particular control algorithm—such that the aggregator's control actions are safe for distribution networks. Second, through analysis and simulation, we compare the conservativeness of two versions of the safety constraint: one that measures the size of deviations with a 2-norm, and the other with a 1-norm. Third, we propose and prove two propositions that enable a substantial reduction in the number of

optimization problems that must be solved to compute a safety constraint and therefore a reduction in overall computation time.

7.2 Methods

7.2.1 Designing the Safety Constraint

There are four desired criteria for the safety constraint: 1) it should ensure the aggregator's actions will not cause unsafe distribution operation, 2) it should not overly restrict the aggregator's control actions, 3) an operator should be able to calculate the constraint without the aggregator's private information, and 4) the aggregator should be able to adhere to the constraint without the operator's private information.

We design the safety constraint to prioritize safety (criteria 1) over the aggregator's range of control (criteria 2). The constraint takes the general form

$$||\boldsymbol{\Delta}\mathbf{P}^{\mathrm{c}}|| < ||\boldsymbol{\alpha}||, \tag{7.1}$$

where $||\cdot||$ represents the 1-norm or 2-norm, variable $\boldsymbol{\Delta}\mathbf{P}^{\mathrm{c}}$ is the vector of controllable loads' power deviations at each bus, and parameter $\boldsymbol{\alpha}$ is the minimum-sized vector of power deviations that causes unsafe operation somewhere on the network. For brevity, we refer to $\boldsymbol{\alpha}$ as the minimum unsafe deviation. The controllable loads' power deviations are relative to the power the aggregation would have consumed in the absence of control, referred to as the aggregation's "baseline".

The safety constraint (7.1) ensures network safety by the definition of $\boldsymbol{\alpha}$. The constraint is conservative; that is to say, some *safe* values of $\boldsymbol{\Delta}\mathbf{P}^{\mathrm{c}}$ will not satisfy the constraint. The benefit of the constraint is that if we are able to compute the minimum unsafe deviation $\boldsymbol{\alpha}$, then, by definition, the constraint will guarantee safe distribution operation.

7.2.2 Computing the Safety Constraint's Limit

Overview

To determine the safety constraint's limit, we must find the minimum unsafe deviation for a given network at a nominal operating point. We reduce the scope of the problem by considering only the most likely modes of unsafe operation on a network which, for this application, are out-of-range voltage magnitudes [73]. The voltage constraint for bus i is

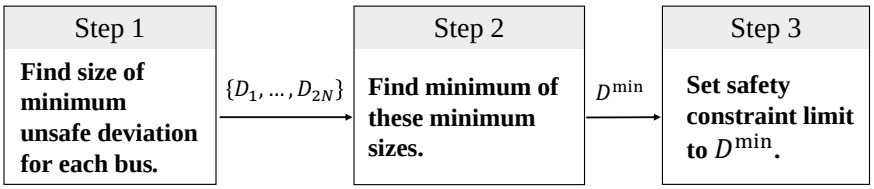


Figure 7.2: Process for computing the safety constraint's limit. Variable D_m is the size of the minimum unsafe deviation that causes an under-voltage at bus m, and D_{m+N} is the size of the minimum unsafe deviation that causes an over-voltage at bus m.

$\underline{V} \leq V_i \leq \overline{V}$, where V_i is the voltage magnitude, $\overline{V} = 1.05$ p.u., and $\underline{V} = 0.95$ p.u. Extending our methods to include all network constraints (e.g., over-currents on lines) is future work.

Fig. 7.2 shows the three main steps of computing the size of the minimum unsafe deviation, referred to as $D^{\min}$. In step 1, we find candidate values for $D^{\min}$ by solving two optimization problems for each bus: the first searches for the minimum deviation that causes an under-voltage at the selected bus; the second does the same but for an over-voltage at the selected bus. In step 2, we set $D^{\min}$ equal to the minimum of the candidate values. In step 3, we set the safety constraint's limit (the right hand side of (7.1)) to $D^{\min}$.

The remainder of Section 7.2.2 proceeds as follows. First, we define the models that will be used in the optimization problems. Then we formulate the optimization problems used to compute the 2-norm safety constraint's limit. Finally, we formulate the problems used to compute the 1-norm safety constraint's limit.

Modeling

We model an N-bus distribution network with aggregator-controlled loads at each bus. We use a single-phase equivalent line model, which assumes balanced power flow and symmetric lines. The network's $N \times N$ conductance and susceptance matrices are denoted as $\mathbf{G}$ and $\mathbf{B}$. We aggregate loads at the bus-level and use a constant power model: each bus has real and reactive power consumption P_i and Q_i, respectively. We separate a bus's power consumption into two components: a controllable component (P_i^{c}, Q_i^{c}) that represents aggregator-controlled loads, and an uncontrollable component (P_i^{uc}, Q_i^{uc}) that represents all other loads. Thus, we have that $P_i = P_i^{\mathrm{c}} + P_i^{\mathrm{uc}}$ and $Q_i = Q_i^{\mathrm{c}} + Q_i^{\mathrm{uc}}$.

We model the aggregator's control actions in terms of bus-level power deviations. When providing balancing, an aggregator controls the aggregation's deviation from baseline such that it tracks a balancing signal. In terms of bus-level power deviations, the aggregator controls loads such that sum of real-power deviations across all buses, $\sum_{i=1}^{N} \Delta P_i^{\mathrm{c}}(k)$, tracks

the balancing signal with sufficient accuracy in every time step k. The variable ΔP_i^{c} is the deviation in P_i^{c} from the bus's baseline value $\widehat{P}_i^{\mathrm{c}}$ (i.e., $\Delta P_i^{\mathrm{c}}(k) = P^{\mathrm{c}}(k) - \widehat{P}_i^{\mathrm{c}}$). We assume the controllable loads have a constant power factor ζ_i, such that any deviation in real power ΔP_i^{c} is accompanied by a deviation in reactive power given by $\Delta Q_i^{\mathrm{c}} = \Delta P_i^{c} \tan(\arccos \zeta_i)$. Finally, the deviations at each bus are naturally constrained by the physical capacities of the loads at that bus; this constraint is given by

$$\underline{P}_i^{\mathrm{c}} \leq (\widehat{P}_i^{\mathrm{c}} + \Delta P_i^{\mathrm{c}}) \leq \overline{P}_i^{\mathrm{c}}, \tag{7.2}$$

where $\underline{P}_i^{\mathrm{c}}$ and $\overline{P}_i^{\mathrm{c}}$ reflect the loads' aggregate physical capacity.

2-Norm Safety Constraint

In *step 1*, we find the minimum-sized deviation (as measured by the squared 2-norm) that causes an under-voltage at bus m by solving the following optimization problem:

$$\text{minimize} \quad \sum_{i=1}^{N} \left(\Delta P_i^{\mathrm{c}}\right)^2 \tag{7.3a}$$

subject to

$$V_m \leq \underline{V}, \tag{7.3b}$$

$$\underline{P}_i^{\mathrm{c}} \leq (\widehat{P}_i^{\mathrm{c}} + \Delta P_i^{\mathrm{c}}) \leq \overline{P}_i^{\mathrm{c}} \qquad \forall i \in \mathcal{N}, \tag{7.3c}$$

$$\Delta Q_i^{\mathrm{c}} = \beta_i \Delta P_i^{c} \qquad \forall i \in \mathcal{N}, \tag{7.3d}$$

$$P_i = \Delta P_i^{\mathrm{c}} + \widehat{P}_i^{\mathrm{c}} + P_i^{\mathrm{uc}} \qquad \forall i \in \mathcal{N}, \tag{7.3e}$$

$$Q_i = \Delta Q_i^{\mathrm{c}} + \widehat{Q}_i^{\mathrm{c}} + Q_i^{\mathrm{uc}} \qquad \forall i \in \mathcal{N}, \tag{7.3f}$$

$$P_i = V_i \sum_{k=0}^{N-1} V_k (G_{ik} \cos \phi_{ik} + B_{ik} \sin \phi_{ik}) \qquad \forall i \in \mathcal{N}, \tag{7.3g}$$

$$Q_i = V_i \sum_{k=0}^{N-1} V_k (G_{ik} \sin \phi_{ik} - B_{ik} \cos \phi_{ik}) \qquad \forall i \in \mathcal{N}, \tag{7.3h}$$

$$\phi_0 = 0, \tag{7.3i}$$

$$V_0 = V_{\mathrm{set}}. \tag{7.3j}$$

The optimal objective value for bus m's is denoted D_m. The set $\mathcal{N}$ is the set of all buses in the network. The decision variables in this problem are ΔP_i^{c}, ΔQ_i^{c}, P_i, Q_i, V_i $\forall$ $i \in \mathcal{N}$ and ϕ_{ik} $\forall$ $i, k \in \mathcal{N}$, where ϕ_{ik} is the voltage angle difference between buses i and k.

The objective function (7.3a) and constraint (7.3b) are opposing forces on the size of $\mathbf{\Delta P}^{\text{c}}$: the objective function minimizes the size of the deviations, but the deviations must be large enough such that an under-voltage occurs at bus m. A deviation in power is necessary to create an under-voltage at bus m because we assume that the network's voltages are within the operational range $[\underline{V}, \overline{V}]$ at the nominal operating point.

Constraints (7.3c) and (7.3d) model the controllable loads. Constraint (7.3c) restricts power deviations at each bus according to the controllable loads' physical capacities; the baseline value $\widehat{P}_i^{\text{c}}$ is assumed known. Constraint (7.3d) enforces a constant power factor for controllable loads, where $\beta_i = \tan(\arccos \zeta_i)$.

Constraints (7.3e)-(7.3f) sum the controllable and uncontrollable components of the power consumption at each bus. The uncontrollable components P_i^{uc} and Q_i^{uc} are assumed known.

Constraints (7.3g)-(7.3j) model the power flow in the network and define the slack bus. Constraints (7.3g)-(7.3h) are the standard power flow equations, where the real and reactive power consumption at bus i must be balanced by the sum of all real and reactive power flows into bus i. Constraint (7.3j) sets the substation bus as the reference for voltage angles. Constraint (7.3j) fixes the substation bus's voltage magnitude as a constant; we assume the value of parameter V_{set} is set by the distribution operator (e.g., $V_{\text{set}} = 1.0$ p.u.).

We also solve an over-voltage problem for each bus m in the network. The over-voltage problem is given by

$$\text{minimize} \quad \sum_{i=1}^{N} \left(\Delta P_i^{\text{c}}\right)^2 \tag{7.4a}$$

subject to constraints (7.3c)-(7.3j) and

$$V_m \geq \overline{V}. \tag{7.4b}$$

The optimal objective value for bus m's over-voltage problem is denoted D_{N+m}.

The optimization problems (7.3) and (7.4) are non-convex because of the non-linear constraints (7.3g)-(7.3h) that model the network's AC power flow. In this chapter, we use a non-linear solver to solve (7.3) and (7.4); the solver finds locally optimal solutions, so global optimality is not guaranteed. In future work, we plan to apply a convex relaxation to the AC power flow equations in order to identify a lower bound on the globally optimal objective value of the original problem; this lower bound will make D_m conservative but will ensure no smaller-sized unsafe deviation exists.

In *step 2*, we find D^{min} the size of the minimum of the minimum unsafe deviations found

in step 1. We simply take the minimum across the set of optimal objective values:

$$D^{\min} = \min_{m \in \mathcal{M}} D_m, \tag{7.5}$$

where $\mathcal{M}$ is the set of problems for which a feasible solution was found.

In *step 3*, we set the limit of the 2-norm safety constraint to $D^{\min}$. The constraint is then given by

$$\sum_{i=1}^{N} \left(\Delta P_i^{\mathrm{c}}\right)^2 < D^{\min}. \tag{7.6}$$

1-Norm Safety Constraint

We use the same general 3-step method to compute the limit of the 1-norm safety constraint. In *step 1*, the optimization problem formulation is identical to (7.3) except the objective function is in terms of the 1-norm. The under-voltage problem is given by

$$\begin{aligned} &\text{minimize} \quad \sum_{i=1}^{N} \left|\Delta P_i^{\mathrm{c}}\right| \\ &\text{subject to constraints (7.3b)} - \text{(7.3j)}. \end{aligned}$$

The optimal objective values are denoted by F_m for the under-voltage problems and F_{N+m} for the over-voltage problems. To improve solvability, we reformulate the absolute value terms in the objective function such that the objective function is linear (see chapter IX of [55] for details). In *step 2*, we find $F^{\min}$ the minimum of the minimum unsafe deviations found in *step 1*. We calculate $F^{\min}$ as $F^{\min} = \min_{m \in \mathcal{M}} F_m$, where $\mathcal{M}$ is the set of problems for which a feasible solution was found. Finally, in *step 3*, we define the 1-norm safety constraint as

$$\sum_{i=1}^{N} \left|\Delta P_i^{\mathrm{c}}\right| < F^{\min}. \tag{7.8}$$

7.2.3 Conservativeness of 1 and 2-Norm Safety Constraints

The less conservative of the two constraints should be used because it will allow the aggregator more feasible control actions. However, determining which constraint is less conservative can be a challenge. The conservativeness of a constraint cannot be determined *a priori* because it depends on the minimum unsafe deviation that has been found. Fig. 7.3 demonstrates this point with illustrations of the 1-norm and 2-norm approximations of the set of

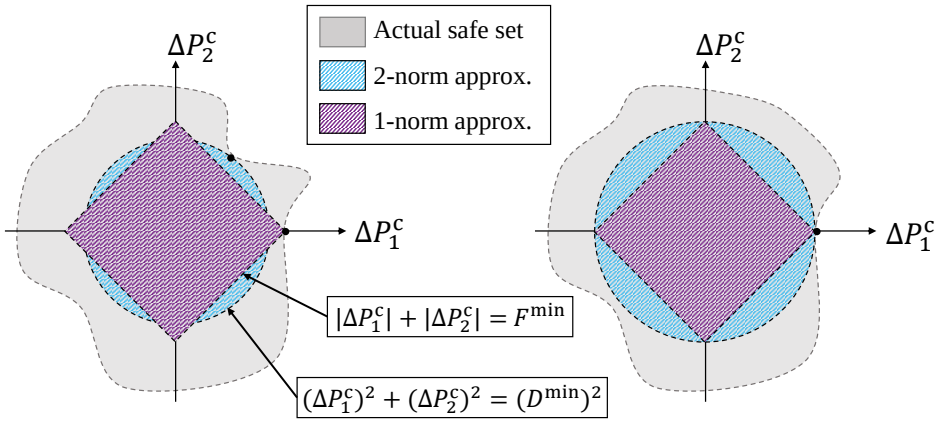


Figure 7.3: Approximations of the set of safe deviations given by the 1-norm and 2-norm constraints. Left: neither the 2-norm or 1-norm approximation is uniformly less conservative than the other. Right: the 2-norm approximation is uniformly less conservative.

safe power deviations for two different operating points. (Note, for illustration purposes, we consider deviations in only 2 dimensions.) On the right side of Fig. 7.3, both methods have found the same minimum unsafe deviation; this is the point where the boundaries of the approximations intersect with the boundary of the actual set. Because the minimum unsafe deviation lies on the x-axis, the 2-norm constraint is uniformly less conservative than the 1-norm constraint; in this case, the 2-norm constrain should be used since it allows the aggregator more feasible control actions. On the left side of Fig. 7.3, the 1-norm and 2-norm methods have found different minimum unsafe deviation points, and neither constraint is uniformly less conservative than the other.

For cases in which neither constraint is uniformly less conservative than the other, we propose choosing the constraint that allows the larger maximum balancing capacity. In general, an aggregator is compensated for the size of its capacity and would prefer the safety constraint that maximizes its balancing capacity. We find C^{II}, the maximum balancing capacity that the 2-norm safety constraint will allow, by maximizing the capacity $\sum_{i=1}^{N} \Delta P_i^c$ subject to (7.6). We find that the solution must have all deviations equal (i.e., $\Delta P_i^c = \Delta P_j^c \ \forall \ (i,j) \in \mathcal{N}$). Setting both sides of (7.6) equal and all deviations equal gives us the following result

$$C^{\mathrm{II}} = N\sqrt{D^{\min^-}/N}, \tag{7.9}$$

where $D^{\min^-}$ is just slightly less than $D^{\min}$. The value of C^{I}, the maximum balancing capacity

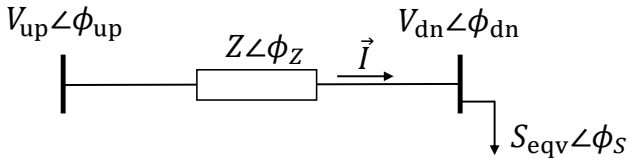


Figure 7.4: Two-bus equivalent system. Every pair of adjacent buses in a network can be represented by this two bus system.

that the 1-norm constraint will allow, is clear upon inspection and is given by

$$C^{\text{I}} = F^{\text{min}^-}, \tag{7.10}$$

where F^{min^-} is just slightly less than F^{min}. To maximize balancing capacity, if $C^{\text{II}} > C^{\text{I}}$, the 2-norm constraint should be used; otherwise, the 1-norm constraint should be used.

7.2.4 Reducing Problem Size

Solving $2N$ optimization problems to compute a safety constraint may be too computationally intensive for real-time operations. We propose reducing computation time by reducing the number of buses for which an optimization problem must be solved. We can eliminate a problem if we can guarantee it is infeasible or its solution will not have the minimum objective value of the full set of problems.

We propose two conditions—one for the under-voltage problem and one for over-voltage problem—that if satisfied enable the problem to be eliminated. To derive these conditions, we find the loading conditions that guarantee two adjacent buses in a radial network, referred to as a "bus-pair", have decreasing voltage magnitudes in the downstream direction (i.e., away from the substation). We represent a generic bus-pair with the two-bus equivalent system in Fig. 7.4. (Note we use vector notation $\vec{X}$ and polar notation $X\angle\phi_X$ interchangeably to represent a vector in the complex plane.) In Fig. 7.4, $\vec{S}_{\text{eqv}}$ is the apparent power of the equivalent load at the downstream bus and is equal to the sum of all loads connected to and downstream of the bus in the actual network, as well as their associated line-losses. Variables $\vec{V}_{\text{up}}$ and $\vec{V}_{\text{dn}}$ are the voltages at the upstream and downstream buses, respectively. Parameter $\vec{Z}$ is the line's impedance, and $\vec{I}$ is the line's current flow.

We derive the loading conditions that ensure $V_{\text{dn}} \leq V_{\text{up}}$ as follows. By Ohm's law we have

that $\vec{V}_{\text{dn}} = \vec{V}_{\text{up}} - \vec{I}\vec{Z}$. After a few simple operations, we transform this expression into

$$V_{\text{dn}}^2 = V_{\text{dn}} V_{\text{up}} \cos(\phi_{\text{up}} - \phi_{\text{dn}}) - ZS_{\text{eqv}} \cos(\phi_{\text{Z}} - \phi_{\text{S}}) \tag{7.11a}$$

$$0 = V_{\text{dn}} V_{\text{up}} \sin(\phi_{\text{up}} - \phi_{\text{dn}}) - ZS_{\text{eqv}} \sin(\phi_{\text{Z}} - \phi_{\text{S}}), \tag{7.11b}$$

where the voltage angle of the downstream bus has been defined as the reference angle. Next, we make the small angle approximation: $\cos(\phi_{\text{up}} - \phi_{\text{dn}}) \approx 1$, which is suitable for two adjacent buses on a distribution network. With this approximation, (7.11a) simplifies to

$$V_{\text{dn}}^2 = V_{\text{dn}} V_{\text{up}} - ZS_{\text{eqv}} \cos(\phi_{\text{Z}} - \phi_{\text{S}}). \tag{7.12}$$

If the last term in (7.12) satisfies

$$ZS_{\text{eqv}} \cos(\phi_{\text{Z}} - \phi_{\text{S}}) \geq 0, \tag{7.13}$$

then $V_{\text{dn}} V_{\text{up}} - ZS_{\text{eqv}} \cos(\phi_{\text{Z}} - \phi_{\text{S}}) \leq V_{\text{dn}} V_{\text{up}}$. Substituting (7.12) into the latter inequality, we have $V_{\text{dn}}^2 \leq V_{\text{dn}} V_{\text{up}}$ and thus $V_{\text{dn}} \leq V_{\text{up}}$, since V_{dn} is positive by definition.

We have derived the following loading condition: if (7.13) is satisfied, then $V_{\text{dn}} \leq V_{\text{up}}$. Since the magnitudes S_{eqv} and Z are positive by definition, (7.13) is satisfied if $\cos(\phi_{\text{Z}} - \phi_{\text{S}}) \geq 0$. Thus, the loading condition simplifies to: if

$$-\frac{\pi}{2} < \phi_{\text{Z}} - \phi_{\text{S}} < \frac{\pi}{2}, \tag{7.14}$$

then $V_{\text{dn}} \leq V_{\text{up}}$.

We state the first of two propositions for reducing the overall problem size:

> *Proposition 1*: If a bus-pair satisfies (7.14) for all possible operating points, then the under-voltage problem for the upstream bus in the pair can be eliminated from the set of problems that are necessary to solve.

The proof follows. Let us assume (7.14) is satisfied for a given bus-pair for all operating points. According to the loading condition, we have $V_{\text{dn}} \leq V_{\text{up}}$ for all operating points. Thus $V_{\text{dn}} \leq V_{\text{up}}$ for the optimal solution of the under-voltage problem for the upstream bus. This solution is also a feasible solution for the downstream bus's under-voltage problem because $V_{\text{dn}} \leq V_{\text{up}} \leq \underline{V}$, which satisfies constraint (7.3b). Thus the optimal objective value D_{up} is a candidate value for D_{dn} and D_{up} must be greater than or equal to D_{dn}. If

$D_{\text{up}} \geq D_{\text{dn}}$, then the upstream bus's optimal objective value is not a candidate for $D^{\min}$ and the under-voltage problem for the upstream bus can be eliminated.

> *Proposition 2*: If all bus-pairs on the path between bus m and the substation satisfy (7.14) for all possible operating points, then the over-voltage problem for bus m can be eliminated from the set of problems that are necessary to solve.

The proof follows. Let us assume (7.14) is satisfied for all bus-pairs between bus m and the substation. The voltage magnitude is necessarily non-increasing along this path since $V_{\text{dn}} \leq V_{\text{up}}$ for all of the bus-pairs. Thus the voltage magnitude of bus m must be less than or equal to that of the substation. Since the voltage at the substation is regulated within operational limits (i.e., $\underline{V} \leq V_0 \leq \overline{V}$), bus m's voltage magnitude cannot be greater than $\overline{V}$. Thus constraint (7.4b) cannot be satisfied and the over-voltage problem is infeasible for bus m. Because the problem is infeasible, it can be eliminated.

7.3 Case Study

7.3.1 Study Setup

In the case study, we use a 56-bus distribution feeder model that is a modified version of the IEEE 123-bus test feeder. The 56-bus model has balanced loads and symmetric lines, enabling a single-phase equivalent model. Full details of the model are provided in [5]. Fig. 7.5 shows the network's radial topology and its range of voltage magnitudes at the nominal operating point. In most of the case study, we assume there are no capacitor banks on the network and set the substation voltage to 1.02 p.u., which ensures that there are no under-voltage violations at the nominal operating point. We also assume there are no voltage regulators except at the substation. Extending the optimization problem (7.3) to include in-line voltage regulators is future work. We solve the proposed optimization problems using the non-linear solver Ipopt. We initialize the solver at the network's nominal operating point.

We use the network's nominal loading data to determine the operating points for our model's uncontrollable and controllable loads. At each bus, we assume 50% of the nominal real-power consumption is controllable and set $\widehat{P}_i^{\text{c}}$ equal to it. We assume the controllable load's power factor ζ_i is 0.95 lagging for all buses. At each bus, the remaining nominal power consumption is assigned to the uncontrollable load (P_i^{uc}, Q_i^{uc}). When the controllable loads are at baseline, the network's loading matches that of the nominal data. Finally, we assume

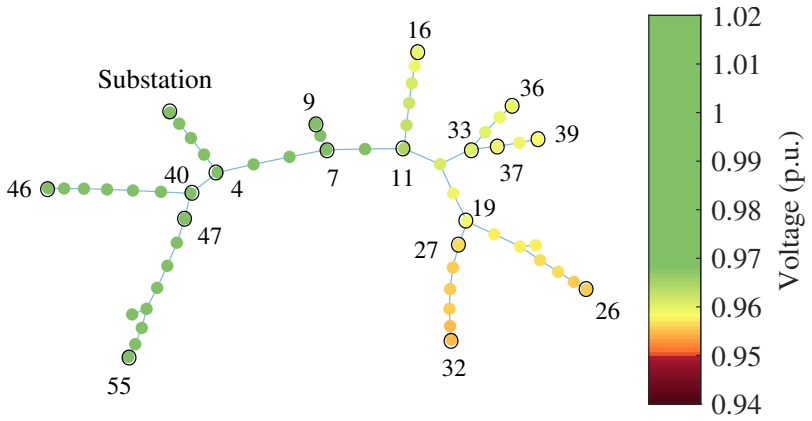


Figure 7.5: 56-Bus distribution network used in case study. Voltage magnitudes are shown for the nominal operating point.

the physical capacity of the controllable loads at each bus is $\pm 80\%$ of their baseline power (i.e., $\underline{P}_i^{\mathrm{c}} = 0.2\widehat{P}_i^{\mathrm{c}}$ and $\overline{P}_i^{\mathrm{c}} = 1.8\widehat{P}_i^{\mathrm{c}}$).

7.3.2 Demonstration of 2-Norm Method

We demonstrate step 1 of the 2-norm method by showing the optimal solution of the over-voltage optimization problems for a particular bus. The top plot of Fig. 7.6 shows the network's voltage magnitudes for the optimal solution to the under-voltage, 2-norm problem for bus 20. Voltages at and around bus 20 are close to or under the lower voltage limit, as indicated by dark red. Fig. 7.6 (middle) shows the exact voltage magnitudes at each bus. At its nominal operating point, bus 20's voltage is 0.9571 p.u. and decreases to 0.9500 p.u for the optimal solution. Because bus 20 is not a terminal bus, under-voltages also occur at downstream buses 21-26, as well as at adjacent buses 27-32. Fig. 7.6 (bottom) shows the change in real power at each bus with respect to nominal for this solution. Power consumption increases at all load-buses: the largest increases occur at or downstream of bus 20, and the smallest increases occur close to the substation (buses 1-4). This pattern shows which buses' power deviations have the most influence over bus 20's voltage, with larger deviations indicating larger influence.

We demonstrate step 2 of the method by selecting the bus/problem that has the minimum optimal objective value. Fig. 7.7 shows the optimal objective value for each under-voltage problem for the network. Problem 32 has the minimum objective value, and corresponds to a terminal bus far from the substation. For buses close to the substation, no feasible

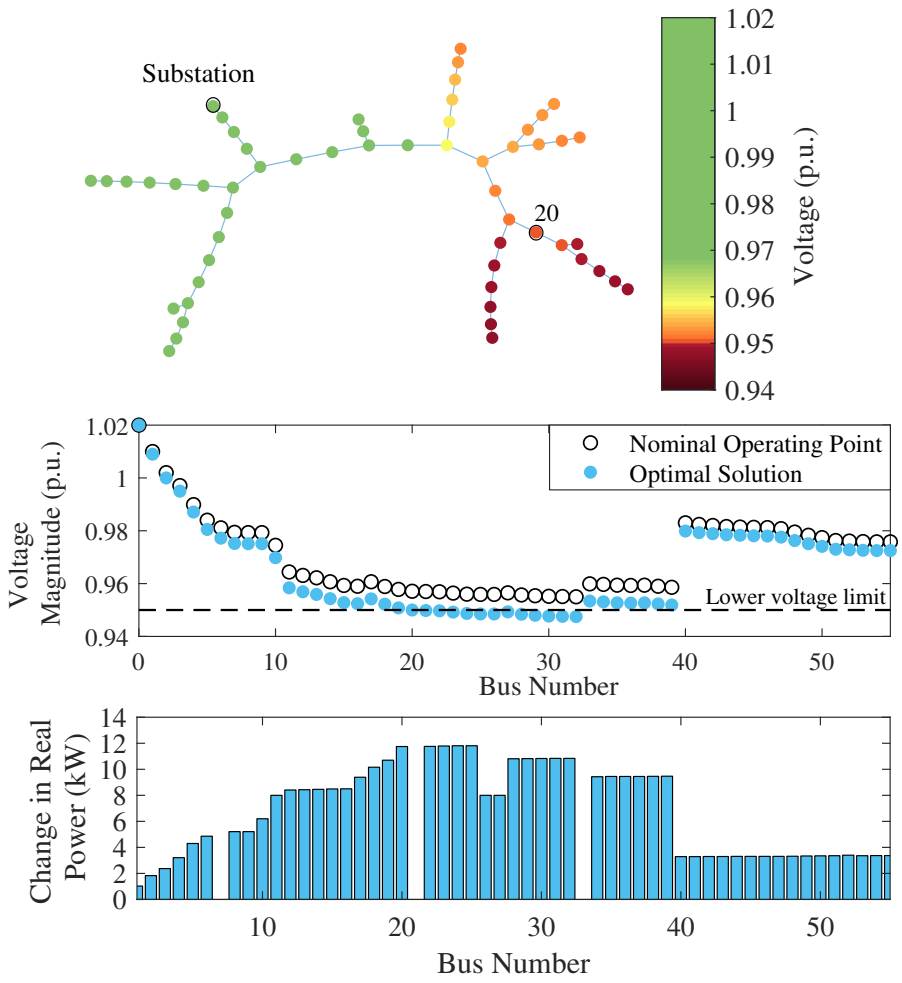


Figure 7.6: Solution to the under-voltage, 2-norm problem for bus 20. Top and middle plots: the voltage at bus 20 is exactly at the lower limit (0.95 p.u.); voltages downstream of bus 20 are below the limit. Bottom plot: Each load-bus contributes some change in real power to achieve the under-voltage at bus 20; change is relative to the nominal operating point.

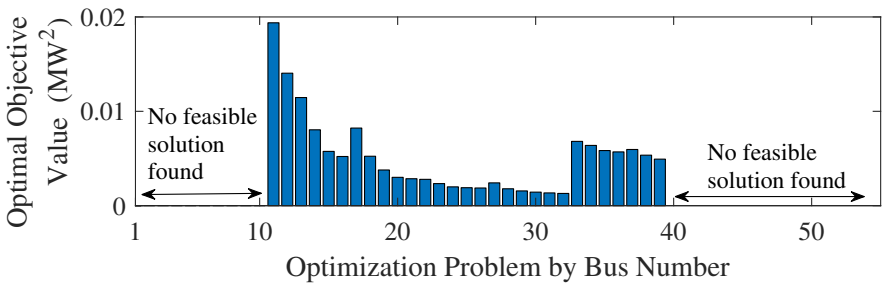


Figure 7.7: Optimal objective values for all 2-norm, under-voltage problems. Minimum value is at bus 32. No feasible solutions are found for buses closer to the substation, i.e., buses 1-10 and 40-55 (see Fig. 7.5 for numbering).

solution is found, and likely none exists, because of the controllable loads' physical capacity limits (see (7.3c)). In addition, no feasible solutions were found for any of the over-voltage problems, which is unsurprising since there are no positive power injections in the network (e.g., from capacitor banks or photovoltaic systems). For all of the optimal solutions found, that of bus 32 has the minimum optimal objective value, so we set $D^{\min} = D_{32}$, which in this case is 0.0013 MW^2. After $D^{\min}$ is known, the last step of the method is simply to set the limit of the 2-norm safety constraint such that $\sum_{i=1}^{N} \left(\Delta P_i^{\mathrm{c}}\right)^2 < D^{\min}$.

7.3.3 Comparing and Testing the 1-Norm and 2-Norm Methods

We compare the 1-norm and 2-norm methods for the under-voltage optimization problems. First, we consider an example problem, again for the under-voltage at bus 20. Fig. 7.8 shows the percent change in real power for both the 1-norm and 2-norm optimal solutions with respect to the nominal operating point. As the figure shows, the two solutions are strikingly different. The solutions differ because of the methods' different objective functions. The 2-norm objective function penalizes an incremental increase to a large deviation more than to a small deviation; this preference for small deviations causes the 2-norm problem to distribute the deviations across all load-buses. In contrast, the 1-norm objective function penalizes all incremental increases equally; this causes the 1-norm problem to concentrate the deviations among buses with the largest influence over the constrained bus's voltage.

Next, we compare the maximum balancing capacities that the two methods allow. For both methods, the problem for bus 32 has the minimum optimal objective value; the values are $D^{\min} = 0.0013$ MW^2 and $F^{\min} = 0.163$ MW. Using (7.9) and (7.10), we find the maximum balancing capacities $C^{\mathrm{II}} = 0.260$ MW and $C^{\mathrm{I}} = 0.163$ MW, for the 2-norm and 1-norm methods, respectively. Thus, for this operating point, the 2-norm method is preferred

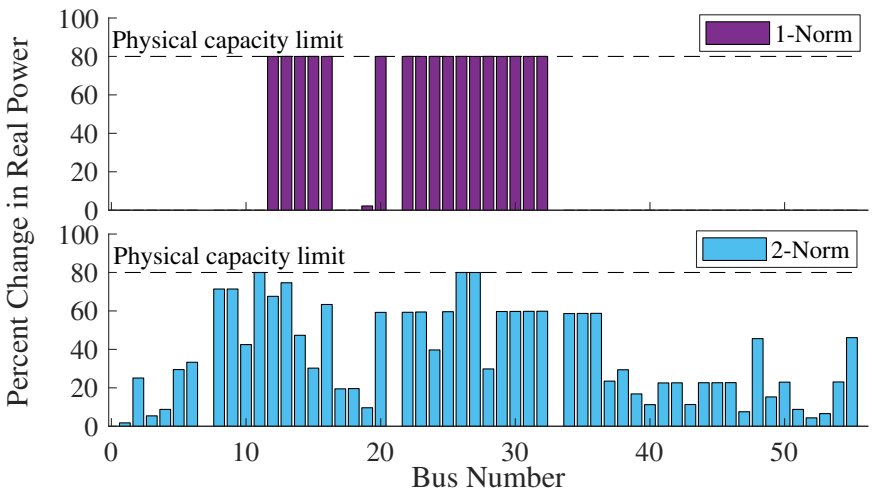


Figure 7.8: Comparison of 1-norm and 2-norm solutions to the under-voltage problem for bus 20. The 1-norm and 2-norm methods find different optimal solutions. Note the physical capacity limit of controllable loads is 80% of their baseline for all buses.

because it allows for a larger balancing capacity.

We numerically test the 1-norm and 2-norm safety constraints with the optimal solutions of the other method. The 2-norm safety constraint should exclude the minimum unsafe deviations found by the 1-norm problem, and the 1-norm safety constraint should exclude those found by the 2-norm problem. Fig. 7.9 shows the results of these tests. The y-axis values are calculated by evaluating the objective function of the method being tested at the optimal solutions of the other method. As shown in the top plot of Fig. 7.9, the 1-norm safety constraint passes its test: all of the 2-norm method's minimum unsafe deviations lie outside of the safe region defined by the 1-norm safety constraint. Similarly, as shown in the bottom plot of Fig. 7.9, the 2-norm safety constraint passes its test.

7.3.4 Reducing Problem Size

We apply Propositions 1 and 2 to reduce the problem size for the 56-bus network, as well as a modified version of the network. The modified network has capacitor banks located at buses 26, 28, 29, and 30 that cause a voltage rise along the line from bus 19 to 26 (see [5] for details). Because of this voltage rise, fewer problems for the modified network should qualify for elimination. For Propositions 1 and 2 to hold for a given bus-pair, constraint (7.14) must be satisfied by both the minimum and maximum possible values of ϕ_S for the bus-pair. In this analysis, we evaluate (7.14) at two extreme operating points that approximate the

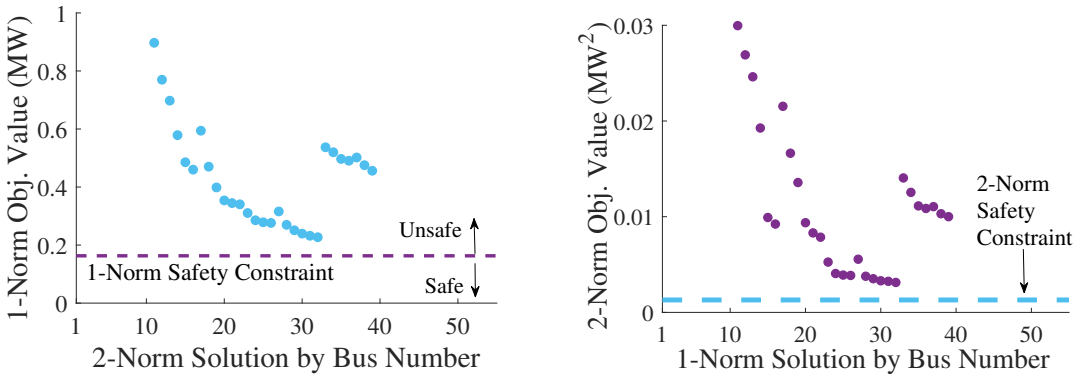


Figure 7.9: Verification of safety constraints. The 1-norm safety constraint correctly excludes the minimum unsafe deviations found by the 2-norm method (left); the 2-norm safety constraint passes a similar test (right).

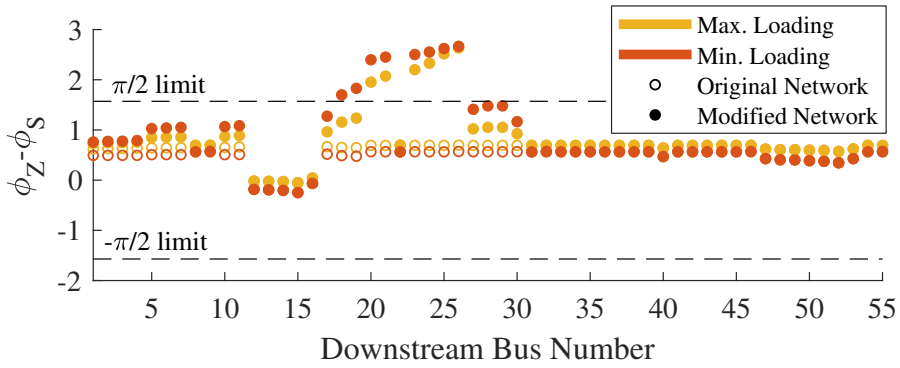


Figure 7.10: Loading conditions for each bus-pair. Bus-pairs with $\phi_Z - \phi_S > \pi/2$ do not satisfy (7.14). Fewer optimization problems can be eliminated when this condition is not met.

operating points that minimize and maximize ϕ_S. These extreme points are: 1) "maximum loading" in which the controllable loads consume maximum power and 2) "minimum loading" in which they consume minimum power. Finding the actual operating points with maximum and minimum ϕ_S is future work.

Fig. 7.10 shows which bus-pairs satisfy (7.14) and indicates which problems can be eliminated. points that lie above the $\pi/2$ limit indicate bus-pairs that do not satisfy the constraint. In the original network, all of the bus-pairs satisfy the constraint. By Proposition 1, we can eliminate the under-voltage problems for the upstream bus of each of these bus-pairs; after this elimination, only under-voltage problems for the network's 11 terminal buses remain. By Proposition 2, we can eliminate all over-voltage problems because voltages are always decreasing. In the modified network, eight bus-pairs do not satisfy (7.14). As expected, these buses are where the voltage rise occurs due to capacitor banks. By Proposition 1, we can eliminate under-voltage problems for all non-terminal buses except for the identified eight upstream buses; after the allowed eliminations, under-voltage problems for 11 terminal buses and 8 non-terminal buses remain. To apply Proposition 2 to the modified network, we identify bus 18 as the downstream bus closest to the substation for which (7.14) is not satisfied. By Proposition 2, all buses—except for bus 18 and those downstream of 18—can be eliminated. Thus, the over-voltage problem must be solved for 15 buses.

Using the proposed reduction techniques, we are able to drastically reduce the size of the overall problem. For the original network, the number of total optimization problems decreased from 104 to 11 for the original network and from 104 to 34 for the modified network.

7.4 Chapter Conclusion

We have proposed a method for constraining a load aggregator's control actions to ensure the safe operation of the distribution network. The proposed safety constraint can be computed by a distribution operator and adhered to by a third-party aggregator with minimal sharing of private information between the two entities. The safety constraint is, by design, conservative. We compared two versions of the constraint to determine if one was uniformly less conservative than the other. Although we were unable to draw general conclusions, in a case study we found that the 2-norm constraint allowed for a larger balancing capacity than the 1-norm constraint. To reduce the time it takes to compute the safety constraint, we proposed two conditions under which optimization problems can be eliminated from the set

of necessary problems. In the case study, we found this reduction technique to be very effective. Because the method's conservatism and computational intensity is network dependent, some networks will be better suited for the safety-constraint method than others.

Chapter 8

Conclusion

Load flexibility will be an essential component of reliable and cost-effective operation of future, low-carbon power systems. This dissertation has developed methods that enable the aggregation and control of loads at a large scale, while avoiding negative effects on the distribution network. We conclude by summarizing the dissertation's key findings and methodological contributions as well as avenues for future research.

8.1 Summary of Key Methods and Results

In Chapter 2, we designed two methods for updating an aggregate TCL model to account for changing conditions and biased disturbances that affect TCL behavior. One method uses state measurements and observed transitions between states to update the model; the other method uses output tracking error. With the state-based update, the proposed control system outperformed the benchmark control systems in all tests, demonstrating that state-based updating is a powerful method for adapting to changing load behaviors. Output-based updating is also promising, but because the update depends on and also influences the controller's tracking performance, the method requires additional testing and analysis to determine if it is stable in all possible scenarios.

In Chapter 3, we conducted a simulation study of realistic distribution networks with a high penetration of loads controlled to provide regulation. The study's results indicate that load-based regulation causes network voltages to vary more and network transformers to age less rapidly on average. Notably, the networks did not experience an increase in constraint violations when subject to load control, suggesting that many networks may be capable of hosting aggregations at a large scale without the use of network-safe control

techniques. However, in a separate study of a test network with a high penetration of EVs, load control caused the prevalence of voltage violations to increase from 0.17% of nodes to 0.84%. Thus, as EV and PV adoption increases, coordinating load control with distribution network operation will become more important.

In Chapter 4, we proposed a network-safe load control strategy in which an operator blocks TCLs when necessary to ensure network safety. We designed a tracking controller that successfully estimates and compensates for the number of loads that are blocked. Here, the key innovation was extending the Markov chain model to include blocked and unblocked states. The proposed controller outperformed a benchmark controller in 11 of 12 scenarios with different blocking and feedback levels. Results indicate that the proposed controller is more advantageous in scenarios with high blocking levels and when feedback to the aggregator about blocking is minimal.

In Chapter 5, we applied the mode-count control algorithm developed in [62] to groups of TCLs to manage at-risk network constraints. We proposed setting a group's upper and lower on-count limits equal to each other to reduce the variability of the group's power consumption. In a case study, this approach successfully reduced voltage deviations at the end of a long distribution line, demonstrating the method's ability to manage network constraints. One drawback of the algorithm is that it can cause TCLs to cycle on and off rapidly. To address this issue, we designed a mode-count algorithm to satisfy specified TCL cycling constraints. We proposed new conditions for the feasibility of upper and lower on-count limits given the new algorithm. In a simulation test, the new algorithm successfully controlled 1000 TCLs with cycling constraints over the course of 12 hours to satisfy an upper on-count limit; the original algorithm failed the same test, demonstrating the usefulness of the new algorithm.

In Chapter 6, we proposed and compared two network-safe control strategies for TCLs with cycling constraints. In the first strategy, the aggregator provides tracking with an aggregate-model based controller, and the operator intervenes by blocking particular TCLs when necessary for network safety; in the second strategy, the aggregator uses an individual-model based controller, and the operator uses the new mode-count control algorithm. In a case study, the second strategy achieved a root-mean-square tracking error of 0.10% of the TCLs' baseline power consumption while removing fewer than 1% of TCLs from the aggregator's control; the first strategy achieved a 0.70% tracking error while removing approximately 15% of TCLs. These results indicate that mode-count control is a more effective safety intervention than blocking and would likely be preferred by aggregators, since

it removes fewer TCLs from the tracking group. However, mode-count control requires substantially more computational effort and communication infrastructure than blocking, and for this reason, operators are likely to prefer blocking. It is more challenging to draw conclusions about the two tracking controllers, since their performance is affected by the number of TCLs removed from the tracking group. The testing of additional strategies (e.g., blocking paired with individual-model based tracking control, and mode-count control paired with aggregate-model based tracking control) would enable like-for-like comparisons of the different controllers.

In Chapter 7, we designed a constraint that ensures network safety by limiting the size of the deviations in power that an aggregator can cause across a network. We proposed two versions of the safety constraint, based on the 2-norm and 1-norm of deviations, and found that neither version is guaranteed to be less conservative than the other. The constraint's upper limit is computed by solving a set of optimization problems, one for each voltage constraint on the network. To improve computational efficiency, we derived conditions that identify problems that will not determine the constraint's limit and therefore do not need to be solved. We applied these conditions to a test network and reduced the number of problems by 67%. Within the context of network-safe load control for energy balancing, the safety constraint could be computed by the distribution operator and sent to the aggregator who would ensure their power deviations for energy balancing satisfies the safety constraint.

8.2 Future Research

This work has generated several avenues for additional research. We list research topics that are direct extensions of this dissertation first and additional topics second.

Extensions to this dissertation:

- Application of Chapter 7's optimization methods to determine which network constraints are most at risk of violation.
- Development of an online method for determining which TCLs to remove from the tracking population to protect at-risk network constraints. This would, in a sense, complete the strategies proposed in Chapter 6.
- Development of a network-safe load control strategy that uses the safety constraint proposed in Chapter 7. This strategy would likely restrict an aggregator more than

the strategies proposed in Chapter 6, resulting in more conservative (safer) distribution operation.

- Proof of Chapter 5's proposed feasibility conditions and mode-count control algorithm that account for TCLs' cycling constraints.

Additional topics:

- Characterization of the mismatch between simulated TCL populations and actual TCL populations. A model of this mismatch would enable researchers to improve the realism of the simulated plant and enhance the reliability of controllers tested on simulated plants.

- Design of a tracking controller for an aggregation of TCLs with cycling constraints that makes the optimal trade-off between immediate tracking accuracy, which may require many switching actions and result in many locked TCLs, and future tracking accuracy, which suffers if too many TCLs are locked from past actions.

- Development of a method to fairly constrain different actors on a distribution network. In this dissertation, only one load-aggregator was considered, but multiple aggregators of different types could be active on the same network.

- Design of network-safe control strategies for distributed energy resource (DER) aggregations, where DERs are resources that can inject as well as consume power. This topic is particularly relevant as the Federal Energy Regulatory Commission (FERC) is currently grappling with how to enable DER aggregations to participate in wholesale markets without burdening distribution operators, who are not under FERC's jurisdiction (see [21]).

Bibliography

[1] K. C. Armel, A. Gupta, G. Shrimali, and A. Albert. "Is disaggregation the holy grail of energy efficiency? The case of electricity". In: *Energy Policy* 52 (2013), pp. 213–234.

[2] K. Astrom and R. Murray. *Feedback Systems: An Introduction for Scientists and Engineers*. Princenton University Press, 2008.

[3] S. Bashash and H. Fathy. "Modeling and control of aggregate air conditioning loads for robust renewable power management". In: *IEEE Transactions on Control Systems Technology* 21.4 (2013), pp. 1318–1327.

[4] J. Bay. *Fundamentals of Linear State Space Systems*. Electrical and Computer Engineering Faculty Scholarship, 1999. URL: `https://orb.binghamton.edu/electrical%5C_fac/3/`.

[5] S. Bolognani and S. Zampieri. "On the existence and linear approximation of the power flow solution in power distribution networks". In: *IEEE Transactions on Power Systems* 31.1 (2016), pp. 163–172.

[6] T. Borsche, F. Oldewurtel, and G. Andersson. "Minimizing communication cost for demand response using state estimation". In: *2013 IEEE Grenoble Conference*. Grenoble, France, June 2013.

[7] D. Callaway. "Tapping the energy storage potential in electric loads to deliver load following and regulation, with application to wind energy". In: *Energy Conversion and Management* 50 (2009), pp. 1389–1400.

[8] D. S. Callaway and I. A. Hiskens. "Achieving controllability of electric loads". In: *Proceedings of the IEEE* 99.1 (2011), pp. 184–199.

[9] E. Can Kara, Z. Kolter, M. Berges, B. Krogh, G. Hug, and T. Yuksel. "A moving horizon state estimator in the control of thermostatically controlled loads for demand response". In: *IEEE International Conference on Smart Grid Comm.* Vancouver, Canada, Oct. 2013.

[10] C. Y. Chang, W. Zhang, J. Lian, and K. Kalsi. "Modeling and control of aggregated air conditioning loads under realistic conditions". In: *IEEE PES Innovative Smart Grid Technologies Conference* (2013).

[11] Y. Chen, A. Bušić, and S. P. Meyn. "State estimation for the individual and the population in mean field control with application to demand dispatch". In: *IEEE Transactions on Automatic Control* 62.3 (2017), pp. 1138–1149.

[12] M. Chertkov, D. Deka, and Y. Dvorkin. "Optimal Ensemble Control of Loads in Distribution Grids with Network Constraints". In: (2017). URL: http://arxiv.org/abs/1710.09924.

[13] K. Clement-Nyns, E. Haesen, and J. Driesen. "The impact of charging plug-in hybrid electric vehicles on a residential distribution grid". In: *IEEE Transactions on Power Systems* 25.1 (2010), pp. 371–380.

[14] A. Cofffman, N. Cammardella, P. Barooah, and S. Meyn. "Aggregate capacity of TCLs with cycling constraints". In: *ArXiv.org E-print Archive* (2019). URL: https://arxiv.org/abs/1909.11497.

[15] M. Cohen and D. Callaway. "Effects of distributed PV generation on California's distribution system, part 1: engineering simulations". In: *Solar Energy* 128 (2016), pp. 126–138.

[16] S. J. Crocker and J. L. Mathieu. "Adaptive state estimation and control of thermostatic loads for real-time energy balancing". In: *American Control Conference (ACC)*. Boston, MA, July 2016, pp. 3557–3563.

[17] E. Dall'Anese, S. S. Guggilam, A. Simonetto, Y. C. Chen, and S. V. Dhople. "Optimal regulation of virtual power plants". In: *IEEE Transactions on Power Systems* 33.2 (2018), pp. 1868–1881.

[18] A. P. Dobos. *PVWatts version 5 manual*. Tech. rep. National Renewable Energy Laboratory, 2014.

[19] P. Du and N. Lu. "Appliance commitment for household load scheduling". In: *IEEE Transactions on Smart Grid* 2.2 (2011), pp. 411–419.

[20] Ergon Energy/Energex. *Technical specification for polemounting distribution transformers*. Datasheet. Ergon Energy, 2012. URL: https://www.ergon.com.au/about-us/tenders/technical-specifications.

[21] Federal Energy Regulatory Commission. *Notice inviting post-technical conference comments*. Tech. rep. Docket No. RM18-9-000. 2018. URL: https://www.ferc.gov/CalendarFiles/20180427135034-notice-for-comments.pdf.

[22] L. P. Fernandez, T. G. S. Román, R. Cossent, C. M. Domingo, and P. Frias. "Assessment of the impact of plug-in electric vehicles on distribution networks". In: *IEEE Transactions on Power Systems* 26.1 (2011), pp. 206–213.

[23] R. Fu, D. Feldman, M. Woodhouse, and K. Ardani. *U.S. solar photovoltaic system cost benchmark: Q1 2017*. Tech. rep. NREL/TP-6A20-68925. National Renewable Energy Laboratory, 2017.

[24] J. Fuller. *GridLAB-D discussion: prototypical feeder loading assumptions*. 2016. URL: https://sourceforge.net/p/gridlab-d/discussion/842562//thread/43529e21/?limit=25#4305.

[25] H. Gerard, E. I. Rivero Puente, and D. Six. "Coordination between transmission and distribution system operators in the electricity sector: a conceptual framework". In: *Utilities Policy* 50 (2018), pp. 40–48.

[26] T. Gonen. *Electric Power Distribution Engineering*. Boca Raton: CRC Press, 2015.

[27] D. Halamay, T. Brekken, A. Simmons, and S. McArthur. "Reserve requirement impacts of large-scale integration of wind, solar, and ocean wave power generation". In: *IEEE Transactions on Sustainable Energy* 2.3 (2011), pp. 321–328.

[28] H. Hao, B. Sanandaji, K. Poolla, and T. Vincent. "Aggregate flexibility of thermostatically controlled loads". In: *IEEE Transactions on Power Systems* 30.1 (2015), pp. 189–198.

[29] Harrison, Davila, Bartlett, and Hahn. *Adopt an Ordinance adding a new Chapter 12.80 to the Berkeley Municipal Code Prohibiting Natural Gas Infrastructure in New Buildings*. Tech. rep. City Council of Berkeley, July 2019. URL: https://www.pjm.com/-/media/committees-groups/committees/mc/20190425-somr/20190425-2018-imm-state-of-the-market-report-presentation.ashx.

[30] A. Hassan, R. Mieth, M. Chertkov, D. Deka, and Y. Dvorkin. "Optimal load ensemble control in chance-constrained optimal power flow". In: *IEEE Transactions on Smart Grid* 10.5 (2019), pp. 5186–5195.

[31] A. D. Hilshey, P. D. Hines, P. Rezaei, and J. R. Dowds. "Estimating the impact of electric vehicle smart charging on distribution transformer aging". In: *IEEE Transactions on Smart Grid* 4.2 (2013), pp. 905–913.

[32] M. Humayun, M. Z. Degefa, A. Safdarian, and M. Lehtonen. "Utilization improvement of transformers using demand response". In: *IEEE Transactions on Power Delivery* 30.1 (2015), pp. 202–210.

[33] IEEE. *IEEE Guide for Loading Mineral-Oil-Immersed Transformers*. Tech. rep. IEEE Std C57.91-1995(R2004). Dec. 2004.

[34] S. Ihara and F. Schweppe. "Physically based modeling of cold load pickup". In: *IEEE Transactions on Power Apparatus and Systems* 9.PAS-100 (1981), pp. 4142–4150.

[35] E. E. Institute. *Electric Vehcile Sales: Fact and Figures*. Tech. rep. Apr. 2018. URL: https://www.eei.org/issuesandpolicy/electrictransportation/Documents/FINAL_EV_Sales_Update_April2019.pdf.

[36] J. Jung, A. Onen, K. Russell, and R. P. Broadwater. "Local steady-state and quasi steady-state impact studies of high photovoltaic generation penetration in power distribution circuits". In: *Renewable and Sustainable Energy Reviews* 43 (2015), pp. 569–583.

[37] E. C. Kara, M. Berges, and G. Hug. "Impact of disturbances on modeling of thermostatically controlled loads for demand response". In: *IEEE Transactions on Smart Grid* 6.5 (2015), pp. 2560–2568.

[38] E. C. Kern and E. M. Gulachenski. "Cloud effects on distributed photovoltaic generation: slow transients at the Gardner, Massachusetts photovoltaic experiment". In: *IEEE Transactions on Energy Conversion* 4.2 (1989), pp. 184–190.

[39] W. Kersting. *Distribution System Modeling and Analysis, Third Edition*. Boca Raton, Florida: CRC Press, 2012.

[40] S. Kundu, N. Sinitsyn, S. Backhaus, and I. Hiskens. "Modeling and control of thermostatically controlled loads". In: *Power Systems Computation Conference*. Stockholm, Sweden, Aug. 2011.

[41] G. S. Ledva, L. Balzano, and J. L. Mathieu. "Real-time energy disaggregation of a distribution feeder's demand using online learning". In: *IEEE Transactions on Power Systems* 33.5 (2018), pp. 4730–4740.

[42] G. S. Ledva, E. Vrettos, S. Mastellone, G. Andersson, and J. L. Mathieu. "Managing communication delays and model error in demand response for frequency regulation". In: *IEEE Transactions on Power Systems* 33.2 (2018), pp. 1299–1308.

[43] R. Leou, C. Su, and C. Lu. "Stochastic analyses of electric vehicle charging impacts on distribution network". In: *IEEE Transactions on Power Systems* 29.3 (2014), pp. 1055–1063.

[44] M. Liu and Y. Shi. "Model predictive control of aggregated heterogeneous second-order thermostatically controlled loads for ancillary services". In: *IEEE Transactions on Power Systems* 31.3 (2016), pp. 1963–1971.

[45] N. Lu, D. Chassin, and S. Widergren. "Modeling uncertainties in aggregated thermostatically controlled loads using a state queueing model". In: *IEEE Transactions on Power Systems* 20.2 (2005), pp. 725–733.

[46] J. MacDonald, P. Cappers, D. S. Callaway, and S. Kiliccote. "Demand response providing ancillary services: a comparison of opportunities and challenges in the US wholesale markets". In: *Grid Interop* (2012).

[47] T. Mai, P. Jadun, J. Logan, C. McMillan, M. Muratori, D. Steinberg, L. Vimmerstedt, R. Jones, B. Haley, and B. Nelson. *Electrification Futures Study: Scenarios of Electric Technology Adoption and Power Consumption for the United States*. Tech. rep. National Renewable Energy Lab, 2018. URL: `https://www.nrel.gov/docs/fy18osti/71500.pdf`.

[48] Y. Makarov, C. Loutan, J. Ma, and P. De Mello. "Operational impacts of wind generation on California power systems". In: *IEEE Transactions on Power Systems* 24.2 (2009), pp. 1039–1050.

[49] J. L. Mathieu, M. Kamgarpour, J. Lygeros, and D. S. Callaway. "Energy arbitrage with thermostatically controlled loads". In: *European Control Conference (ECC)*. Zurich, Switzerland, July 2013, pp. 2519–2526.

[50] J. Mathieu, M. Dyson, and D. Callaway. "Resource and revenue potential of California residential load participation in ancillary services". In: *Energy Policy* 80 (2015), pp. 76–87.

[51] J. Mathieu, M. Kamgarpour, J. Lygeros, G. Andersson, and D. Callaway. "Arbitraging intraday wholesale energy market prices with aggregations of thermostatic loads". In: *IEEE Transactions on Power Systems* 30.2 (2015), pp. 763–772.

[52] J. Mathieu, S. Koch, and D. Callaway. "State estimation and control of electric loads to manage real-time energy imbalance". In: *IEEE Transactions on Power Systems* 28.1 (2013), pp. 430–440.

[53] MATLAB. *Kalman Filtering*. 2018. URL: `https://www.mathworks.com/help/control/ug/kalman-filtering.html`.

[54] C. Maxey and A. Andreas. *Rotating Shadowband Radiometer (RSR)*. Tech. rep. NREL DA-5500-56512. Oak Ridge National Laboratory, 2007.

[55] B. McCarl and T. Spreen. *Applied Mathematical Programming Using Algebraic Systems*. 2003. URL: `http://agecon2.tamu.edu/people/faculty/mccarl-bruce/books.htm`.

[56] D. Molzahn and L. A. Roald. "Grid-aware versus grid-agnostic distribution system control: a method for certifying engineering constraint satisfaction". In: *Proceedings of the 52nd Hawaii International Conference on System Sciences*. Jan. 2019.

[57] More than Smart. *Coordination of Transmission and Distribution Operations in a High Distributed Energy Resource Electric Grid*. Tech. rep. June 2017. URL: `https://www.caiso.com/Documents/MoreThanSmartReport-CoordinatingTransmission_DistributionGridOperations.pdf`.

[58] R. E. Mortensen and K. P. Haggerty. "A stochastic computer model for heating and cooling loads". In: *IEEE Transactions on Power Systems* 3.3 (1988), pp. 1213–1219.

[59] National Electrical Manufacturers Association. *American National Standard for Electric Power Systems and Equipment – Voltage Ratings (60 Hertz)*. Tech. rep. ANSI C84.1-2006. Dec. 2006.

[60] M. S. Nazir and I. A. Hiskens. "Noise and parameter heterogeneity in aggregate models of thermostatically controlled loads". In: *International Federation of Automatic Control Conference (IFAC) World Congress*. Toulouse, France, July 2017.

[61] P. Nilsson and N. Ozay. "Maximizing the time of invariance for large collections of switched systems". In: *Proceedings of the IEEE Conference on Decision and Control (CDC)*. Melbourne, Australia, Dec. 2017.

[62] P. Nilsson and N. Ozay. "On a class of maximal invariance inducing control strategies for large collections of switched systems". In: *Proceedings of the International Conference on Hybrid Systems: Computation and Control (HSCC)*. Pittsburgh, Pennsylvania, USA, Apr. 2017.

[63] Pacific Northwest National Laboratory. *GridLAB-D Software*. URL: http://www.gridlabd.org/.

[64] Pacific Northwest National Laboratory. *Population Script*. 2015. URL: https://sourceforge.net/p/gridlab-d/code/HEAD/tree/Taxonomy_Feeders/PopulationScript/.

[65] Pacific Northwest National Laboratory. *Residential module user's guide*. URL: http://gridlab-d.shoutwiki.com/wiki/Residential_module_user%27s_guide.

[66] Pacific Power Association. *Quantification of Energy Efficiency in the Utilities of the U.S. Affiliate States (Excluding US Virgin Islands)*. Tech. rep. Dec. 2010. URL: http://prdrse4all.spc.int/data/kema-study-quantification-energy-efficiency-utilities.

[67] B. H. Pinckaers. *Minimum off-time circuit*. US Patent 3,619,668. Honeywell, Nov. 1971.

[68] PJM Interconnection LLC. *PJM Ancillary Services*. URL: http://www.pjm.com/markets-and-operations/ancillary-services.aspx.

[69] S. C. Ross, G. Vuylsteke, and J. L. Mathieu. "Effects of load control for real-time energy balancing on distribution network constraints". In: *IEEE Power & Energy Society PowerTech Conference*. Manchester, U.K., June 2017.

[70] S. C. Ross, P. Nilsson, N. Ozay, and J. L. Mathieu. "Managing voltage excursions on the distribution network by limiting the aggregate variability of thermostatic loads". In: *American Control Conference*. Philadelphia, PA, July 2019.

[71] S. C. Ross, N. Ozay, and J. L. Mathieu. "Coordination between an aggregator and distribution operator to achieve network-aware load control". In: *IEEE Power & Energy Society PowerTech Conference*. Milan, Italy, June 2019.

[72] S. C. Ross, G. Vuylsteke, and J. L. Mathieu. *GridLABD-load-control*. 2018. URL: https://github.com/crockerross/GridLABD-load-control.

[73] S. C. Ross, G. Vuylsteke, and J. L. Mathieu. "Effects of load-based frequency regulation on distribution network operation". In: *IEEE Transactions on Power Systems* 34.2 (2019), pp. 1569–1578.

[74] B. Sanandaji, H. Hao, and K. Poolla. "Fast regulation service provision via aggregation of thermostatically controlled loads". In: *Proceeedings of HICSS*. 2014, pp. 2388–2397.

[75] K. Schneider, Y. Chen, D. Chassin, R. Pratt, D. Engel, and S. Thompson. *Distribution Taxonomy Final Report*. Tech. rep. Pacific Northwest National Laboratory, Nov. 2008.

[76] R. Sherick. *Advanced Distribution Analytics Enabling High Penetration Solar PV*. Tech. rep. Southern California Edison, Nov. 2016. URL: http://www.calsolarresearch.ca.gov/final-project-reports.

[77] D. Simon. "Kalman filtering with state constraints: a survey of linear and nonlinear algorithms". In: *IET Control Theory & Applications* 4.8 (2010), pp. 1303–1318.

[78] N. A. Sinitsyn, S. Kundu, and S. Backhaus. "Safe protocols for generating power pulses with heterogeneous populations of thermostatically controlled loads". In: *Energy Conversion and Management* 67 (2013), pp. 297–308.

[79] R. Sonderegger. "Dynamic models of house heating based on equivalent thermal parameters". Ph.D. dissertation. Princeton Univ., N.J., 1978.

[80] State Zero-Emission Vehicle Programs. *Multi-State ZEV Task Force*. URL: https://www.zevstates.us/.

[81] J. Taylor, A. Maitra, M. Alexander, D. Brooks, and M. Duvall. "Evaluations of plug-in electric vehicle distribution system impacts". In: *IEEE PES General Meeting*. July 2010.

[82] L. C. Totu, R. Wisniewski, and J. Leth. "Demand response of a TCL population using switch-rate actuation". In: *IEEE Transactions on Control Systems Technology* 25.5 (2017), pp. 1537–1551.

[83] V. Trovato, S. H. Tindemans, and G. Strbac. "Leaky storage model for optimal multi-service allocation of thermostatic loads". In: *IET Generation, Transmission Distribution* 10.3 (2016), pp. 585–593.

[84] U.S. Energy Information Administration. *September 2019 Monthly Energy Review*. Tech. rep. DOE/EIA-0035(2019/9). Sept. 2019. URL: https://www.eia.gov/totalenergy/data/monthly/index.php.

[85] U.S. Environmental Protection Agency. *Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990-2017*. Tech. rep. EPA 430-R-19-001. Apr. 2019. URL: https://www.epa.gov/ghgemissions/inventory-us-greenhouse-gas-emissions-and-sinks-1990-2017.

[86] U.S. Federal Energy Regulatory Commission. *Notice of Proposed Rulemaking: Electric Storage Participation in Markets Operated by Regional Transmission Organizations and Independent System Operators*. Tech. rep. Nov. 2016.

[87] U.S. Federal Energy Regulatory Commission. *Order No. 719: Wholesale Competition in Regions with Organized Electric Markets*. Tech. rep. Oct. 2008.

[88] U.S. Federal Energy Regulatory Commission. *Order No. 745: Demand Response Compensation in Organized Wholesale Energy Markets*. Tech. rep. Mar. 2011.

[89] U.S. Federal Energy Regulatory Commission. *Order No. 755: Frequency Regulation Compensation in the Organized Wholesale Power Markets*. Tech. rep. Oct. 2011.

[90] U.S. Federal Energy Regulatory Commission. *Order No. 841: Electric Storage Participation in Markets Operated by Regional Transmission Organizations and Independent System Operators*. Tech. rep. Feb. 2018.

[91] VanTran Industries. *VanTran Catalog*. July 2012. URL: http://vantran.com/information/.

[92] T. Voice. “Stochastic thermal load management”. In: *IEEE Transactions on Automatic Control* 63.4 (2017), pp. 931–946.

[93] E. Vrettos and G. Andersson. “Combined load frequency control and active distribution network management with thermostatically controlled loads”. In: *IEEE International Conference on Smart Grid Communications*. Oct. 2013.

[94] E. Vrettos, J. Mathieu, and G. Andersson. “Control of thermostatic loads using moving horizon estimation of individual load states”. In: *Proceedings of the Power Systems Computation Conference (PSCC)*. Wroclaw, Poland, Aug. 2014.

[95] K. Wada and A. Yokoyama. “Load frequency control using distributed batteries on the demand side with communication characteristics”. In: *IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe)*. Berlin, Germany, Oct. 2012.

[96] W. Zhang, J. Lian, C. Chang, and K. Kalsi. “Aggregated modeling and control of air conditioning loads for demand response”. In: *IEEE Transactions on Power Systems* 28.4 (Nov. 2013), pp. 4655–4664.

[97] C. Zhong and J. L. Mathieu. “Relation between overheating of distribution transformers and switching frequency of electric loads used for demand response”. In: *North American Power Symposium*. Oct. 2015.

[98] Z. Ziadi, S. Taira, M. Oshiro, and T. Funabashi. “Optimal power scheduling for smart grids considering controllable loads and high penetration of photovoltaic generation”. In: *IEEE Transactions on Smart Grid* 5.5 (June 2014), pp. 2350–2359.

[99] C. Ziras, S. You, H. W. Bindner, and E. Vrettos. “A new method for handling lockout constraints on controlled TCL aggregations”. In: *2018 Power Systems Computation Conference (PSCC)*. June 2018.