Packetized energy management: asynchronous and anonymous coordination of thermostatically controlled loads

Mads Almassalkhi, *Member, IEEE* Jeff Frolik, *Senior Member, IEEE* Paul Hines, *Senior Member, IEEE*

***Abstract*— Because of their internal energy storage, electrically powered, distributed thermostatically controlled loads (TCLs) have the potential to be dynamically managed to match their aggregate load to the available supply. However, in order to facilitate consumer acceptance of this type of load management, TCLs need to be managed in a way that avoids degrading perceived quality of service (QoS), autonomy, and privacy. This paper presents a real-time, adaptable approach to managing TCLs that both meets the requirements of the grid and does not require explicit knowledge of a specific TCL's state. The method leverages a packetized, probabilistic approach to energy delivery that draws inspiration from digital communications. We demonstrate the packetized approach using a case-study of 1000 simulated water heaters and show that the method can closely track a time-varying reference signal without noticeably degrading the QoS. In addition, we illustrate how placing a simple ramp-rate limit on the aggregate response overcomes synchronization effects that arise under prolonged peak curtailment scenarios.**

I. Introduction

Fast-ramping generators have provided reliable operating reserves for decades. However, power systems with high penetrations of renewable energy challenge this operating paradigm. Indeed, there is a growing consensus that balancing supply and demand in power systems with large amounts of variable renewable energy will require an active role for flexible distributed energy resources (DERs) in addition to balancing services from conventional power plants [1]. While the core concepts underlying modern demand-side management (DSM) have existed for decades [2], [3], the technology for coordinating the activities of DERs is maturing rapidly.

To overcome common privacy, convergence, and quality-of-service (QoS) concerns and enable large-scale penetration of renewable energy, the work herein proposes a novel bottom-up load coordination framework, *packetized energy management* (PEM), that regulates the aggregate power consumption of distributed energy resources (DERs) (e.g., electric water storage heaters, the focus of this present effort) by building on methods used to manage data packets in communication networks. Specifically, the delivery of energy to a load is accomplished using multiple "energy packets" or "packetized energy", just as digital communication networks break data into data packets. In fact, by leveraging internet-like protocols for distributed control, PEM inherits certain "fairness" properties with regard to providing statistically identical grid access to each load.

With the proposed PEM architecture, the grid operator or aggregator only requires a two dimensional measurement from the collection of loads: aggregate power consumption and an aggregate request process. This represents a significant advantage over aggregate model-estimator-controller state-space approaches in [4], which requires an entire histogram of states from the collection of loads to update a state bin transition model. In [4], this is addressed through an observer design to estimate the histogram based on aggregate power consumption; however, in some cases, the model may not be observable [5]. Recent work has extended [4] to include higher order dynamic models and end-user and compressor delay constraints [6] and stochastic dynamical performance bounds [7]. Similar to the mean-field approaches in [5], [8], [9], this work guarantees QoS through opt-out mechanisms and injects randomization based on local state variables, which limits the duration of synchronization effects and promotes equitable access to the grid. However, in contrast to those prior works, PEM does not have the load coordinator broadcast the control signal (in top-down fashion). Instead, PEM is designed to have each load request an energy packet from the coordinator stochastically (in bottom-up fashion) based on the load's local state variables. The coordinator then responds in real-time to each packet request based on grid or market conditions.

The most closely related work on energy packets is found in references [10], [11], where an omniscient centralized *packetized direct load controller* (PDLC) is developed for TCLs. The average controller performance and consumer QoS is analytically investigated and queuing theory is employed by the authors to quantify the centralized controller's performance. In [12], a distributed (binary information) version of PDLC is proposed that requires only (binary) packet request information from the loads. Unlike the proposed PEM scheme, the distributed PDLC assumes that the exact number of participating packetized loads at any given time is known, the allocation of packet requests from the queue is synchronized, and the queue stores packet requests if the packets cannot be allocated, which creates delays in service. Instead, the work herein extends the authors' previous packetized energy results for managing the near-optimal charging of PEVs [13], [14] and presents the first application of this anonymous, asynchronous, and randomized bottom-up control scheme for TCLs.

This work was supported by the U.S. Department of Energy's Advanced Research Projects Agency - Energy grant DE-AR0000694.

Authors are with the Department of Electrical and Biomedical Engineering, University of Vermont, 121 Colchester Avenue, Burlington, VT 05405, U.S.A. {malmassa, jfrolik, phines}@uvm.edu

The specific contributions of this paper are two-fold:

1) Presentation of novel distributed coordination of TCLs under imperfect information and consumer QoS constraints.
2) Performance of the proposed PEM paradigm is investigated with an electric water heater case study under reference tracking and prolonged peak curtailment scenarios.

The remainder of this paper is structured as follows. Section II describes the adaptation of PEM to asynchronous and anonymous coordination of TCLs. In Section III, we present a case-study with simulation results for 1000 packetized TCLs under different PEM schemes with a baseline control case to illustrate QoS constraint management. Section IV, concludes the paper and provides future research directions.

II. Packetized Energy Management

Figure 1 illustrates the cyber-physical interactions needed to realize PEM in a power system. We will separately describe the functions of the grid operator (e.g., a utility), the coordinator (e.g., DER management system or a virtual power plant - VPP), and the packetized energy controller, which is connected to the VPP via a Wi-Fi-enabled TCL and can interact directly with its load to beget the "packetized" behavior. Owing to the proposed bottom-up approach, we will first introduce the concept of a packetized load.

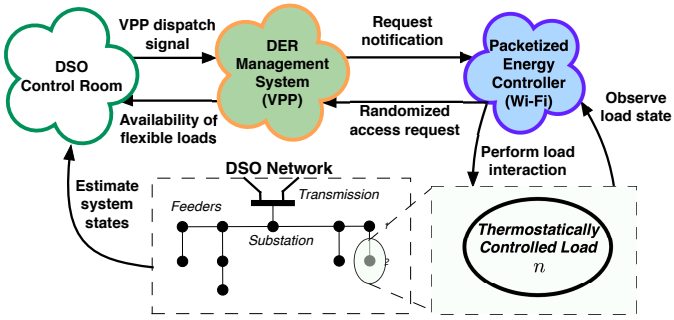

Fig. 1: Cyber-physical infrastructure needed to realize PEM.

A. Packetized Load

As noted, PEM has previously been proposed for coordinated charging of electric vehicles. In this earlier work [13], [15], [14], PEVs asynchronously request the authority to charge with a specific probability according to their logic state in a probabilistic automaton. For example, for a three-state finite-state machine (e.g., Fig. 2(right)), the probability to request access to the grid from state i is P_i and $P_1 > P_2 > P_3$. If there is capacity in the grid, the PEV is granted authority to charge, but only for a fixed duration of time (e.g., 15 minutes), referred to as the control epoch and a state transition takes place: $P_i \to P_{i-1}$, which reduces the mean time-to-request (MTTR). In contrast, if the PEV is denied authority to charge, the mean time-to-request increases with transition $P_i \to P_{i+1}$. Herein, we adapt the PEM concept for the purpose of managing TCLs by specifically leveraging the TCL's local temperature to drive the randomization of requests.

1) TCL modeling: The TCL load type of interest is hot-water heaters (HWHs). The following presents a first-order single-heating-element thermodynamic model motivated by [16], [17] but modified to consider a uniform thermal mass and hot-water withdrawal by the consumer in liters/min rather than as a fixed energy loss. Thus, consider the dynamics of the temperature of TCL n at time $t, T_n(t)$.

$$c\rho L \frac{dT_n(t)}{dt} = u_n(t) - d_n(t) - w_n(t) \tag{1}$$

where $c = 4.186$ [kJ/kg-°C] and $\rho = 0.990$ [kg/liters] represent specific heat capacity and density of water close to 50°C[1]. L [liters] represents the total capacity of the HWH. The heat input u, ambient losses d, and uncontrolled hot-water consumption w are defined by the following physical relations:

$$u_n(t) = \frac{P_n^{\text{rate}}}{\eta_n} z_n(t) \tag{2}$$

$$d_n(t) = c\rho L \frac{(T_n(t) - T_{\text{amb}}(t))}{\tau_n} \tag{3}$$

$$w_n(t) = c\rho \frac{v_n(t)}{60} (T_n(t) - T_{\text{inlet}}(t)) \tag{4}$$

where $P_n^{\text{rate}}, \eta_n$, and z_n are the heating element power transfer rate [kW], heat transfer efficiency, and the binary ON/OFF state of HWH n ($\Rightarrow z_n = 1/0$), respectively. The terms $T_{\text{amb}}, \tau_n, T_{\text{inlet}}$ and v_n are the ambient temperature [°C], time-constant due to ambient insulation losses [s], inlet temperature [°C], and hot water withdrawal rate by consumer n in [liters/min], respectively. Next, to derive the discrete-time model, suppose that the control input z_n and disturbances d_n, w_n are step-wise with step width Δt [s], such that $z_n(t) := z_n[k]$ for $t \in [k\Delta t, (k+1)\Delta t]$. Similarly, define $d_n[k]$ and $w_n[k]$, in which case, (1) becomes:

$$T_n[k+1] = T_n[k] + \frac{\Delta t}{c\rho L} (u_n[k] - d_n[k] - w_n[k]) \tag{5}$$

where $T_n[k] := T_n(k\Delta t)$. By substituting for u_n, d_n, w_n, it is straightforward to see that for numerical stability, the sampling period is required to satisfy:

$$\Delta t < \frac{120L}{\frac{60L}{\tau_n} + v_n[k]}, \tag{6}$$

which through simple analysis implies that for conservative $v_n[k] < 60, L > 100$, and $\tau_n > 24$ hours, sampling faster than every 200 seconds satisfies numerical stability conditions. We will use $\Delta t = 10$s in all simulations presented in Section III.

2) Traditional control of TCLs: The vast majority of existing traditional TCLs operate in a binary (ON/OFF) manner and are already controlled by simple state machines – thermostats that change state based on temperature thresholds. Locally, a TCL is controlled to maintain a desired temperature set-point, T_n^{set}, within a temperature dead-band,

[1]Physically, c and ρ vary with temperature, but this relationship is ignored in this paper as it does not affect the results or conclusion of PEM's local decision making.

$T_n^{\text{set}} \pm T_n^{\text{set,DB}}/2$. This yields the standard TCL hysteretic temperature response according to local discrete-time control logic:

$$z_n[k] = \begin{cases} 1, & T_n[k] \le T_n^{\text{set}} - T_n^{\text{set,DB}}/2 \\ 0, & T_n[k] \ge T_n^{\text{set}} + T_n^{\text{set,DB}}/2 \\ z_n[k-1], & \text{otherwise.} \end{cases} \quad (7)$$

We will refer to the aggregate response under the above fully decentralized control logic as the "no-control" case. The proposed PEM scheme requires only the replacement of the existing state machine with a more sophisticated one (i.e., the equivalent of a firmware upgrade) that interacts with a coordinator/aggregator.

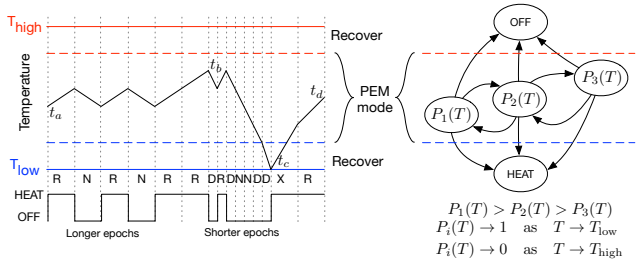

Fig. 2: Water heater managed by PEM. The left figure shows a sequence of events. At time t_a, when grid resources are unconstrainted, loads stochastically request (R) or do not request (N) energy. At t_b, the system approaches a period of constrained supply, in which the system aggregator mostly denies requests (D) and reduces the epoch length. As a result, the automaton transitions to a lower probability state (e.g., $P_1 \to P_2$). At t_c, the temperature hits the QoS bound and the load exits (X) from PEM and rapidly seeks to recover temperature to within the QoS bounds, which occurs at t_d. The right figure shows the state machine that changes its request probabilities ($P_i(T)$) and its epoch lengths, based on responses the local temperature state. Also embedded in the automaton is the epoch lengths between state transitions/making requests.

3) Adaptation of PEM for TCLs: Figure 2(right) illustrates a TCL automaton under PEM. When the local temperature of the TCL, T, is between its upper and lower temperature limits for PEM operation, the TCL's time-to-request is driven by an exponential distribution whose mean is inversely proportion to T relative to the upper limit. That is, TCLs with temperatures very close to the lower threshold will make requests with near certainty (i.e., $P_i(T \to T_{\text{low}}) \approx 1$) and those near the upper limit in temperature will make requests with low probability (i.e., $P_i(T \to T_{\text{high}}) \approx 0$). Upon transmitting a request and, if there is capacity in the grid, the TCL will be given the authority to turn ON for a fixed control epoch length δ_t (i.e., $z_n(t) = 1$ for $t \in (t_0, t_0 + \delta_t)$), and a state transition occurs: $P_i(T) \to P_{i-1}(T)$. If the request is denied, the TCL finite state machine transitions to a state with lower MTTR, $P_i(T) \to P_{i+1}(T)$, but will immediately resume requesting with temperature-dependent probability. If access is denied repeatedly, T reaches T_{low}, which causes the TCL to exit (i.e., opt-out of) the PEM scheme to guarantee that temperature bounds are satisfied. An illustrative ON/OFF cycle of a packetized water heater is illustrated in Fig. 2(left).

In addition to the TCL receiving an "Yes/No" response to a request, the TCL may also receive an updated (global) control epoch length, δ_t, thus enabling tighter tracking in the aggregate, which is helpful during ramping events. Clearly, while the TCL is ON, it does not make requests. Furthermore, we require $\delta_t \ge \Delta t$.

Remark II.1 *Since all TCLs operate in this manner, the DER aggregator granting ("Yes") or denying ("No") the authority to turn on does not require any knowledge/tracking of a particular TCL. Furthermore, the aggregator does not even track which TCL is making a particular request. As each TCL runs the same automaton logic and its ability to turn on depends only on the real-time system capacity, any TCL making a request at the same point in time will be treated the same by the aggregator. As such, the PEM approach inherently maintains privacy while still being fair to its customers.*

4) The stochastic request rate with PEM: In the discrete-time implementation of PEM, the probability that TCL n with local temperature $T_n[k]$ in automaton state i requests access to the grid during time-step k (over interval Δt) is defined by the cumulative exponential distribution function:

$$P_i(T_n[k]) := 1 - e^{-\mu(T_n[k],i)\Delta t}$$

where rate parameter $\mu(T_n[k], i) > 0$ is dependent on the local temperature and the probabilistic automaton's logic state i. This dependence is established by considering the following boundary conditions:

- $P_i\left(\text{TCL } n \text{ requests access at } k \,|\, T_n[k] \le T_i^{\min}\right) = 1$
- $P_i\left(\text{TCL } n \text{ requests access at } k \,|\, T_n[k] \ge T_i^{\max}\right) = 0,$

which give rise to the following natural design of a PEM rate parameter:

$$\mu(T_n[k], i) = \begin{cases} 0, & \text{if } T_n[k] > T_n^{\max} \\ \frac{T_n^{\max} - T_n[k]}{T_n[k] - T_n^{\min}} M_i, & \text{if } T_n[k] \in (T_n^{\min}, T_n^{\max}] \\ \infty, & \text{if } T_n[k] \le T_n^{\min} \end{cases} \quad (8)$$

where $M_i > 0$ [1/sec] is a design parameter that depends on the TCL's automaton state i and defines the MTTR. For example, if one desires a MTTR of 5 minutes, $M_i = \frac{1}{600}$Hz.

Remark II.2 *If we consider a symmetric dead-band, i.e., $T_n^{\min}, T_n^{\max} := T_n^{set} \mp T_n^{set,DB}/2$, then the mean time-to-request (MTTR) for TCL n with $T_n[k] = T_n^{set}$ is exactly described by $1/M_i$ (in seconds), which represents a useful parameter for design of the finite-state machine. Figure 3 illustrates a TCL's stochastic request rate for a three-state automaton, where $P_1(T_n[k]) > P_2(T_n[k]) > P_3(T_n[k])$ are defined by the bold blue, red, and green lines, respectively. In the case of asymmetric dead-band, M_i can still be an effective design parameter by generalizing the middle condition of Eq.* (8)*:*

$$\mu(T_n[k], i) = \left(\frac{T_n^{\max} - T_n[k]}{T_n[k] - T_n^{\min}}\right)\left(\frac{T_n^{set} - T_n^{\min}[k]}{T_n^{\max}[k] - T_n^{set}}\right) M_i.$$

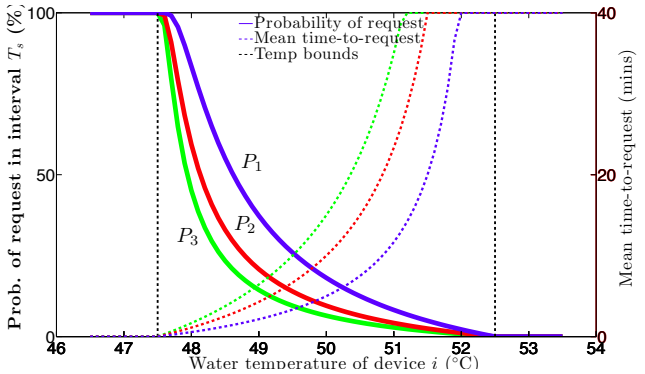

Fig. 3: Illustrating the effect of local temperature on the stochastic access request rates (bold) and MTTR (dashed) of a three-state TCL under PEM. For graphical purposes only, the mean time-to-request have been truncated to 40 minutes.

5) Quality of service under PEM: With the stochastic nature of TCLs under PEM, it is entirely possible that a disturbance (e.g., a large hot water withdrawal rate) can drive $T_n[k]$ below $T_n^{\min}$. Therefore, to maximize QoS to the consumer (i.e., avoid cold showers), a TCL under PEM can temporarily exit (i.e., opt-out of) PEM and operate under traditional TCL control (e.g., turn ON and stay ON). This is illustrated in Fig. 2(left) at event t_c and with automaton states HEAT and OFF in Fig. 2(right). That is, once a TCL under PEM exceeds temperature bounds, the traditional control logic is employed temporarily to bring the local temperature within PEM "recovery bounds" $T_n^{\text{set}} \pm T_n^{\text{set,PEM}}/2$ with $T_n^{\text{set,PEM}} < T_n^{\text{set,DB}}$ when PEM control logic is reinstated (i.e., TCL opts back into PEM). The recovery bounds are helpful to avoid excessive exit/re-entry cycling at the min/max bounds. While cold showers are undesirable, overheating HWHs is dangerous to consumer and damaging to TCL. As such, a TCL under PEM will *never* actuate if $T_n[k] > T_n^{\max}$.

Remark II.3 *Clearly, if TCLs exit PEM en-masse, the available flexibility can be greatly reduced and, therefore, will impact ability of a coordinator to track a given reference balancing signal. Hence, an on-going effort is focused on optimal design of the finite-state machine to reduce the need for opting out and to improve ramping capabilities.*

B. Coordinating TCLs under PEM: Virtual Power Plant

As shown in Fig. 1, a packetized energy controller enables bidirectional Wi-Fi communication between a load and the virtual power plant (VPP). The VPP receives balancing dispatch signals akin to Automatic Generation Control (AGC) from a grid operator and coordinates flexible energy resources to track the balancing command[2]. Within the proposed PEM scheme, the VPP tracks the balancing signal by responding to individual downstream asynchronous and stochastic load access requests (i.e., pings) with "Yes" or "No" notifications based on real-time output error between actual aggregate output, $y(t)$, and the VPP reference signal, $r(t)$: $e(t) := r(t) - y(t)$. This is illustrated in Fig. 4. The VPP is similar to a relay controller in the sense that it accepts a request ("Yes") if $e(t) > 0$, otherwise, "No." However, unlike standard relay control of a single plant, the VPP responds to asynchronous, stochastic requests from N plants, which overcomes common drawbacks associated with relay control (e.g., oscillations) and permits accurate tracking.

Input: Balancing reference signal, asynchronous request;
Output: Yes/No access notification to individual load.

[2]While the VPP needs to estimate and predict the aggregate flexibility from available loads, this paper will focus on the tracking control problem as the estimation problem represents ongoing research

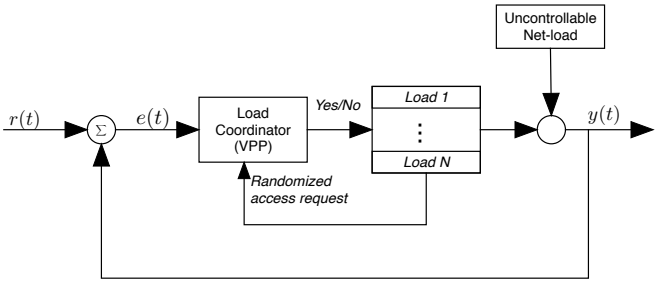

Fig. 4: The closed-loop feedback system for PEM with the reference $r(t)$ provided by the Grid Operator and the aggregate packetized TCL output response $y(t)$ measured by VPP.

C. Providing grid-level service with PEM

The transmission (e.g., ISO New England) or distribution utility system operator (e.g., the DSO Control Room in Fig. 1) is able to measure or estimate the states of the grid, such as voltage, frequency, and power flows. Under scenarios with high penetration of renewable energy, the grid operator will find it ever more difficult to balance demand and supply and, therefore, seeks to leverage the flexible packetized DERs sitting in customer homes and industrial/commercial facilities. This is achieved by signaling individual balancing requests to VPPs across the grid in near real-time akin to Automatic Generation Control (AGC) signals, which are transmitted every 4-5 seconds today.

Input: Grid states and net-load forecasts;
Output: Balancing request signal;

In summary, by managing the anonymous, fair, and asynchronous pings of packetized loads via a VPP that receives grid or market-based balancing signals from the grid operator, PEM represents a bottom-up distributed control scheme that has been adapted for TCLs in this paper. The next section implements and provides numerical analysis of a VPP tracking various balancing signals.

III. Case Study: Tracking with Packetized TCLs

Next, we simulate the response of $N = 1000$ packetized electric hot water heaters under PEM over a period of $T_{sim} = 6$ hours in accordance with (5). PEM employs a 3-state automaton for each HWH with opt-out capability as described in Fig. 2 and with mean time-to-request (MTTR) defined by (8) and set to 300 seconds for all logic states. Table I summarizes the simulation, where the bracket notation $[a, b]$ denotes randomly distributed values

TABLE I: Simulation parameters

Parameter	Value	Unit
Simulation period, T_{sim}	6	hours
Sampling period, Δt	10	s
Control epoch length, δ_t	300,1800	s
Specific heat capacity (Water), c	4.186	kJ / (kg-°C)
Water density, ρ	0.99	kg / liter
Ambient insulation losses, τ_n	150	hr
Heater Capacity, L	[250,300]	liters
Set-point temperature, T_n^{set}	[52, 58]	°C
Dead-band temperature, $T_n^{\text{set,DB}}$	$0.12T_n^{\text{set}}$	°C
PEM temperature bounds, $T_n^{\text{set,PEM}}$	$0.08T_n^{\text{set}}$	°C
PEM request parameters $\{M_{1,2,3}\}$	$\frac{1}{300}$	-
Input heat transfer rate, P_n^{rate}	[4.5, 5.5]	kW
Heating efficiency, η_n	100	%
Initial temperature, $T_n[0]$	[49,61]	°C
Initial ambient temperature, $T_{\text{amb}}[0]$	[14,18]	°C
Initial remaining control epoch	$[0, \delta_t]$	s
Initial automata state, i	2	-
Average HW events/hr, HW_{avg}^{hr}	1	event/hr
HW event starting times, k_0^{HW}	$[0, 5.83]$	hours
HW event duration, μ_{HW}, σ_{HW}	700, 300	secs
HW event magnitude, $v_n[k]$	[1, 30]	liters/min

over the closed interval $[a, b]$. Note that the TCLs represent a heterogeneous population. The ambient conditions for TCLs are heterogenous, as well, but considered constant over the time of the simulation. To capture hot-water consumption profiles of TCL n, $w_n[k]$ is given by the following simplistic stochastic process:

- Choose the average number of hot-water (HW) events per hour, HW_{avg}^{hr}.
- For each TCL, uniformly distribute the total number of HW events with mean $2T_{sim}HW_{avg}^{hr}$.
- Randomly select hot-water events' starting times from available times, k_0^{HW}
- For each HW event, choose duration Δk^{HW} from normally distributed $\min\{\max\{\mathcal{N}(700, 300)/\Delta t, 1\}, 3600/\Delta t\}$.
- From the duration of each HW event, choose a constant hot-water withdrawal rate $v_n[k]$ $[liters/min]$ based on the exponential distribution with mean $1200/(\Delta t \Delta k^{HW})$, which is inversely proportional to duration. A capacity of 30 liters/min is imposed on $v_n[k]$, which represents a high residential flow-rate [18].

Next, we show how PEM compares with the traditional TCL control (i.e., "no control") of HWHs and how PEM can employ a coordinator to provide balancing in near-real-time with no prediction and only limited and imperfect information. Finally, we will discuss PEM under stressed operating conditions and show how to avoid the synchronization of TCLs under PEM.

A. Load Choreography with PEM

The $N = 1000$ TCLs are subjected to the hot-water events described above and the overarching goal is now to understand how well a packetizing VPP can track a balancing signal from a grid operator by accepting/denying requests from packetized TCLs. The following four control schemes

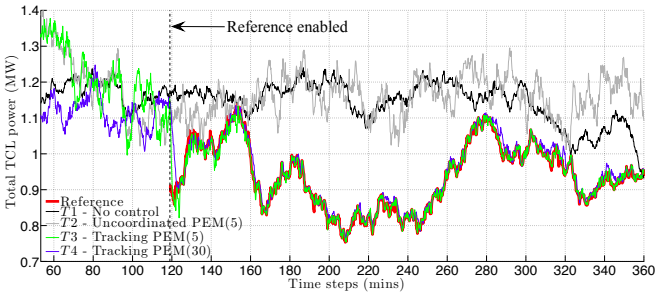

Fig. 5: The evolution of 1000 TCLs under the four different control schemes. Red is the reference, which is enabled after two hours and tracked for four hours with PEM(5) and PEM(30) ($T3$ and $T4$). The black and gray trajectories represent uncoordinated response of TCLs under hysteretic and PEM control schemes ($T1$ and $T2$).

are compared:

$T1$: Baseline "no control" - standard hysteretic TCL control logic is employed as described in Section II.A.2.

$T2$: Uncoordinated PEM(5) - the 3-state probabilistic finite-state machine with control epoch $\delta_t = 5$ minutes described in Table I. However, TCLs are uncoordinated and unresponsive to the reference signals (i.e., every request is accepted by VPP). That is, this case represents a fully decentralized PEM.

$T3$: Reference-tracking PEM(5) - the VPP-coordinated 3-state probabilistic finite-state machine with control epoch $\delta_t = 5$ minutes described in Table I.

$T4$: Reference-tracking PEM(30) - the VPP-coordinated 3-state probabilistic finite-state machine with control epoch $\delta_t = 30$ minutes described in Table I.

1) Comparing Uncoordinated PEM to traditional TCL scheme: In Fig. 5, the uncoordinated schemes (in black and gray) are unaware of the reference signal (red) and, therefore, evolve stochastically according to their internal plant dynamics rather than the reference tracking error. In the case of the uncoordinated PEM(5) scheme, the stochasticity is not only due to the random hot-water usage events (and resulting stochastic consumer profiles), but also due to the randomizing nature of PEM scheme, which is evident by comparing Fig. 6(a) and (b). The standard TCL control logic leads to long ON/OFF-periods. However, PEM(5) showcases the TCLs' availability throughout the period. This randomized flexibility is leveraged in the coordinated schemes to tightly track a reference signal.

2) Comparing closed-loop reference tracking PEM schemes: In the case of the tracking PEM schemes, the VPP accepts/rejects anonymous pings based on the real-time difference between aggregate load draw and the reference signal. The difference represents a power imbalance but the VPP is unaware of the specific power rating from each request. To make up for this lack of information, the VPP uses only the estimated average power rating of TCLs (5kW) to determine how many requests to accept each discrete-time sampling period k.

Recall, if the local temperature conditions precipitate the need to opt-out of the PEM scheme, the TCL will do so to

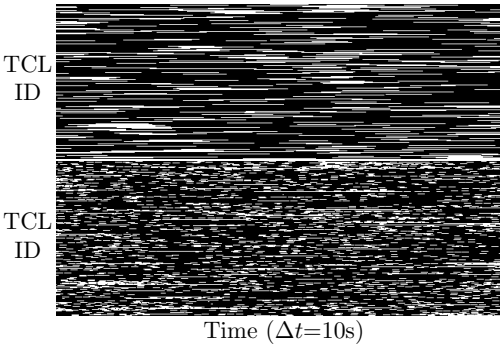

Fig. 6: The randomization of access is evident from the TCL ON/OFF status. White means TCL is ON; Black means TCL is OFF. Only every fifth TCL is displayed for graphical purposes.

preserve quality of service. The evolution of the number of TCLs that have their requests accepted and denied along with those that have opted out and did not request is provided for tracking PEM(5) in Fig. 7.

In terms of quality of service, we consider three metrics (see Table II): average, ΔT_μ, and standard deviation, ΔT_σ, of the absolute deviation from temperature set-points for all time after reference is enabled and the average cycling per hour (ACPH). A cycle is considered any OFF-to-ON or ON-to-OFF event. From the table, it is clear that T1 and T2 are very similar except in terms of the ACPH. It is expected that PEM with a 5-minute control epoch length will cycle more often. However, as epoch length increases ACPH decreases significantly as seen with T4. Interestingly, T3 deviates the least from desired set-point of resistive water heater while also tracking the reference signal. Notice that the average temperature dead-band is $\pm 3.3°$C, which results in 90% of TCLs staying within their limits at all time for all four schemes. Indeed, the results further indicate that the PEM approaches (T2,T3,T4) do not negatively impact quality of service when compared with standard TCL hysteresis control (T1).

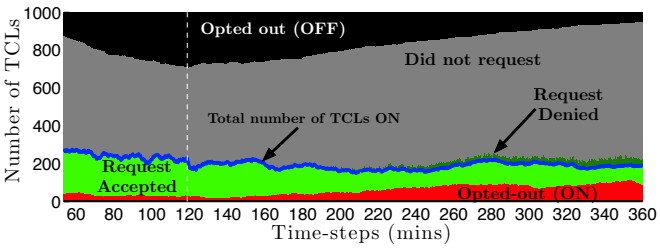

Fig. 7: The evolution of TCL requests and opt-out decisions. Each color represents a different sub-class of TCLs. Having a large population of "request denied" is helpful for tracking control, while a large "Opted-out (ON)" can interfere with tracking performance.

As shown in Fig. 5, the coordinated PEM(5) and PEM(30) schemes are both able to track the signal. With the shorter control epoch length, PEM(5) is able to better track large step-changes when the reference is first enabled at minute 120 and faster ramping events, such as, at minutes 160, 280, and 340. The reason PEM(30) is unable to track larger ramping events and step-changes in the reference

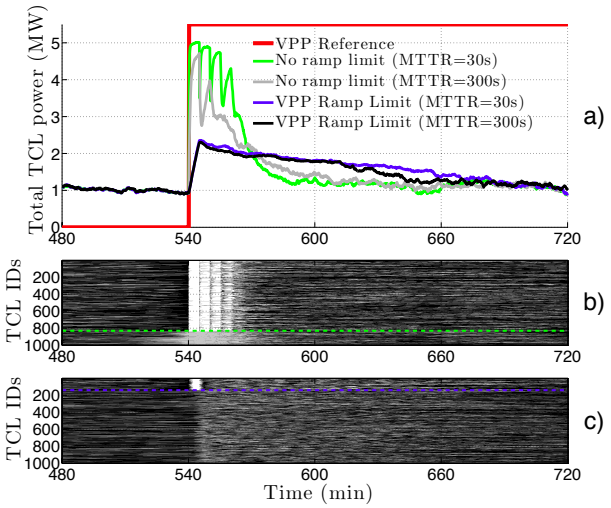

Fig. 8: Examining PEM under extreme operating conditions reveals in (a) a potential for oscillations due to local (in-time) synchronization between different packetized TCLs. However, a 300kW/min ramp-rate limit on the VPP alleviates synchronization concerns and reduces the rebound effects as shown in (b)(No ramp limit - MTTR= 30) and (c)(VPP Ramp limit - MTTR= 30). In (b) and (c), TCLs under PEM that have not opted out and have been turned ON for at least 30 seconds in the first three minutes after the change in reference are considered synchronized. The green and blue lines in (b) and (c) separate synchronized and unsynchronized groups. From (b), the devices that are ON before the reference step-change have clearly opted out and do not feature in the synchronizing group.

is due to the 30-minute control epoch length, which allocates the TCLs in larger blocks of time. The average, E_M, and root-mean-square, E_{RMS}, tracking errors are tabulated in Table II for each control scheme.

Remark III.1 *Note that PEM(5) and PEM(30) are able to track the time-varying reference signal with an average tracking error below 2% and an RMS of less than a handful of water-heaters for over four hours without optimizing, pre-positioning, or predicting the loads or the reference. Furthermore, it should be emphasized that the TCLs are coordinated with only anonymous pings sent to the VPP.*

3) Rebound effects under PEM: A final investigation of PEM performance entails operation under severe operating conditions. More specifically, four versions of PEM(5) are employed for a load reduction event to investigate the rebound effect of packetized water heaters when all requests over a 6-hour period are rejected. Figure 8(a) illustrates this 6-hour reduction event (minutes 180-540). After the six hours of the VPP rejecting every request, the water heaters are allowed full access to the grid (i.e., all requests are accepted by VPP). Once the 6-hour period concludes, the VPP accepts all requests as indicated by the reference[3].

[3]This scenario can be thought of as an extreme peak reduction event by a utility or load aggregator.

TABLE II: Performance metrics for control schemes

Scheme	$E_M(\%)$	E_{RMS}(kW)	$\Delta T_\mu \pm \Delta T_\sigma(^\circ\text{C})$	ACPH
T1	23	231	2.02 ± 0.56	0.6 ± 0.3
T2	25	243	2.15 ± 0.40	5.0 ± 2.0
T3	0.6	15	1.91 ± 0.49	4.0 ± 1.6
T4	1.3	25	2.34 ± 0.41	1.4 ± 0.6

While the VPP declines all packet requests, the packetized water heaters must opt out of PEM to maintain QoS. Thus, the only power consumed during the 6-hour period is from water heaters that have opted out (i.e., with temperature below $T_n^{\text{set}} - T_n^{\text{set,PEM}} = 0.92 \times T_n^{\text{set}}$). As illustrated in Fig. 8(a), the aggregate opt-out consumption stabilizes around 1000kW for minutes 360-540. However, during this opt-out period, large groups of water-heaters naturally become synchronized and, at minute 540, when the reference abruptly changes, the VPP can experience large MW-scale (damped) oscillatory ramping events as shown in Fig. 8(a,b) for the "No ramp limit" cases where the VPP just accepts all incoming requests (e.g., green and gray). The large oscillations occur at a period equal to the control epoch until the randomizing nature of PEM desynchronizes the population. Note that shorter MTTR begets increased oscillations as more frequent requests prevent desynchronization. To prevent this large spike, the VPP is equipped with a ramp-rate limit (e.g., 300kW per minute), which effectively limits the number of packets that can be accepted during an interval (e.g., ca. 60 packets per minute). As displayed in Fig. 8(a,c), the VPP ramp-rate limit is clearly successful in mitigating the rebound effect as it prevents re-synchronization of packetized loads between control epochs and limits the initial rebound peak by about 75%. The drawback of the VPP ramp-rate limit is a longer recovery period, which could impact future availability of PEM for tracking or additional peak reduction services. Thus, synchronization effects that can plague certain load control schemes does not seem a concern in PEM. In fact, the tracking PEM(5) and PEM(30) from Fig. 5 both employ the ramp-rate limited VPP during the tracking periods without impacting tracking performance.

IV. Conclusion

This paper presents and demonstrates a novel and scalable distributed coordination scheme, known as packetized energy management (PEM), for dispatching thermostatically controlled loads (TCLs). This approach builds on core ideas from digital communications theory, in which users of a resource make probabilistic requests to access this resource. Under the PEM scheme, each TCL asynchronously and stochastically requests the authority to turn on for a fixed amount of time (i.e., request an "energy packet") and the aggregate population is coordinated in the sense that if there is a surplus of load, incoming requests are denied. The anonymous load requests enable PEM to accurately track an aggregate reference signal.

Simulation results for $N = 1000$ electric hot water heaters are presented. To fully take advantage of PEM's randomization and control flexibility, it is important to develop an analytical framework wherein conditions on automata design and VPP interaction can guarantee certain performance. Thus, work is currently focusing on developing the analytical underpinnings of PEM to understand the packetized approach's fundamental limitations in control and stability [19]. Additional future work will also implement packetized energy management on a physical test-bed and to extend modeling efforts to other dynamic net-loads (e.g., flexible energy storage and compressor-driven loads).

References

[1] D. S. Callaway and I. A. Hiskens, "Achieving Controllability of Electric Loads," *Proceedings of the IEEE*, vol. 99, pp. 184–199, Jan. 2011.

[2] M. Morgan and S. Talukdar, "Electric power load management: Some technical, economic, regulatory and social issues," *Proceedings of the IEEE*, vol. 67, no. 2, pp. 241 – 312, 1979.

[3] F. C. Schweppe, R. D. Tabors, J. L. Kirtley, H. R. Outhred, F. H. Pickel, and A. J. Cox, "Homeostatic utility control," *IEEE Trans. Power Apparatus and Systems*, no. 3, pp. 1151–1163, 1980.

[4] J. L. Mathieu, S. Koch, and D. S. Callaway, "State Estimation and Control of Electric Loads to Manage Real-Time Energy Imbalance," *IEEE Transactions on Power Systems*, vol. 28, no. 1, pp. 430–440, 2013.

[5] Y. Chen, A. Busic, and S. P. Meyn, "State estimation and mean field control with application to demand dispatch," in *IEEE Conference on Decision and Control*, pp. 6548–6555, Dec. 2015.

[6] W. Zhang, J. Lian, C.-Y. Chang, and K. Kalsi, "Aggregated Modeling and Control of Air Conditioning Loads for Demand Response," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 4655–4664, 2013.

[7] S. Esmaeil Zadeh Soudjani and A. Abate, "Aggregation and Control of Populations of Thermostatically Controlled Loads by Formal Abstractions," *IEEE Transactions on Control Systems Technology*, vol. 23, pp. 975–990, May 2015.

[8] S. P. Meyn, P. Barooah, A. Busic, Y. Chen, and J. Ehren, "Ancillary Service to the Grid Using Intelligent Deferrable Loads," *IEEE Transactions on Automatic Control*, vol. 60, pp. 2847–2862, Nov. 2015.

[9] Y. Chen, A. Busic, and S. P. Meyn, "State Estimation for the Individual and the Population in Mean Field Control with Application to Demand Dispatch," *IEEE Transactions on Automatic Control*, vol. 62, no. 3, pp. 1138 – 1149, 2017.

[10] B. Zhang and J. Baillieul, "A packetized direct load control mechanism for demand side management," in *IEEE Conference on Decision and Control*, pp. 3658–3665, 2012.

[11] B. Zhang and J. Baillieul, "A novel packet switching framework with binary information in demand side management," in *IEEE Conference on Decision and Control*, pp. 4957–4963, 2013.

[12] B. Zhang and J. Baillieul, "Control and Communication Protocols Based on Packetized Direct Load Control in Smart Building Microgrids," *Proceedings of the IEEE*, vol. 104, no. 4, pp. 837–857, 2016.

[13] J. Frolik and P. Hines, "Urgency-driven, plug-in electric vehicle charging," in *IEEE Innovative Smart Grid Technology Conference (Europe)*, 2012.

[14] P. Rezaei, J. Frolik, and P. D. Hines, "Packetized plug-in electric vehicle charge management," *IEEE Transactions on Smart Grid*, vol. 5, no. 2, pp. 642–650, 2014.

[15] J. Frolik and P. Hines, "Random access, electric vehicle charge management," in *IEEE International Electric Vehicle Conference*, 2012.

[16] C. Goh and J. Apt, "Consumer strategies for controlling electric water heaters under dynamic pricing," *Proceeding of Carnegie Mellon Electricity Industry Center*, 2004.

[17] J. Kondoh, N. Lu, and D. J. Hammerstrom, "An Evaluation of the Water Heater Load Potential for Providing Regulation Service," *IEEE Transactions on Power Systems*, vol. 26, no. 3, pp. 1309–1316, 2011.

[18] ASHRAE, "Chapter 49: Service Water Heating," in *ASHRAE Applications Handbook*, pp. 49.1–49.22, ASHRAE, Oct. 2002.

[19] L. A. Duffaut Espinosa, M. Almassalkhi, P. Hines, S. Heydari, and J. Frolik, "Towards a Macromodel for Packetized Energy Management of Resistive Water Heaters," in *Conference on Information Sciences and Systems*, Mar. 2017.