

A Novel Packet Switching Framework with Binary Information in Demand Side Management*

Bowen Zhang¹ and John Baillieul²

Abstract—Inspired by the success of packet switched data transfer that improves the fairness and efficiency in communication networks, this paper proposes a novel concept of packetized energy that is capable of improving the performance of the smart grid in providing demand response. Here the term *packetized* refers to a temporal quantization into fixed length intervals of energy authorization based on binary information – energy is either requested by an appliance if it wishes to consume or withdrawn if the desired consumption level is reached. We model the energy request and withdrawal process as a queuing system with multiple servers and probabilistic returns. We derive an analytical expression for the mean waiting time (MWT) per authorization as a function of the packet length. We show that with short packet duration the MWT of the packet switching framework is smaller than the system without packet switching, and that the total waiting time (TWT) to complete the service remains the same. Consequently, the packet switched framework guarantees fairness in energy delivery when limited resources are available. We also provide a sensitivity analysis of the MWT and the TWT in terms of the packet length. Results show that the TWT is more robust to the change of packet length than the MWT. Simulations are provided to verify theoretical results.

I. INTRODUCTION

Packet switched digital communication technologies have enabled almost all modern networks for transmission of data, voice, and various kinds of media. One of the important benefits of packet switching is the guarantee of fairness in data delivery [1]. Data from multiple sources are cut into pieces of small data packets and then sent alternatively from the source to the destination sharing a common communication link in such a way that the communication channel is equitably shared. This success and the fairness it provides has led us to consider the possibility of quantizing the energy that utilities provide to their customers. Our goal has been to understand how the concept of *energy packets* could be developed in such a way that energy systems and communication systems might share similar beneficial features – such as fairness. We note that both systems provide service in which certain quantities are delivered from a source to a destination, and that the resources needed to carry out delivery will typically be limited in both. We note that even if the energy resources are available, a power system operator may still wish to provide reduced amounts of energy

so as to avoid consumption spikes in peak hours. Limiting the energy that is supplied to end users is generally known as *demand response*. It is briefly reviewed below.

Research has been done in demand response to evaluate load shedding with different control approaches and associated loss of consumer utility. The present paper investigates direct load control (DLC) in which an electricity provider has direct access by which to control the operation of individual appliances in a building. The operator also has direct knowledge of individual room temperatures. It is assumed that the system operator can directly shift the appliances' set point to change the appliances duty cycle and the aggregated consumption. For example, the appliances' set points can be adjusted automatically as a function of outside temperature [2] [3] or electricity price [4] [5]. We can also directly control the on/off switch of appliances' actuators to instantly shed loads. In previous work, Pacific Northwest Laboratory implemented a *Grid Friendly Appliances Project* to evaluate the performance of under-frequency load shedding by controlling the on/off switch of water heaters and dryers [6]. Queueing theory has been previously used to evaluate the disutility characterized by system waiting time when different resource levels are available in demand response [7].

Following our earlier work [8], this paper continues to discuss the use of energy packets where we define a packet as a specified time interval of authorized electricity use. When a packet is scheduled, an appliance gets the authorization to consume electricity for a fixed amount of time Δt which is the length of the temporal packet. Different from [8] where the building operator can obtain continuous state information from local appliances, this paper focuses on a scenario where the building operator receives binary information – an energy request by an appliance if it wishes to consume or an energy withdrawal if the desired consumption is completed. With the recent appearance of Internet-enabled electric energy components like the Nest[®] thermostat, it appears that the networked control technologies [9] that are needed for effective demand response are already in place.

The contribution of this paper is to propose a new energy distribution protocol for rationing limited amounts of energy to a pool of appliances. The key to realizing this protocol is the concept of an energy packet. It is shown that we can control the number of expected appliances and the appliances' arrival rate by designing energy packets with the appropriate authorization length. A design consideration is that with a small value of packet length Δt , the appliances' mean waiting time (MWT) per packet authorization can be reduced while 1) the total waiting time (TWT) to consume the desired

*The authors gratefully acknowledge support of the U.S. National Science Foundation under EFRI Grant 1038230

¹Corresponding author. Division of Systems Eng., Boston University, 15 St. Marys St., Brookline, MA 02446, email: bowenz@bu.edu

²Dept. of Electrical and Computer Eng., Dept. of Mechanical Eng., and Division of Systems Eng., Boston University, 110 Cummings St., Boston, MA 02215, email: johnnb@bu.edu

number of packets, and 2) the duration of consumption per unit time of waiting remain the same compared with the system without the packet-based allocation protocol. This in turn guarantees fairness in energy dispatch because no appliance has the priority to consume a long duration of energy while letting other appliances wait a comparatively long time to get energy. The key to the protocol's reducing the MWT for a pool of appliances is its operation as a FIFO queue with client re-entry at the end of each time interval Δt . A sensitivity analysis shows that changing the energy packet length Δt has more impact on the MWT than on the TWT, which means reducing Δt provides greater fairness in allocation.

The paper proceeds as follows. In Sec.II, we formulate the packet switching framework as a queuing system with deterministic service time and probabilistic returns. In Sec.III, we focus on a single packet system and show how the MWT per packet authorization depends on Δt . In Sec.IV we extend the analysis to multi-packet systems and propose two evaluation metrics for the new protocol. Sec.V provides sensitivity analysis, Sec.VI provides simulation, and Sec.VII concludes.

II. PROBLEM FORMULATION

A. Energy Packet Queuing System

We consider thermostatic appliances with two operating states either on or off. We assume as in [7], [10] that each appliance has exponential holding time with rate λ if the operating state is off and with rate μ if the state is on. Denote the total number of appliances in the building by N_c , and consider a demand response where a maximum of m appliances can be turned on at any time. When energy is provided according to the packet protocol with packet length Δt , we can formulate the problem as a queuing system having deterministic service time Δt with probabilistic returns as in Fig.1. An appliance will re-enter the queue to request additional energy if the desired level of consumption is not reached within Δt . The probability of re-entry p can be determined by parameters μ and Δt ,

$$p = \int_{\Delta t}^{+\infty} \mu e^{-\mu t} dt = e^{-\mu \Delta t}. \quad (1)$$

To transform the non-Markov queuing system, we use the *method of stages* approach [11] to approximate the deterministic packet duration Δt with M series connected identical sub-packets each having exponential duration with rate $M/\Delta t$, see Fig.2. Upon arrival, an appliance will consume the first sub-packet and proceed to the next when the previous sub-packet is consumed. The service is completed when an appliance consumes all M sub-packets. It can be verified that the total time $\bar{\Delta t}$ to consume the M sub-packets has mean and variance,

$$E(\bar{\Delta t}) = \Delta t, \sigma^2(\bar{\Delta t}) = \frac{\Delta t^2}{M}.$$

The variance of $\bar{\Delta t}$ approaches 0 as M increases and $\bar{\Delta t} = \Delta t$ with probability 1, which means we can approximate the

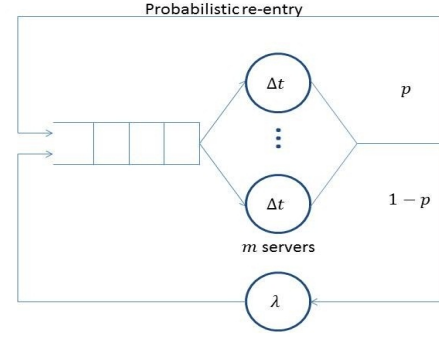


Fig. 1. Closed loop queuing system with probabilistic feedback that depends on packet length Δt .

deterministic energy packet with large number of exponentially distributed sub-packets. It should be noted that the concept of sub-packet is only used for the Markov system approximation, there is no real sub-packet in practical applications.

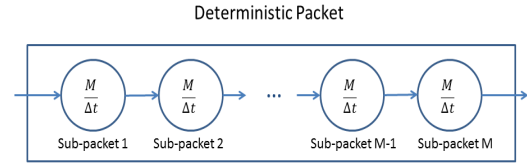


Fig. 2. Approximation of deterministic energy packet with M series connection of exponentially distributed sub-packets.

B. Energy Packet Markov Chain

To simplify analysis, we first formulate the Markov single server system with $m = 1$, and compare its performance with traditional energy distribution protocol without the energy packet in Sec.III. We will extend the result to multi-server systems with $m > 1$ in Sec.IV.

Define the state i as the number of sub-packets that need to be consumed to satisfy all the requests in the queue at a given time. Possible states are integers from 0 to $N_c M$. (For example, if 5 appliances wait in the queue and 1 appliance is served with 7 sub-packets left to consume, then the state of the system is $5M + 7$.) For a given state i , the number of appliances requesting energy packets is $X_i = \lceil \frac{i}{M} \rceil$, and $N_c - X_i$ is the number of appliances in the off state. There are four types of events that can possibly happen for state i :

Event A: an arrival of a packet request from appliances in the off operating state. This event happens with rate $\lambda_i = (N_c - X_i)\lambda$, and the state increases by M because additional M sub-packets are requested. ($X_i \rightarrow X_{i+M}$)

Event B: an appliance consumes one sub-packet which is not the last sub-packet. This event happens with rate $M/\Delta t$, and the state decreases by 1. ($X_i \rightarrow X_{i-1}$)

Event C: an appliance consumes the last sub-packet and switches to the off operating state. This event happens with rate $(1-p)M/\Delta t$, and the state decreases by 1. ($X_i \rightarrow X_{i-1}$)

Event D: an appliance consumes the last sub-packet and requests additional energy packet. This event happens with rate $pM/\Delta t$, and the state increases by $M-1$. ($X_i \rightarrow X_{i+M-1}$)

Note that we need to consider events *C* and *D* only for state $i = kM + 1, k = 0, 1, \dots, N_c - 1$, where the appliance is consuming the last sub-packet. Fig.3 illustrates detailed transitions where events *A, B, C, D* are represented by black, green, black dashed, and green dashed arrows, respectively.

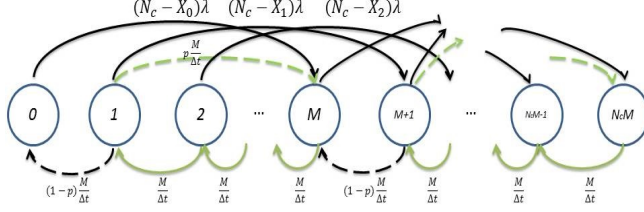


Fig. 3. Transition rate diagram for the single server system. Events *A, B, C, D* are represented by black, green, black dashed, and green dashed arrows, respectively.

III. SINGLE PACKET SYSTEM ANALYSIS

In this section we derive the MWT, $W_{M/D/1}$, in terms of Δt . According to Little's Law,

$$W_{M/D/1} = s - \Delta t = \frac{E(X)}{\lambda_{ex}} - \Delta t, \quad (2)$$

where s is the system time, $E(X)$ and λ_{ex} are the expected number of appliances and appliance arrival rate of the system in a long term. Denoting the steady state probability distribution by $\{\pi_0, \pi_1, \dots, \pi_{MN_c}\}$, we can express $E(X)$ as

$$E(X) = \sum_{i=0}^{MN_c} \pi_i X_i. \quad (3)$$

The arrival rate is the sum of arrival from appliances in the off operating state and probabilistic re-entry,

$$\lambda_{ex} = \sum_{i=0}^{M(N_c-1)} \pi_i (N_c - X_i) \lambda + p \frac{M}{\Delta t} \sum_{k=0}^{N_c-1} \pi_{kM+1}, \quad (4)$$

where the first term is the arrival from the off state, and the second is the arrival by re-entry. Before deriving $E(X)$ and λ_{ex} as a function of Δt , we prove the following lemma that will be used in proposition 1.

Lemma 1 Assuming that the steady state probability distribution $\{\pi_0, \pi_1, \dots, \pi_{MN_c}\}$ exists, then the probability that an appliance is consuming at the j^{th} sub-packet is,

$$P_j = \sum_{k=0}^{N_c-1} \pi_{kM+j} = \frac{1}{M} (1 - \pi_0), \quad j = 1, \dots, M, \quad (5)$$

where k is the number of appliances waiting in the queue.

Proof. The only state such that no packet request exists is state $i = 0$ with probability π_0 . Hence the following equation holds,

$$\sum_{j=1}^M P_j = 1 - \pi_0. \quad (6)$$

Based on Fig.3, we have the following equations for state $i = 1$ as the transition rate into the state is equal to the rate

out of the state after the steady state probability distribution is reached,

$$\left[\frac{M}{\Delta t} + (N_c - X_1) \lambda \right] \pi_1 = \frac{M}{\Delta t} \pi_2. \quad (7)$$

Similarly for states $kM + 1, k = 1, \dots, N_c - 1$, we have

$$\left[\frac{M}{\Delta t} + (N_c - X_{kM+1}) \lambda \right] \pi_{kM+1} = \frac{M}{\Delta t} \pi_{kM+2} + (N_c - X_{(k-1)M+1}) \lambda \pi_{(k-1)M+1}. \quad (8)$$

Summing (7) and (8) for k from 0 to $N_c - 1$ on both sides

$$\sum_{k=0}^{N_c-1} \pi_{kM+1} = \sum_{k=0}^{N_c-1} \pi_{kM+2},$$

which is equivalent to $P_1 = P_2$. Similarly we can prove that all P_j are equal, therefore (5) holds. ■

A straightforward way to understand this lemma is the following: since the M sub-packets are identical and independent of the number of appliances that are operating, having the same exponential service time, the probability that an appliance is consuming a specific sub-packet should all be equal given that the appliance is being served.

Proposition 1 The expected number of appliances $E(X)$ and arrival rate λ_{ex} in steady state for our packetized energy scheduling system are given by

$$E(X) = N_c - \frac{(1-p)(1-\pi_0)}{\lambda \Delta t}, \quad \lambda_{ex} = \frac{1}{\Delta t} (1 - \pi_0), \quad (9)$$

respectively.

Proof. The proof is based on the steady state transition rate flow conservation shown in Fig.3. We have the conservation flow starting for state $i = 0$,

$$(N_c - X_0) \lambda \pi_0 = (1-p) \frac{M}{\Delta t} \pi_1, \quad (10)$$

for states $i = 1, \dots, M-1$,

$$\left[\frac{M}{\Delta t} + (N_c - X_i) \lambda \right] \pi_i = \frac{M}{\Delta t} \pi_{i+1}, \quad (11)$$

and for state $i = M$,

$$\left[\frac{M}{\Delta t} + (N_c - X_M) \lambda \right] \pi_M = (1-p) \frac{M}{\Delta t} \pi_{M+1} + p \frac{M}{\Delta t} \pi_1 + (N_c - X_0) \lambda \pi_0. \quad (12)$$

Based on (10) – (12)

$$\sum_{i=1}^M (N_c - X_i) \lambda \pi_i = (1-p) \frac{M}{\Delta t} \pi_{M+1}. \quad (13)$$

Similarly for states $i > M$ and $k = 1, \dots, N_c - 2$ number of appliances waiting in the queue

$$\sum_{i=kM+1}^{(k+1)M} (N_c - X_i) \lambda \pi_i = (1-p) \frac{M}{\Delta t} \pi_{(k+1)M+1}. \quad (14)$$

According to (10), (13), and (14)

$$\sum_{i=0}^{M(N_c-1)} (N_c - X_i) \lambda \pi_i = (1-p) \frac{M}{\Delta t} \sum_{k=0}^{N_c-1} \pi_{kM+1}. \quad (15)$$

Rearranging terms, we get

$$\sum_{i=0}^{M(N_c-1)} (N_c - X_i) \lambda \pi_i + p \frac{M}{\Delta t} \sum_{k=0}^{N_c-1} \pi_{kM+1} = \frac{M}{\Delta t} \sum_{k=0}^{N_c-1} \pi_{kM+1}. \quad (16)$$

According to Lemma 1, the R.H.S of (16) equals $(1 - \pi_0)/\Delta t$. The L.H.S is the steady state arrival rate λ_{ex} , and therefore $\lambda_{ex} = (1 - \pi_0)/\Delta t$. From (15), it is easy to show that $E(X) = N_c - \frac{(1-p)(1-\pi_0)}{\lambda \Delta t}$. ■

From Proposition 1,

$$W_{M/D/1} = \frac{N_c \Delta t}{(1 - \pi_0)} - \frac{1 - p}{\lambda} - \Delta t. \quad (17)$$

To compare the performance of our packetized scheduling system with a queuing system that does not apply quantized energy packets, we have the MWT of the M/M/1 system, $W_{M/M/1}$, with total N_c appliances having the same duty cycle parameters given by [11],

$$W_{M/M/1} = \frac{N_c}{\mu(1 - \pi_0)} - \frac{1}{\lambda} - \frac{1}{\mu}. \quad (18)$$

From (17) and (18), the MWT for the M/D/1 and the M/M/1 systems have structural similarities. When the total number of appliances N_c is large, the server utilization must approach 1 and therefore $\pi_0 \rightarrow 0$ in both systems. (17) and (18) become

$$W_{M/D/1}(\Delta t) = (N_c - 1)\Delta t - \frac{1-p}{\lambda}, \quad (19)$$

$$W_{M/M/1} = (N_c - 1)\frac{1}{\mu} - \frac{1}{\lambda},$$

respectively. We will use (19) for the $W_{M/D/1}$ and $W_{M/M/1}$ in the rest of the paper which rests on the assumption of large number of appliances in the system.

Corollary 1 A necessary condition for choosing Δt such that the MWT of the packetized M/D/1 system is smaller than the corresponding non-packet M/M/1 system is that $\Delta t < 1/\mu$.

Proof. From (19), the MWT of M/D/1 system increases as Δt increases. Let $\Delta t = 1/\mu$,

$$W_{M/D/1}(\Delta t) = \frac{N_c - 1}{\mu} - \frac{1 - p}{\lambda} > \frac{N_c - 1}{\mu} - \frac{1}{\lambda} = W_{M/M/1}.$$

Therefore $\Delta t < 1/\mu$ is a necessary condition. ■

The intuition behind the corollary is that it is a waste of opportunity to authorize a energy packet longer than the expected holding time.

IV. MULTI-PACKETS SYSTEM ANALYSIS

To evaluate the MWT of the packetized scheduling system with multiple m servers, we use the following approximation for $W_{M/D/m}$ [12],

$$W_{M/D/m} \cong \frac{W_{M/M/m}}{W_{M/M/1}} W_{M/D/1}. \quad (20)$$

It should be noted that the approximation (20) is based on the assumption of Poisson arrival which is correct for our closed system in a *steady state* sense. The actual instant-by-instant arrival is not Poisson and is affected by λ, μ . This will result in approximation error and we will discuss the error with numerical simulation in Sec.VI-A.

To solve for $W_{M/D/m}$ according to (20), we first calculate $W_{M/M/m}$ by constructing the transition rate between adjacent states. If i denotes the number of appliances requesting energy packets, the arrival λ_i and the departure μ_i rate are

$$\lambda_i = (N_c - i)\lambda, \mu_i = \begin{cases} i\mu & \text{if } 0 < i \leq m \\ m\mu & \text{if } i > m \end{cases}.$$

The state i is increased or decreased by one per arrival (departure), this is a birth-and-death process with steady state probability distribution,

$$P_i = \frac{\prod_{j=0}^{i-1} \lambda_j}{\prod_{k=1}^i \mu_k} P_0.$$

Adding all P_i and using the normalization condition

$$P_0 = \left(1 + \sum_{i=1}^{N_c} \frac{\prod_{j=0}^{i-1} \lambda_j}{\prod_{k=1}^i \mu_k}\right)^{-1},$$

we are able to calculate $E(X)$, λ_{ex} . $W_{M/M/m}$ is then calculated by Little's Law.

Corollary 2 A necessary condition for choosing Δt such that the MWT of the packetized M/D/m system is smaller than the corresponding non-packet M/M/m system is that $\Delta t < 1/\mu$.

Proof. The proof is based on (20) and corollary 1. ■

The advantage of using energy packets is that the MWT per authorization is reduced, but consuming short packets per authorization requires appliance to enter the queue multiple times to have enough number of packets to reach the desired consumption. Note that an appliance re-enters the queue with probability p , then the expected number of entries n is the mean of a geometric random variable $E(n) = (1 - p) + 2p(1 - p) + \dots = 1/(1 - p)$. We define the expected total waiting time (TWT) $T_{M/D/m}$ for the completion of consumption as the product of the MWT per packet authorization and the expected number of entries needed to complete packet utilization, namely

$$T_{M/D/m}(\Delta t) = W_{M/D/m}(\Delta t) E(n). \quad (21)$$

We will use $T_{M/D/m}(\Delta t)$ as the first metric to evaluate the system performance in terms of Δt . The second metric to evaluate the system is the fraction of service time in the overall system time, $P_{M/D/m}(\Delta t)$, which can be interpreted as the energy duration that an appliance can be authorized per unit time in the queuing system, namely

$$P_{M/D/m}(\Delta t) = \frac{\Delta t}{W_{M/D/m}(\Delta t) + \Delta t}. \quad (22)$$

Similarly we define $P_{M/M/m}$ for the non-packet system,

$$P_{M/M/m}(\Delta t) = \frac{1/\mu}{W_{M/M/m} + 1/\mu}. \quad (23)$$

The following proposition gives properties of $T_{M/D/m}(\Delta t)$ and $P_{M/D/m}(\Delta t)$.

Proposition 2 The following two properties hold:

(1) $T_{M/D/m}(\Delta t)$ is a monotonically increasing function of Δt and $\lim_{\Delta t \rightarrow 0} T_{M/D/m}(\Delta t) = W_{M/M/m}$.

(2) $P_{M/D/m}(\Delta t)$ is a monotonically decreasing function of Δt and $\lim_{\Delta t \rightarrow 0} P_{M/D/m}(\Delta t) = P_{M/M/m}$

Proof. (1) From (1), (19), (20), and (21)

$$\begin{aligned} T_{M/D/m}(\Delta t) &= \frac{W_{M/M/m}}{W_{M/M/1}} W_{M/D/1} E(n) \\ &= \frac{W_{M/M/m}}{W_{M/M/1}} \left(\frac{(N_c - 1)\Delta t}{1 - e^{-\mu\Delta t}} - \frac{1}{\lambda} \right). \end{aligned} \quad (24)$$

The monotonicity of $T_{M/D/m}(\Delta t)$ is consistent with the monotonicity of the function $f(\Delta t) = \Delta t / (1 - e^{-\mu\Delta t})$ that can be verified as an increasing function. It should be noted that $f(\Delta t)$, which is the product of Δt and $1/(1 - e^{-\mu\Delta t})$, is the expected amount of authorization that is needed to reach the desired level of comfort setting. Based on L'Hôpital's rule,

$$\lim_{\Delta t \rightarrow 0} \frac{(N_c - 1)\Delta t}{1 - e^{-\mu\Delta t}} - \frac{1}{\lambda} = \lim_{\Delta t \rightarrow 0} \frac{N_c - 1}{\mu} - \frac{1}{\lambda} = W_{M/M/1}. \quad (25)$$

Substitute (25) into (24) yields $\lim_{\Delta t \rightarrow 0} T_{M/D/m}(\Delta t) = W_{M/M/m}$.

(2) From (1), (19), (20), (22)

$$\begin{aligned} P_{M/D/m}(\Delta t) &= \frac{\Delta t}{\frac{W_{M/M/m}}{W_{M/M/1}} [(N_c - 1)\Delta t - \frac{1 - e^{-\mu\Delta t}}{\lambda}] + \Delta t} \\ &= \frac{1}{\frac{W_{M/M/m}}{W_{M/M/1}} [(N_c - 1) - \frac{1 - e^{-\mu\Delta t}}{\lambda\Delta t}] + 1}. \end{aligned} \quad (26)$$

Since $f(\Delta t)$ is a monotonically increasing function of Δt , $P_{M/D/m}(\Delta t)$ is monotonically decreasing. Note that $\lim_{\Delta t \rightarrow 0} \frac{1 - e^{-\mu\Delta t}}{\lambda\Delta t} = \frac{\mu}{\lambda}$. We thus have

$$\begin{aligned} \lim_{\Delta t \rightarrow 0} P_{M/D/m}(\Delta t) &= \frac{1}{\frac{W_{M/M/m}}{W_{M/M/1}} [(N_c - 1) - \frac{\mu}{\lambda}] + 1} \\ &= \frac{1}{\frac{W_{M/M/m}}{W_{M/M/1}} \mu + 1} = P_{M/M/m}, \end{aligned} \quad (27)$$

where the second equation is gotten by substituting (19). ■

Proposition 2 proves that we should choose small Δt to decrease the total authorization time as well as the TWT. In the limit when $\Delta t \rightarrow 0$, the TWT for each appliance is identical to the system where we do not use energy packets, while the MWT per authorization approaches 0. The use of small packets greatly increases the fairness of energy dispatch because energy authorization is not dedicated to any appliance for a long time duration, and appliances only wait for a small MWT to get energy. The advantage of the packetized scheduling system is similar to the advantage of packet switching in communication systems: the approach enhances the fairness. In addition, the fraction of service time in the overall system time $P_{M/D/m}$ increases when smaller Δt is chosen, which means appliances wait less to get per unit time of packet utilization. In the limit when $\Delta t \rightarrow 0$, $P_{M/D/m}$ is identical to the corresponding metric in the non-packet system, indicating that appliances do not lose any utility by participating in the packetized distribution protocol.

Remark It can be verified based on (19) and (21) that the monotonicity properties as well as the equalities when $\Delta t \rightarrow 0$ in Proposition 2 also hold for a single packet system with $m = 1$. Thus we should use small packet duration to reduce the waiting time per packet authorization as well as to increase the amount of energy that can be consumed per unit time of waiting.

V. SENSITIVITY ANALYSIS

A. Sensitivity in terms of Δt

We explore the sensitivity of $W_{M/D/m}(\Delta t)$ and $T_{M/D/m}(\Delta t)$ in terms of the change in Δt . We take derivatives

$$\begin{aligned} W_{M/D/m}(\Delta t)' &= \frac{W_{M/M/m}}{W_{M/M/1}} [(N_c - 1) - \frac{\mu}{\lambda} e^{-\mu\Delta t}] \\ &\approx \frac{W_{M/M/m}}{W_{M/M/1}} (N_c - 1) \end{aligned}$$

with large value of N_c , and

$$\begin{aligned} T_{M/D/m}(\Delta t)' &= \frac{W_{M/M/m}}{W_{M/M/1}} (N_c - 1) \frac{1 - e^{-\mu\Delta t} - \frac{\mu\Delta t}{\lambda} e^{-\mu\Delta t}}{(1 - e^{-\mu\Delta t})^2} \\ &= \frac{W_{M/M/m}}{W_{M/M/1}} (N_c - 1) \left[1 + \frac{e^{-\mu\Delta t} (1 - \mu\Delta t - e^{-\mu\Delta t})}{(1 - e^{-\mu\Delta t})^2} \right] \\ &< \frac{W_{M/M/m}}{W_{M/M/1}} (N_c - 1). \end{aligned}$$

The last inequality follows from $1 - \mu\Delta t - e^{-\mu\Delta t} < 0$ by Taylor series with positive energy packet $\Delta t > 0$. From the above two expression, we have $W_{M/D/m}(\Delta t)' > T_{M/D/m}(\Delta t)'$, which means the MWT is more sensitive in terms of the change in energy packet duration than the TWT, and that decreasing Δt is more helpful in providing a much fairer packet share than in reducing the TWT. This will also be numerically verified in the Sec.VI-B where we find the TWT is robust in terms of Δt . An interesting finding is the sensitivity of Δt when it is small. It can be shown that

$$\begin{aligned} \lim_{\Delta t \rightarrow 0} T_{M/D/m}(\Delta t)' &= \frac{1}{2} \frac{W_{M/M/m}}{W_{M/M/1}} (N_c - 1) \\ &\approx \frac{1}{2} \lim_{\Delta t \rightarrow 0} W_{M/D/m}(\Delta t)'. \end{aligned}$$

This indicates that the change in MWT is approximately twice as much as the change in TWT for small Δt .

B. Delay in the Probabilistic Return Channel

We modify the basic model in Sec.II to prevent appliances from frequent cycling by requiring that an appliance needs to maintain a certain duration in its duty off cycle after consuming an energy packet. Then the modified model becomes the queuing system in Fig.4, where v is the exponential delay if an appliance requests an additional packet.

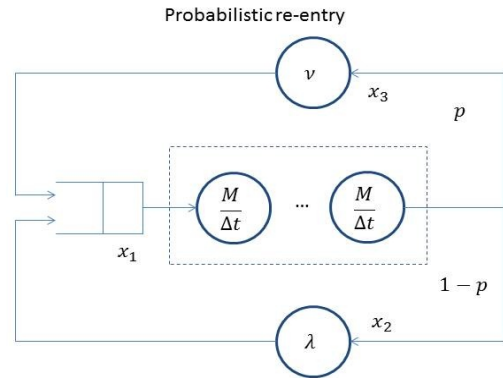


Fig. 4. Probabilistic feedback with exponential delay in probabilistic return channel to prevent appliances from frequent duty cycling.

Denote the number of appliances in the energy packet server, the duty cycle off server, and the probabilistic delay

server by x_1, x_2 , and x_3 respectively. Since the queuing system is closed, $\sum_{i=1}^3 x_i = N_c$, and we can describe the system by $\{x_1, x_2\}$ with a two dimensional transition rate diagram. For each state, there are possible sub-states describing the number of sub-packets that need to be consumed. For illustration purposes, we plot the diagram for $N_c = 3$ in Fig.5. The main figure is the two dimensional transition rate diagram. Each circle represents one specific state and the two numbers $\{i, j\}$ inside the circle are the values of the pair $\{x_1, x_2\}$. Arrows between states are the transition rate expressions based on $\lambda, \mu, \Delta t$. The upper left sub-figure is the illustration for the sub-states inside the state $\{x_1, x_2\} = \{1, 1\}$ where an appliance consumes different sub-packets in the sequence of M sub-packets. The structure of the diagram is similar for large value of N_c .

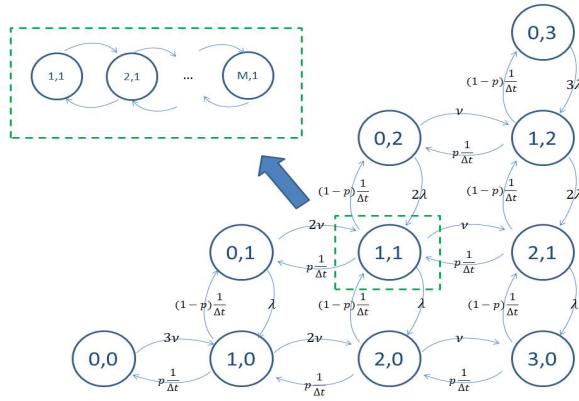


Fig. 5. Two dimensional transition rate diagram.

Based on the two dimensional transition rate diagram, we can establish results similar to lemma 1, namely the probability that an appliance is being served at the j -th sub-packet is constant and does not depend on j . As for the analysis of $E(x_1)$, λ_{av} , and the MWT, we summarize below.

Corollary 3 For the system having delay with rate v in probabilistic returns, the long term average expected arrival rate λ_{av} , average number of appliances $E(x_1)$, and the MWT are respectively given by

$$\begin{aligned}\lambda_{av} &= \frac{1}{\Delta t}(1 - \pi_0), \\ E(x_1) &= N_c - \frac{(1-p)(1-\pi_0)}{\lambda \Delta t} - \frac{p(1-\pi_0)}{v \Delta t}, \\ \bar{W}_{M/D/1} &= \frac{N_c \Delta t}{1 - \pi_0} - \frac{1-p}{\lambda} - \frac{p}{v} - \Delta t.\end{aligned}$$

Proof. The proof is based on the two dimensional transition flow conservation in Fig.5. ■

Compared with (17), the structure of the MWT remains the same. The only difference is the term p/v introduced by the exponential delay with rate v in the duty off cycle. This can be viewed as a sensitivity analysis in terms of the delay in the re-entry path, indicating that $\bar{W}_{M/D/1}$ changes linearly with the expected delay duration $1/v$.

VI. NUMERICAL SIMULATION

A. Validation of the Systems' MWT

We calculate the theoretical MWT for an M/D/1 system according to (19) and run a Monte Carlo simulation to verify the MWT for M/D/1 system with parameter combination of mean duty cycle off time $1/\lambda$ and on time $1/\mu$ ranging from 6 to 10.5 minutes and 10 to 14.5 minutes with step sizes of 0.5 minutes, respectively. The MWT increases as $1/\mu$ increases and as $1/\lambda$ decreases, and the error is within 1% of the theoretical value. This indicates that the M/D/1 modelling is accurate.

We validate the MWT approximation of a multi-packet system in Fig.6. The error is within 8% which is caused by the approximation (20). We note that the error is small when we have large value of $1/\mu$ and small value of $1/\lambda$. This may be explained that with large $1/\mu$ the feedback probability becomes large and the arrival process is dominated by the probabilistic return process, where with small $1/\lambda$ the off operational state becomes short as if it is a direct return with minimum delay. These two factors make the arrival process appear to be Poisson. Overall, the approximation (20) has satisfactory accuracy.

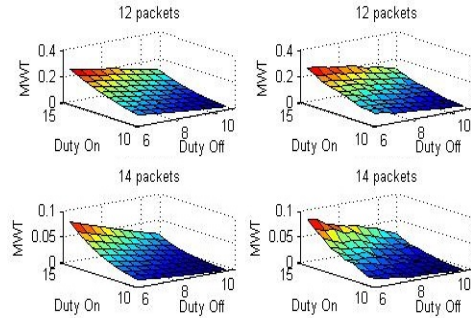


Fig. 6. M/D/m system MWT validation: MWT on left column is theoretical approximation by (20) and MWT on right column shows Monte Carlo simulation results.

B. Sensitivity of the MWT and the TWT

We validate the sensitivity of the MWT and the TWT as we vary Δt while keeping m fixed in Fig.7. Both the MWT and the TWT increase as Δt increases, but the MWT is more sensitive to the change of Δt than the TWT. For example, the MWT doubles from $\Delta t = 5$ to $\Delta t = 9$ minutes when $1/\lambda = 6$ and $1/\mu = 14.5$ minutes while the TWT remains approximately the same for the same parameters. This indicates that fairness is greatly affected by the choice of packet length while the aggregated time to reach the desired consumption is comparatively insensitive to Δt .

C. The Role of Information

The flow of real time information between energy providers and their customers will become increasingly important for realizing the full potential of the smart grid. The design of the information channels that will link providers with customers remains a work in progress, and studies of

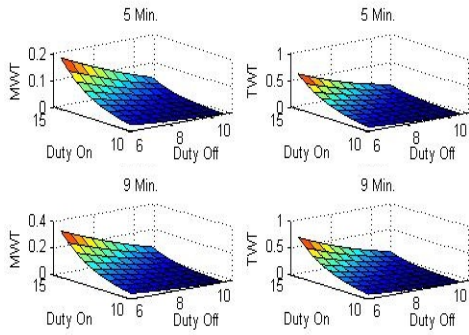


Fig. 7. Robustness analysis: the MWT and the TWT changes as Δt varies from 5 to 9 minutes.

promising alternatives are now being undertaken by a number of researchers [4], [8], [13]. In this paper we have considered the simplest form of communication in which a single binary value is communicated periodically from a local thermostatic appliance (e.g. air conditioner) to an energy provider. The binary value indicates whether the appliance senses the need to operate over the next interval or whether it is willing and able to be idle. We can compare this simplest of *binary* communication protocols with the packet scheduling protocol considered in [8] where we have *continuous* state information of local appliances. Since the building operator does not have continuous information, (s)he cannot actively schedule appliance consumptions in advance. This in turn results in a loss of consumption opportunity as shown in Fig.8 (between 15-25, 30-40 minutes, etc, the building does not fully use the resource it has), and loss of utility (10-15, 60-65 minutes, etc, appliances have to wait to be get packet authorization). If the operator has continuous state information, it can be shown that the number of energy packets utilized is smoothed and the desired comfort band provided by each appliance can be maintained by proper choice of Δt .

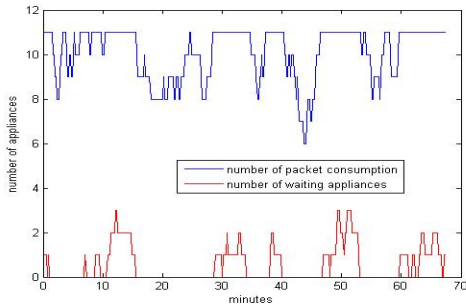


Fig. 8. Number of connected appliances and waiting appliances when the operator receives binary information.

D. The Issue of Frequent Duty Cycling

We monitored the duty cycles of individual appliances in the packetized energy scheduling protocol in order to investigate the potential issue of frequent cycling when different values of Δt are used. We find that small duration of

packet authorization yields more high frequency cycling with a reduced duration in both duty cycle on and off operation. This phenomenon indicates that the protocol of packetized energy scheduling should be applied with caution in choosing Δt , such as the minimum on/off time proposed in [14], to have reachable fairness in practical applications.

VII. CONCLUSION

This paper proposes a novel energy distribution protocol that uses the concept of temporally quantized energy packets to ensure fairness in the sharing of energy by a pool of appliances. We show that by choosing a small value of packet length, the MWT to get energy authorization is reduced with this new protocol while the TWT to get the desired amount of energy remains the same compared with a traditional energy distribution protocol. In addition, the amount of energy that an appliance can get per unit time of waiting remains the same with small Δt . These features guarantee fair energy dispatch to large number of customers when limited resources are available. Future work will compare energy packet solutions with different communication protocols and information levels.

REFERENCES

- [1] Leonard Kleinrock. *Information Flow in Large Communication Nets*. PhD thesis, Massachusetts Institute of Technology, 1961.
- [2] Ning Lu and David P. Chassin. A state-queueing model of thermostatically controlled appliances. *IEEE Trans. Power Systems*, 19(3):1666–1673, August 2004.
- [3] Ning Lu, David P. Chassin, and Steve E. Widergren. Modeling uncertainties in aggregated thermostatically controlled loads using a state queueing model. *IEEE Trans. Power Systems*, 20(2):725–733, May 2005.
- [4] Michael C. Caramanis, Ioannis Ch. Paschalidis, Christos G. Cassandras, Enes Bilgin, and Elli Ntakou. Provision of regulation service reserves by flexible distributed loads. In *the 51th IEEE Conference on Decision and Control*, pages 3694–3700, 2012.
- [5] Frauke Oldewurtel, Andreas Ulbig, Alessandra Parisio, Gran Andersson, and Manfred Morari. Reducing peak electricity demand in building climate control using real-time pricing and model predictive control. In *the 49th IEEE Conference on Decision and Control*, pages 1927–1932, 2010.
- [6] Don J. Hammerstrom. Grid friendly appliances project. Pacific Northwest GridWise Testbed Demonstration Projects, October 2007.
- [7] S. C. Lee, S. J. Kim, and S. H. Kim. Demand side management with air conditioner loads based on the queueing system model. *IEEE Trans. Power Systems*, 26(2):661–668, May 2011.
- [8] Bowen Zhang and John Baillieu. A packetized direct load control mechanism for demand side management. In *the 51st IEEE Conference on Decision and Control*, pages 3658–3665, 2012.
- [9] Panos Antsaklis and John Baillieu. Special issue on the technology of networked real-time systems. *Proceedings of the IEEE*, 95(1):5–8, January 2007.
- [10] Ioannis Ch. Paschalidis, Binbin Li, and Michael C. Caramanis. A market-based mechanism for providing demand-side regulation service reserves. In *the 50th IEEE Conference on Decision and Control and European Control Conference*, pages 21–26, December 2011.
- [11] Christos G. Cassandras and Stéphane LaFortune. *Introduction to Discrete Event Systems*. Springer, 2007.
- [12] Leonard Kleinrock and Richard Gail. *Queueing Systems, Problems and Solutions*. John Wiley & Sons, Inc., 1996.
- [13] Arman C. Kizilkale, Shie Mannor, and Peter E. Caines. Large scale real-time bidding in the smart grid: A mean field framework. In *the 51st IEEE Conference on Decision and Control*, pages 3680–3687, 2012.
- [14] Badri Ramanathan and Vijay Vittal. A framework for evaluation of advanced direct load control with minimum disruption. *IEEE Trans. Power Syst.*, 23(4):1681–1688, November 2008.