Demand Response Using Service Curves

Jean-Yves Le Boudec, *Fellow, IEEE,* Dan-Cristian Tomozei, *Member, IEEE*

Abstract—Existing methods for demand response either assume direct control of appliances by supplier, or assume that consumers adapt their load by reacting to pricing signals. The former are intrusive and might not scale well; the latter expose consumers to price volatility and require detailed awareness of time varying prices. We propose an alternative approach, based on "service curves", which uses the following two ingredients. (1) The rate at which a consumer may draw power from the grid may be controlled by real time signals. Typical concerned devices would be heating systems, air conditioners and e-car batteries. (2) However, consumers are guaranteed that, over any window of time t, the amount of energy that may be drawn is at least equal to b(t), where the function b() (called the service curve) is agreed upon by contract at subscription time. The contract also specifies the maximum power that may be drawn, as well as a fixed price per unit of energy. Hence users are protected from price variability, at the expense of possible, but upper bounded, delays. With a proper service curve definition, an operator may distribute small service reductions in order to alleviate the impact of massive incoming demand onto the power grid. Consumers are able to observe the past service control signals that they received and can compute optimal load schedules from this and their service curve contracts, using only local information. Thus, this provides a distributed, scalable and robust demand response mechanism.

Index Terms—Demand Response, Service Curves, Load Switches

I. INTRODUCTION

Demand response is often assumed to be triggered by pricing signals [1]. However, dynamic pricing comes with a number of issues, in particular, consumer exposure to large price volatility might be socially undesirable. Further, it has been shown [4] that, even in a honest market, dynamic pricing may result in high volatility on the energy market. An alternative to end consumer pricing is implemented in load switches or centralized remote control. For example, the Voltalis *bluepod* switches off thermal electrical loads (boiler, thermal electrical heating) [5, chapter 4]. When and for how long the load may be switched off is under the control of the Voltalis company, however, consumers receive the guarantee that such service inhibitions occur for a total duration of at most 30 mn per day. Pricing is not involved, and consumers pay per unit of consumed energy. In short, consumers are exposed to delayed supply rather than price volatility.

In this paper we propose an extension of such a scheme, called "service curves"; it is more general, and could serve

non thermal loads, in particular plug-in electric vehicle (PEV) loads. It is inspired by the service curve approach used in the integrated services Internet framework [3]. With our scheme, the control exercised on demand need not be On/Off, as it is with Voltalis's devices. We show in Section VI that this may have benefits for reducing variance of the load.

Unlike centralized remote control, the service curves scheme does not require distribution systems operators (DSOs) or market aggregators to directly interfere with consumers' appliances, it provides one level of separation between consumers and the distribution system.

The contributions of our paper are as follows.

1) We introduce the concept of service curves. This framework supports demand response without exposing consumers to price volatility, at the expense of some limited delay or throttle.

2) We show how the framework allows a DSO[1] to exercise some control on demand in order to smoothen the load curve, or even delay some demands in case of stochastic load surge or supply reduction. In the case of random but stationary load, it is best for the DSO to send smooth control signals. In contrast, for reducing peak demand, it is best for the DSO to send binary signals. We also show how consumer appliances can compute online schedules using only locally observed information.

3) We relate the framework to max-plus calculus and derive rules and methods to define and implement service curve constraints.

We define our framework in Section II, and give two possible examples in Section III. In Section IV we discuss some properties of service curves. The implications for consumers and distribution systems operators are described in Sections V to VII.

II. DEFINITION OF THE SERVICE CURVE APPROACH

We assume that the rate at which a consumer may draw power from the grid may be controlled through load control signals. Such signals can be captured by a smart home controller who then uses them to compute a schedule for this home's appliances. The scheduled appliances are those that may have some elasticity, for example heating systems with or without heat pumps, air conditioners and PEV batteries. Non elastic appliances would not be controlled. The controls can use cellular networks or the smart grid; protocols and communication systems required to implement them are outside the scope of this paper (see [2] for a description of such a system).

---

[1] Demand response may be exercised by the DSO or other players, such as market aggregators. In the context of this paper it is irrelevant to know which entity issues the control signal; we use the term DSO as a general concept for the entity that issues the load control signal $u(t)$ defined in Section II.
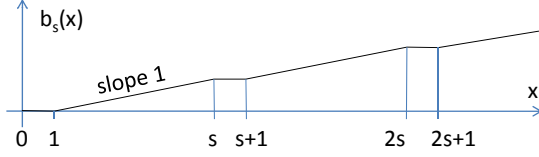
Fig. 1. The function b_s

Let $u(t)$ be the *control signal* received by one specific consumer and $z(t)$ the net power drawn by this consumer from the grid at time t (in Watts). [2] The contract between consumer and distribution system operator specifies the following constraints. First

$$z(t) \leq u(t) \leq z_{\max} \text{ for all } t' < t \qquad (1)$$

i.e. the consumed energy may not exceed the level $u(t)$ specified by the distributor; $z_{\max}$ is the maximum that may ever be allowed.

Second, the consumer receives the guarantee that throttling cannot be excessive:

$$\int_{t'}^{t} u(\tau)d\tau \geq \beta(t - t') \text{ for all } t' < t \qquad (2)$$

where $\beta : \mathbb{N} \to [0, +\infty)$ is a function defined at contract subscription time. We say that β is a *service curve*, and Eq. (2) can be paraphrased by saying that the control signals are constrained by the service curve β.

III. EXAMPLES

In this section we motivate and describe two examples of service definitions. Later in the paper, we compare the impact of these definitions on consumer perceived quality of service and on the amount of load smoothing that can be achieved by a distribution operator.

(a) Load Switching

It is convenient to introduce the function $b_s : [0, +\infty) \to [0, +\infty)$ defined by

$$\begin{cases} b_s \text{ is continuous and } b_s(0) = 0 \\ \frac{\partial b_s(x)}{\partial x} = 0 \text{ for } 0 < x < 1, \quad \frac{\partial b_s(x)}{\partial x} = 1 \text{ for } 1 < x < s \\ \frac{\partial b_s(x+s)}{\partial x} = \frac{\partial b_s(x)}{\partial x} \end{cases}$$

so that

$$b_s(x) \stackrel{\text{def}}{=} \max\left(x, \lfloor x/s \rfloor s + 1\right) - \left(\lfloor x/s \rfloor + 1\right) \qquad (3)$$

where $\lfloor x \rfloor$ is the floor function, i.e. the largest integer $\leq x$ (see Figure 1). Voltalis's bluepod can be expressed in our framework with the service curve

$$\beta_1(t) = t_0 z_{\max} b_{t_1/t_0}(t/t_0) \qquad (4)$$

with $t_0 = 30$ mn and $t_1 = 24$ hours (see Figure 2).

One may think of this service curve constraint as the possibility to impose switch off periods for a total duration of

[2] We assume $z(t)$ is positive, i.e. we focus on controlling demand only. Extension of this framework to consumers who are also producers is for further study.

30 mn every 24 hours, while otherwise energy may be drawn at a rate equal to the power $z_{\max}$. However, the service curve constraint allows for more general controls than on-off; for example, we might impose a control $u(t) = z_{\max}/2$ for 60 mn instead of $u(t) = 0$ for 30 mn; there are many combinations compatible with Eq. (2).

(b) Two-Level Load Control

The service curve β_1 in Eq. (4) and Figure 2 may be appropriate when the distribution system operator directly controls one or several appliances, but may not be so if we wish to leave such a control to the customer's smart home controller. For such cases, the control applies to the aggregate power drawn by one customer and it may be desirable to allow for a constant minimum power at any time (in order to serve non elastic appliances). Assume therefore that we want to define a service curve that (1) allows for power $z_{\max}$ except perhaps for t_0 hours every t_1 hours, but (2) also allows for a power $z_{\min}$ at any time. This corresponds to the following service curve:

$$\beta_2(t) = z_{\min}t + (z_{\max} - z_{\min})\, t_0 b_{t_1/t_0}\left(t/t_0\right) \qquad (5)$$

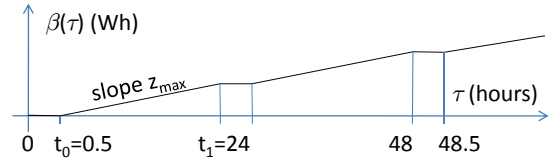(See Figure 3). Note that the service curve β_1 is a special case of β_2 with $z_{\min} = 0$.

Fig. 2. The service curve β_1, shown here, allows the distributor to switch off the load for at most 30 mn every day, or (for example) to reduce the load to $z_{\max}/2$ for 60 mn every day.
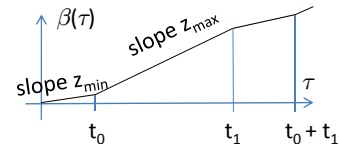

Fig. 3. The service curve β_2, shown here, corresponds to two-level load control, with minimum power $z_{\min}$ allowed at any time and maximum power $z_{\max}$ allowed at any time except perhaps for t_0 hours per period of t_1 hours.

IV. THE CALCULUS OF SERVICE CURVES

For using service contracts as in Section III, it is important to understand how to design and implement service curve constraints. Since service curves define a maximum value over any sliding window, they should be *super-additive* [3, Section 3.1.8], which means that $\beta(t + t') \geq \beta(t) + \beta(t')$ for any nonnegative t and t'.

More precisely, if we define a contract using a service curve β that is not super-additive, we obtain an equivalent definition if we replace β by its *super-additive closure* $\bar{\beta}$, which is

defined as the smallest super-additive function lower-bounded by β. Since $\bar{\beta} \geq \beta$ we have an apparently more stringent constraint [3, Theorem 3.1.10]. Note that the functions β_1 and β_2 introduced in Section III are super-additive, as is the function b_s (as we show in the proof of Lemma 1).

This has the following implication. If $\beta(t) = z_0 t$ for $0 \leq t \leq t_0$ for *some* positive t_0 and z_0 (i.e. we want to allow for some power z_0 over some period of time), then the super-additive closure satisfies $\bar{\beta}(t) \geq z_0 t$ for *all* $t \geq 0$ (i.e. we must guarantee z_0 over any period of time).

Also, assume that we want to design a service contract that several several constraints, such as

$$\int_{t'}^{t} u(\tau)d\tau \geq \beta(t - t')$$
$$\int_{t'}^{t} u(\tau)d\tau \geq \beta'(t - t')$$

for all $t' < t$. Then we must in fact impose a service curve equal to the super-additive closure of $\max(\beta, \beta')$, which, in general, is not the same as $\max(\beta, \beta')$. For example, we will see in Lemma 1 that the service curve β_2 in Figure 3 is equivalent to simultaneously imposing the two constraints $\beta_2'(t) = z_{\min} t$ and $\beta_2''(t) = t_0' z_{\max} b_{t_1/t_0'}(t/t_0')$ with $t_0' = (1 - z_{\min}/z_{\max}) t_0$.

In the rest of this section we use these results to propose practical ways to implement the service curves β_1 and β_2 introduced in Section III.

(a) Load Switching

The following theorem shows that the service curve constraint β_1 can be implemented by counting the control signals over sliding windows of duration t_1:

Theorem 1. *Let $u(t)$ be a sequence of control signals defined up to some time horizon T. Assume that $u(t) \leq z_{\max}$ for all $t \leq T$. The two properties are equivalent:*

(i) *$\int_{t'}^{t} u(\tau)d\tau \geq \beta_1(t - t')$ for all $t' < t \leq T$, where β_1 is the service curve in Eq. (4).*

(ii) *$\int_{t}^{t+t_1} u(\tau)d\tau \geq z_{\max}(t_1 - t_0)$ for all t such that $t + t_1 \leq T$.*

Proof: (i)$\Rightarrow$(ii) is immediate (take $t = t' + t_1$ and note that $\beta_1(t_1) = z_{\max}(t_1 - t_0)$).

(ii)$\Rightarrow$(i) First we show an alternative representation for β_1. Let $t \geq 0$ and let q be the quotient in the euclidian division of t by t_1, and r the remainder. In other words, $q = \lfloor \frac{t}{t_1} \rfloor$ and $r = t - qt_1$. We claim that

$$\beta_1(t) = q\beta_1(t_1) + z_{\max}(r - t_0)^+ \tag{6}$$

where the notation x^+ means $\max(x, 0)$. This can easily be shown by induction on q.

Second, for any nonnegative integer k, we claim that

$$\int_{t'}^{t'+kt_1} u(\tau)d\tau \geq k\beta_1(t_1) \tag{7}$$

To see why, note that

$$
\begin{aligned}
\int_{t'}^{t'+kt_1} u(\tau)d\tau &= \int_{t'}^{t'+t_1} u(\tau)d\tau + ... \\
&+ \int_{t'+(k-1)t_1}^{t'+kt_1} u(\tau)d\tau \\
&\geq k\beta_1(t_1) = \beta_1(kt_1)
\end{aligned}
$$

where the last equality is by Eq. (6).

Third, take some arbitrary $t' < t$ and let $q = \lfloor \frac{t-t'}{t_1} \rfloor$, $r = t - t' - qt_1$. Assume first that $t_0 \leq r$. Then:

$$
\begin{aligned}
\int_{t'}^{t} u(\tau)d\tau &= \int_{t'}^{t'+(q+1)t_1} u(\tau)d\tau - \int_{t}^{t'+(q+1)t_1} u(\tau)d\tau \\
&\geq (q+1)\beta_1(t_1) - z_{\max}(t_1 - r)
\end{aligned}
$$

where the former term is by Eq. (7) and the latter because $u(\tau)d\tau \leq z_{\max}$ by hypothesis. Recall that $\beta_1(t_1) = z_{\max}(t_1 - t_0)$, it follows that

$$
\begin{aligned}
\int_{t'}^{t} u(\tau)d\tau &\geq q\beta_1(t_1) + z_{\max}(t_1 - t_0) - z_{\max}(t_1 - r) \\
&= q\beta_1(t_1) + z_{\max}(r - t_0) = \beta_1(t - t')
\end{aligned}
$$

where we used Eq. (6) and the condition $t_0 \leq r$.

Consider now the case where $t_0 > r$. In this case $\beta_1(t - t') = q\beta_1(t_1)$. Further, $t \geq t' + qt_1$ thus

$$
\begin{aligned}
\int_{t'}^{t} u(\tau)d\tau &\geq \int_{t'}^{t'+qt_1} u(\tau)d\tau \\
&\geq q\beta_1(t_1) = \beta_1(t - t')
\end{aligned}
$$

∎

Theorem 1 gives a practical method for enforcing or verifying the service curve constraint β_1. It is sufficient to keep in memory the history of the control signals over the last t_1 time units; their integral should never drop below $z_{\max}(t_1 - t_0)$.

(b) Two-Level Load Control

We have a similar theorem for the service curve β_2. Recall that β_1 is a special case of β_2 with $z_{\min} = 0$, therefore Theorem 1 can be viewed as a consequence of Theorem 2. However the proof of Theorem 2 uses Theorem 1, which explains why we present the theorems in this order. We start with a lemma:

Lemma 1. *A sequence of control signals is constrained by the service curve β_2 in Eq. (5) if and only if it is simultaneously constrained by each of the two service curves defined by*

$$
\begin{aligned}
\beta_2'(t) &= z_{\min} t \\
\beta_2''(t) &= t_0' z_{\max} b_{t_1/t_0'}(t/t_0') \text{ with } t_0' = (1 - z_{\min}/z_{\max}) t_0
\end{aligned}
$$

Proof: (i) We show that b_s is super-additive. Note that $b_s(t) = q(s-1) + (r-1)^+$ as in Eq. (6) where q is the quotient in the euclidean division of t by s, and r the remainder. Thus for any nonnegative integer k:

$$b_s(t + ks) = k(s-1) + b_s(t)$$

Thus for any any nonnegative integers k, k':

$$b_s(t + t' + (k + k')s) - b_s(t + ks) - b_s(t + k's) =$$
$$b_s(t + t') - b_s(t) - b_s(t')$$

and it is sufficient to show

$$b_s(t + t') \geq b_s(t) + b_s(t') \text{ for } 0 \leq t < s \text{ and } 0 \leq t' < s \quad (8)$$

This is easily done by considering the four cases obtained by testing $t < 1$ and $t' < 1$ (with $0 \leq t < s$ and $0 \leq t' < s$).

(ii) We show that β_2 is super-additive. This follows directly from the representation

$$\beta_2(t) = z_{\min}t + t_0(z_{\max} - z_{\min})b_{t_1/t_0}(t/t_0)$$

and from (i).

(iii) We show that β_2 is equal to the super-additive closure $\bar{\beta}'$ of $\beta' = \max(\beta_2', \beta_2'')$. Note that $\beta_2 \geq \beta_2'$ and $\beta_2 \geq \beta_2''$ thus $\beta_2 \geq \beta'$ and since β_2 is super-additive, $\beta_2 \geq \bar{\beta}'$. Now we show that $\beta_2 \leq \bar{\beta}'$.

Note that β_2' and β_2'' are super-additive (this follows from (i)), thus [3, Theorem 3.1.11] the super-additive closure of β' is $\beta_2'\bar{\otimes}\beta_2''$, where $\bar{\otimes}$ denotes the max-plus convolution:

$$(\beta_2'\bar{\otimes}\beta_2'')(t) = \sup_{t' \leq t} (\beta_2'(t') + \beta_2''(t - t'))$$

Therefore we need to show that for all nonnegative t there exists some nonnegative $t' \leq t$ such that

$$\beta_2(t) \leq \beta_2'(t') + \beta_2''(t - t') = z_{\min}t' + \beta_2''(t - t') \quad (9)$$

Let q be the quotient in the euclidean division of t by t_1, and r the remainder. If $r \geq t_0$ then $\beta_2(t) = \beta_2''(t)$ and Eq. (9) is satisfied with $t' = 0$. Else, $\beta_2(t) = \beta_2''(qt_1) + z_{\min}r$ and Eq. (9) is satisfied with $t' = r$.

(iv) We are now in a position to prove the statement in the lemma. Note that $\beta_2 \geq \beta_2'$ and $\beta_2 \geq \beta_2''$ therefore the "only if" part of the theorem is obvious. Assume now that a sequence of control signals $u(\tau)$ is constrained by the service curves β_2' and β_2''. It is therefore constrained by $\beta' = \max(\beta_2', \beta_2'')$ and, therefore by the super-additive closure of β', which by (iii) is equal to β_2. ■

Theorem 2. *Let $u(t)$ be a sequence of control signals defined up to some time horizon T. Assume that $u(t) \leq z_{\max}$ for all $t \leq T$. The two properties are equivalent:*

(i) *$\int_{t'}^t u(\tau)d\tau \geq \beta_2(t - t')$ for all $t' < t \leq T$, where β_2 is the service curve in Eq. (5).*

(ii) *$u(\tau) \geq z_{\min}$ for all $\tau \leq T$,*
    and $\int_t^{t+t_1} u(\tau)d\tau \geq z_{\max}t_1 - (z_{\max} - z_{\min})t_0$ for all t such that $t + t_1 \leq T$.

Proof: (i)⇒(ii) is immediate as in the proof of Theorem 1, after noticing that $\beta_2(t_1) = z_{\max}t_1 - (z_{\max} - z_{\min})t_0$.

(ii)⇒(i) By Lemma 1, it is sufficient to show that the constraints with service curves β_2' and β_2'' are both satisfied. The constraint with service curve β_2' is obviously satisfied because $u(\tau) \geq z_{\min}$ for every τ. For the second constraint, note that β_2'' is the same as β_1 if we replace t_0 by $t_0' = (1 - z_{\min}/z_{\max}) t_0$ and that $z_{\max}t_1 - (z_{\max} - z_{\min})t_0 = z_{\max}(t_1 - t_0')$. Then apply Theorem 1. ■

Theorem 2 gives a practical method for enforcing or verifying the service curve constraint β_2. It is sufficient to check that the control signals satisfy $u(\tau) \geq z_{\min}$ at every time slot τ and to keep in memory the history of the control signals over the last t_1 time units; their integral should never drop below $z_{\max}t_1 - (z_{\max} - z_{\min})t_0$. Note that the constraint $u(\tau) \leq z_{\max}$ is always implicitly enforced, since we assume that the consumed power can never exceed $z_{\max}$.

V. CONSUMER SIDE PROBLEM

In this section we illustrate how a consumer may optimize her use of a service contract with service curve. Assume a consumer wants to recharge her PEV battery, using a power plug controlled by a smart home controller, with a contract as defined above, with service curve β_2. Assume the minimum guaranteed power $z_{\min}$ can be used at any time by other, non elastic appliances. Assume the PEV is the only elastic appliance (the more general problem involves scheduling of multiple elastic appliances by the smart home controller and is left for further study). Assume the consumer is interested in having a fully charged battery at time horizon T, given that at time 0 the current charge level is $B(0)$. Since the contract specifies a fixed price per unit of energy, she is interested in minimizing the consumed power, subject to feasibility. In other words, she wants to have a battery fully loaded at time T; if this is possible, she further wants to minimize the electricity cost. Let $B(t)$ be the battery level at time t, and $B_{\max}$ the full battery level. We can express the consumer's problem as the desire to minimize the cost function

$$K(B_{\max} - B(T)) + \int_0^T z(\tau)d\tau \quad (10)$$

where K is a large constant ($K > Tz_{\max}$). The equations for battery level are

$$B(t) - B(0) = \int_0^t \left(\eta z(\tau) - \rho z(\tau)^2 - \gamma\mathbb{1}_{\{B(\tau)>0\}} \right) d\tau \quad (11)$$

where the notation $\mathbb{1}_{\{condition\}}$ means 1 when *condition* is true and 0 otherwise, and

$$\eta = \text{efficiency}$$
$$\gamma = \text{battery leakage}$$
$$\rho = \text{thermal loss}$$

Further, the load $z(t)$ must be such that $B(t) \leq B_{\max}$ at all times t.

The consumer's problem must be solved online; to make the problem tractable, we discretize time. The consumer's smart home controller needs to decide at time t for a value $z(t+1)$, not knowing values of the future control signals $u(\tau)$, $\tau > t$, using a "policy" π. The policy must be causal, i.e. decision taken by π at time t depends on $z(\tau), u(\tau)$ for $\tau \leq t$ and on the battery level $B(t)$. Further, the service curve contract must be obeyed, i.e. the policy must choose a load $z(t+1)$ such that

$$z(t+1) + n(t+1) \leq u(t+1)$$

where $n(t+1)$ is the non elastic load at time $t+1$. In this paper, we assume that the non-elastic load cannot be forecast

or controlled, however it may not exceed $z_{\min}$. We assume therefore that the policy enforces the constraint

$$z(t+1) \leq u(t+1) - z_{\min}$$

which indeed allows non-elastic load to operate without hindrance.

The future control signals $u(t+1)$ are not known either at the time of decision making, but their worst case can be predicted, depending on the past. We also assume that the consumer's goal is to minimize the worst case total cost. In other words, let $C(\pi, u)$ be the total cost, given by Eq. (10) (u denotes the complete control signal for all times up to time T). The goal is to find a policy π that minimizes

$$\max_{u} C(\pi, u)$$

where the maximum is over all possible control signals that are compatible with the service curve constraint.

Finding an optimal policy seems to be an open problem and is left for further study. However we can provide some preliminary insight. The first question that can be addressed is whether a close to full charge can be achieved within the time horizon T. Note that a full charge cannot be guaranteed, since control signals sent towards the end of the charging period may stop the charging process, and some battery leakage is assumed in our model. Also note that the thermal loss term in Eq. (11) may imply that for large enough load, e.g., for $z(\tau) > \eta/\rho$, the net charge decreases; clearly, the battery charging system should not operate in this region, it is sufficient to assume that $z_{\max} - z_{\min} < \eta/2\rho$, with batteries such that $\eta^2 > 4\gamma\rho$. We also assume that the horizon T is less than t_1 (this is reasonable if we assume, as we did earlier, that t_1 is of the order of 24 hours).

Theorem 3. *Consider the consumer side problem described earlier. Let $B(0) < B_{\max}$ be the initial battery level and $u(t)$, $t = 0, -1, -2, ...$ the past control signals (known to the consumer).*

Let V^* be the optimal value of the convex programming problem:

minimize $V = \sum_{\tau=1}^{T} \left(\eta x(\tau) - \rho x(\tau)^2 \right)$
over the variables $x(1), ..., x(T)$
subject to the constraints
$$\begin{cases} 0 \leq x(\tau) \leq z_{\max} - z_{\min} \text{ for } 1 \leq \tau \leq T, \\ \sum_{\tau=t+1}^{t+t_1} x(\tau) \geq (z_{\max} - z_{\min})(t_1 - t_0) \\ \qquad \text{for } 1 - t_1 \leq t \leq T - t_1 \end{cases}$$

If $V^* \geq B_{\max} - B(0) + T\gamma$, there exists a policy π that achieves $B(T) \geq B_{\max} - \gamma T$. This policy achieves full battery level at some time $t \in [0, T]$.

Proof: Let $u(t)$ be the actual value of the controls (not known in advance). Let $x(t) = u(t) - z_{\min}$ and let $B'(t)$ be the level of a virtual battery, that would start at $B(0)$, would receive the load $x(t)$, would have 0 leakage, and infinite maximum level. In other words, $B'(0) = B(0)$ and for $t = 0, ..., T - 1$:

$$B'(t+1) - B'(t) = \eta x(t) - \rho x(t)^2$$

The sequence $x(t)$ satisfies the constraints in the theorem, therefore

$$\begin{aligned} B'(T) - B'(0) &= \sum_{\tau=1}^{T} \left(\eta x(\tau) - \rho x(\tau)^2 \right) \\ &\geq V^* \geq B_{\max} - B(0) + T\gamma \end{aligned}$$

and thus

$$B'(T) \geq B_{\max} + T\gamma \tag{12}$$

Now define the policy π by

$$z(t+1) = u(t+1) - z_{\min} \text{ if } B(t) \leq B_{\max}$$
$$z(t+1) = \min\left(u(t+1), \gamma\right) \text{ if } B(t) = B_{\max}$$

i.e. we set the consumed power $z(t)$ to the maximum possible considering the constraint of non elastic load and of maximum battery capacity. Let $B(t)$ be the battery level with this policy. As long as $B(t) < B_{\max}$, the rate at which B and B' vary differ by at most the leakage term γ, thus as long as $B(t) < B_{\max}$:

$$B'(t) - \gamma t \leq B(t) \leq B'(t) \tag{13}$$

Assume that $B(t)$ never reaches $B_{\max}$. It follows that $B'(T) \leq B(T) + \gamma T < B_{\max} + \gamma T$, which contradicts Eq. (12), thus there exists some time t at which $B(t) = B_{\max}$. The rest of the theorem follows from the fact that leakage occurs at rate γ. ∎

VI. A MINIMUM VARIANCE POLICY (MCAP)

In this and the next section we consider two scenarios where DSOs could make use of the service curve approach. In this section, we consider a case where load is random but stationary, and we study how the variance of the load can be reduced by using appropriate control signals. Reducing the variance reduces DSOs' hedging costs. The scenario considered here is highly idealized scenario, but it serves the purpose of showing that simple binary policies are not always optimal.

Consider an infinite population of identical customers having contracts specified by the service curve $\beta_2(\cdot)$. Assume that customers arrive according to a Poisson process of constant rate λ with arrival times $\{S(n), n \in \mathbb{Z}\}$, consume an amount $\beta_2(T)$ of energy, for a fixed T satisfying $t_0 < T < t_1$, and subsequently leave the system forever. The quantity of energy consumed by each customer can be written as

$$w := \beta_2(T) = z_{\max}T - (z_{\max} - z_{\min})t_0.$$

The system operator chooses signal $u(s)$ satisfying the contract agreement

$$\int_{t'}^{t} u(s)\, ds \geq \beta_2(t - t'), \ \forall t' < t,$$

that it sends to each customer upon arrival. Customers behave greedily and immediately consume their w units of energy, as soon as the control signals allow them to do so. Denote the instantaneous power consumption of a customer at t time units after her arrival in the system by
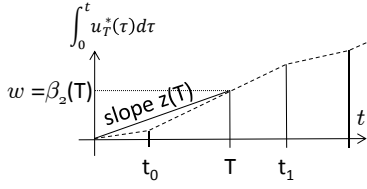
$$z(t) := u(t)\mathbb{1}_{0 \leq t \leq T}.$$

Fig. 4. Cumulative control $\int_0^t u_T^*(\tau)\,d\tau$ for the optimal policy u_T^*. The time $t = 0$ is a customer arrival time.

Assume that the DSO knows w. If $\int_0^T z(s)\,ds < w$, then the contract is violated. On the other hand, if $\int_0^T z(s)\,ds > w$, then the DSO exceeds its obligations. Hence, the DSO wishes to have

$$\int_0^T z(s)\,ds = \beta_2(T). \tag{14}$$

The aggregate instantaneous energy consumption observed by the DSO is given by the stochastic process

$$Z(t) = \sum_{n=-\infty}^{+\infty} z(t - S(n)).$$

By Campbell's theorem, the mean and variance of $Z(t)$ are given by

$$\mathbb{E}Z = \lambda \int_{-\infty}^{+\infty} z(s)\,ds = \lambda w,$$

$$\mathrm{Var}(Z) = \lambda \int_{-\infty}^{+\infty} z^2(s)\,ds.$$

Since the users have the same energy consumption, the DSO wishes to minimize the variance of the process $Z(t)$, i.e., to solve the following problem:

$$\text{Minimize } \int_{-\infty}^{+\infty} z^2(s)\,ds \text{ over } u(\cdot) \leq z_{\max} \tag{15}$$

$$\text{such that } \int_{t'}^t u(s)\,ds \geq \beta_2(t - t'), \ \forall 0 \leq t' < t \leq T. \tag{16}$$

A first choice is to take binary control $u(\cdot)$ equal to

$$u_b(t) = \begin{cases} z_{\min}, & 0 < t \leq t_0 \\ z_{\max}, & \text{otherwise}, \end{cases}$$

a choice which, by Theorem 2, satisfies the constraint (16). By Jensen's inequality, we can write

$$\int_0^T u_b^2(s)\,ds = t_0 z_{\min}^2 + (T - t_0) z_{\max}^2$$

$$\geq T\left(\frac{t_0}{T} z_{\min} + \frac{T - t_0}{T} z_{\max}\right)^2. \tag{17}$$

Denote

$$z(T) := \frac{t_0}{T} z_{\min} + \frac{T - t_0}{T} z_{\max}, \tag{18}$$

and define the policy

$$u_T^*(t) = \begin{cases} z(T), & 0 < t \leq T \\ z_{\max}, & \text{otherwise}. \end{cases} \tag{19}$$

By Theorem 2, u_T^* also satisfies constraint (16) and further-more achieves a lower variance, by (17). We shall refer to this policy as the Minimum Constant Allocation Policy (MCAP). Let us prove that MCAP is optimal.

Theorem 4. *For identical customers with contract given by $\beta_2(\cdot)$ arriving according to a Poisson process of rate λ and consuming the same amount of energy $w = \beta_2(T)$ for a fixed value $t_0 < T < t_1$, the MCAP control signal u_T^* minimizes the variance of the aggregate demand $Z(t)$. In other words, it achieves the optimum of* (15)-(16).

Proof: For any control u, by Cauchy-Schwartz,

$$\int_0^T u^2(s)\,ds \int_0^T 1\,ds \geq \left(\int_0^T u(s)\,ds\right)^2.$$

Thus, using (14), we obtain a lower bound on the variance of $Z(t)$:

$$\int_{-\infty}^{+\infty} z^2(s)\,ds \geq \frac{1}{T}\beta_2^2(T).$$

Since

$$\int_0^T u_T^{*\,2}(s)\,ds = \frac{1}{T}\beta_2^2(T),$$

we can conclude. ∎

Service curves give freedom of control choice for the DSO. As discussed in Section III, service curves such as $\beta_2(\cdot)$ allow the implementation of binary controls u_b (which are similar to the ones implemented by Voltalis), but also of more complex ones. In the setting of steady arrival of identical users, we saw (Theorem 4) that binary controls are not the best suited for reducing peak consumption, and found that the smoother MCAP control u_T^* achieves the goal, while satisfying the contract conditions.

In the next section we evaluate a heuristic for performing demand response using service curves in a setting of bursty arrivals.

VII. BURSTY ARRIVALS

Like in Section VI, we consider consumers that have a guaranteed minimal service characterized by the service curve $\beta_2(\cdot)$. We have seen that, when arrivals are static and consumers homogeneous, the best policy for the system operator is to use MCAP, which imposes a constant power consumption for users. In this section, we define the Quota Binary Allocation Policy (QBAP), a heuristic which distributes temporary service reductions to the users of the system through binary power signals, while satisfying the contract specified via the service curve. We consider a setting in which arrivals are bursty. We show through numerical evaluations that QBAP outperforms MCAP in this scenario.

A. Assumptions

We consider a population of N users $\mathcal{U} = \{1, \ldots, N\}$ with contracts specified by service curve $\beta_2(\cdot)$. At time t each user $i \in \mathcal{U}$ receives control signal $u_i(t)$ from the DSO and consumes power $z_i(t)$, such that $z_i(t) \leq u_i(t)$.

We address a specific scenario where the N consumers (or users) charge their PEV batteries upon arriving home after a work day. The arrival times are normally distributed and the

car batteries are identical, ideal (i.e., no thermal losses, or leakage), have a maximum charging power z_b and capacity w. Like in Section V, consumers additionally impose a non-elastic load on the grid, which may not exceed $z_{\min}$, and the PEV is their only elastic appliance. We set $t_1 = 24h$ and $z_{\max} = z_{\min} + z_b$, such that it is possible to satisfy the non-elastic load and to charge the battery at maximum power for a duration of at least $t_1 - t_0$ per day. The burst of consumption due to batteries causes the load on the grid to peak. We wish to design a policy which makes use of the freedom of decision provided by the service curve to reduce the peak load.

For presentation ease, in this section we consider discrete time. We note that the results presented in the previous sections directly apply for discrete time.

B. Quota Binary Allocation Policy

Theorem 2 states that in order to satisfy the constraints imposed by the service curve $\beta_2(\cdot)$, it is necessary and sufficient to guarantee an energy output of at least $z_{\max}(t_1 - t_0) + z_{\min}t_0$ over any time window of length t_1, while at any time t, the control signal $u_i(t)$ sent to each user i is such that $z_{\min} \le u_i(t) \le z_{\max}$. We use this result to design a technique that enforces the service curve.

For each user i, fix $u_i(t)$ for the duration of time step t, and allow it to take only two possible values $\{z'_{\min}, z_{\max}\}$, where $z'_{\min} = z_{\min} + \epsilon$, for a fixed $\epsilon > 0$. We call $u_i(t)$ a discrete *binary* policy. We introduce ϵ for a signaling purpose: Since the non-elastic load is limited at $z_{\min}$, power consumption greater than $z_{\min}$ is an indication for the DSO that elastic load is present. Let us make sure that the control $u_i(t)$ satisfies the contract. By Theorem 2, it is sufficient to guarantee that $u_i(t)$ will be set to $z'_{\min}$ over no more than t_0 time steps within any time window of length t_1.

Setting $u_i(t)$ to $z'_{\min}$ rather than $z_{\max}$ can be interpreted as a "penalty". Define a *penalty budget sequence* $(p_{i,t}(t'))_{t' \ge 0}$ for each user i and at each time step t. The values $p_{i,t}(t')$ represent the number of successive time steps, starting at time $t + t'$, during which user i can be limited at $z_{\min}$ without violation of the contract, if no low-power signals are sent between times t and $t + t'$. For example, if $p_{i,t}(1) = 2$, then the control sequence $u_i(t..t+3) = (z_{\max}, z'_{\min}, z'_{\min}, z_{\max})$ is feasible, while the control sequence $u_i(t..t+3) = (z_{\max}, z'_{\min}, z'_{\min}, z'_{\min})$ violates the penalty budget.

Implicitly, if $p_{i,t}(0) \ge 1$, then the control $u_i(t)$ may take any of the values $z'_{\min}$ or $z_{\max}$, whereas if $p_{i,t}(0) = 0$, then $u_i(t)$ must be set to $z_{\max}$. In the latter case, we say that the penalty budget of user i is depleted.

Let us describe how the penalty budget sequence can be computed. By definition, $p_{i,t}(t') = t_0$, for all $t' \ge t_1$. Therefore, the operator needs to maintain the penalty budget over a window of only t_1 time steps in the future: $(p_{i,t}(0), p_{i,t}(1), \ldots, p_{i,t}(t_1 - 1))$.

At $t = 0$ (i.e., no consumption history), for each user i, we have that $p_{i,t}(t') = t_0$, for all $t' \ge 0$. At time t, the DSO can perform two operations: either send a maximum power signal $z_{\max}$ to user i, or, if the penalty budget allows

it (i.e., $p_{i,t}(0) \ge 1$), a low-power signal $z'_{\min}$. The operation performed by the DSO modifies the penalty budget sequence as follows:

- If $u_i(t) = z_{\max}$, then the penalty budget is not affected. Hence, at time step $t + 1$ we have

$$p_{i,t+1}(t') = \begin{cases} p_{i,t}(t' + 1), & 0 \le t' < t_1 - 1, \\ t_0, & t' \ge t_1 - 1. \end{cases}$$

- If $p_{i,t}(0) \ge 1$, then the DSO may send a low-power signal to user i, $u_i(t) = z'_{\min}$. In this case, the penalty budget is decremented and at time step $t + 1$ we have:

$$p_{i,t+1}(t') = \begin{cases} p_{i,t}(t' + 1) - 1, & 0 \le t' < t_1 - 1, \\ t_0, & t' \ge t_1 - 1. \end{cases}$$

In Algorithm 1 we propose a heuristic for reducing aggregate peak consumption, the Quota Binary Allocation Policy (QBAP). The DSO enforces the service curve using a sliding window of penalties P_i of length t_1 for each user i, as described above. The algorithm takes as input a fixed "best-effort" *quota* $N_a > 0$. At each time step, it aims for at most N_a users consuming $z_{\max}$. We pick these among the most deprived users in terms of penalty budget. Note that it might not always be possible to satisfy the quota, since users with depleted penalty budgets *must* be served at $z_{\max}$. An adaptive quota is left for future study. For now, we simply wish to show that there exist values of the quota N_a^*, such that QBAP(N_a^*) outperforms MCAP.

Algorithm 1 Quota Binary Allocation Policy

QBAP($N_a, \epsilon, T_{\max}$)
Penalty budget windows (t_1-dimensional): $P_i(0..t_1-1) \; \forall i \in \mathcal{U}$

1: Initialise $P_i := (t_0, \ldots, t_0)$, for all $i \in \mathcal{U}$
2: $t := 1$
3: **for** $t = 1$ **to** $T_{\max}$ **do**
4:      Retreive consumption at time $t - 1$: $\{z_i(t-1)\}_{i \in \mathcal{U}}$. (by default, $z_i(0) = 0$)
5:      Find users i that exceeded $z_{\min}$ at time $t - 1$: $\mathcal{A} := \{i \in \mathcal{U} : z_i(t-1) > z_{\min}\}$.
6:      Sort users in $\mathcal{A}$ increasingly according to their penalty budget $P_i(0)$: $\pi : \mathcal{A} \to \{1, \ldots, |\mathcal{A}|\}$, $P_i(0) \le P_j(0) \iff \pi(i) \le \pi(j), \; \forall i, j \in \mathcal{A}$.
7:      **for** $i = 1$ **to** N **do**
8:          **if** $i \in \mathcal{A}$ **and** $\pi(i) > N_a$ **and** $P_i(0) \ge 1$ **then**
9:              $u_i := z_{\min} + \epsilon$
10:              Decrement penalty budget sequence: $P_i := P_i - 1$
11:          **else**
12:              $u_i := z_{\max}$
13:          **end if**
14:          **for** $t' = 0$ **to** $t_1 - 2$ **do**
15:              Slide window: $P_i(t') := P_i(t' + 1)$
16:          **end for**
17:          $P_i(t_1 - 1) := t_0$.
18:      **end for**
19:      Send signals $\{u_i\}_{i \in \mathcal{U}}$ to the consumers for time step t
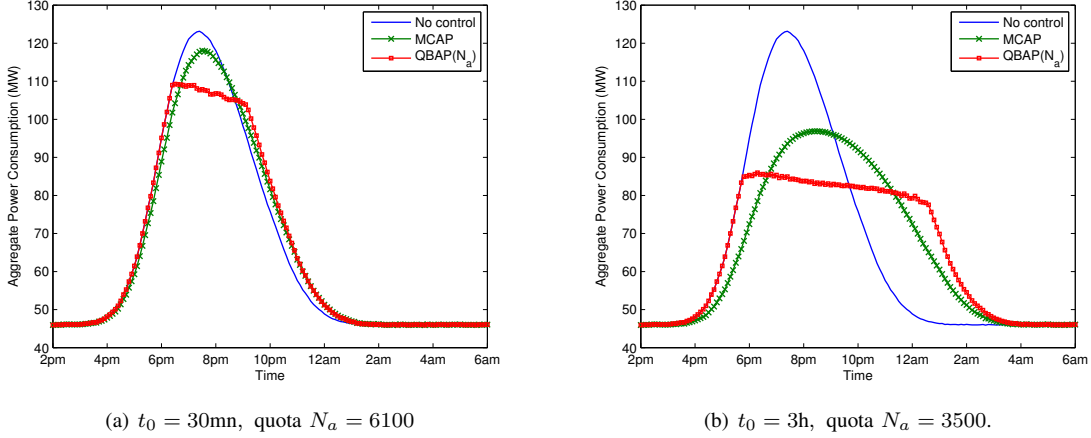20: **end for**

(a) $t_0 = 30$mn, quota $N_a = 6100$

(b) $t_0 = 3$h, quota $N_a = 3500$.

Fig. 5. The peak-shaving effect of MCAP (Section VI) and QBAP (Section VII). $N = 10000$ users, $w = 52$kWh, $z_{max} = 14.6$kW, $z_{min} = 5$kW, $\epsilon = 100$W, time step $= 6$mn.

C. Numerical Evaluation

Customers are guaranteed minimum power of $z_{min} = 5$kW. The non-elastic appliances may consume at most z_{min} power (e.g., physically limited by a fuse box). Non-elastic consumption level for user i at time t is given by independent normally distributed random variables $X_{i,t}$ of mean z_{min} and standard deviation 1 capped at z_{min}: $\min\{z_{min}, X_{i,t}^+\}$.

The N consumers start charging their batteries at random times normally distributed, centered at 6pm, with standard deviation of 1 hour. We consider identical PEV batteries of capacity $w = 52$kWh, with charge levels $(w_i)_{i \in \mathcal{U}}$ uniformly distributed between 0 and $4/5$ of the total battery capacity w. We consider ideal batteries, without losses or charging inefficiencies. The batteries charge at the power imposed by the control signals $z_{min} + \epsilon$ and z_{max} (for $\epsilon = 100W$) without hindering the non-elastic appliances. The maximum power at which the batteries can be charged is $z_b = 9.6$kW. The overall consumption is capped at $z_{max} = z_{min} + z_b = 14.6$kW.

Under a service curve $\beta_2(\cdot)$, battery i may take up to T_i time steps to recharge, where $\beta_2(T_i) = w_i + T_i z_{min}$. A back-of-the-envelope calculation gives $T_i = \frac{w_i}{z_{max} - z_{min}} + t_0$.

In Figure 5 we show the aggregate consumption for two values of t_0: a Voltalis-like setting $t_0 = 30$mn (Figure 5(a)) and a less constraining value $t_0 = 3$h (Figure 5(b)). We consider two possible controls. In the first case, the PEVs transmit their charge levels w_i to the DSO, which in turn implements the smooth MCAP(T_i) customized for each consumer $i \in \mathcal{U}$. This control gives an aggregated consumption shown by the curves labeled "MCAP". In the second case, the DSO uses QBAP(N_a) for the specified versions of the quota N_a. This control gives an aggregated consumption shown by the curves labeled "QBAP(N_a)". For comparison we also plot the aggregate demand when power consumption per user is capped at z_{max} and no control is used.

We notice that in the considered settings QBAP outperforms MCAP. In the Voltalis-like setting (Figure 5(a)) QBAP reduces peak consumption by 11.3% (vs. 4.1% for MCAP). For $t_0 = 3$h (Figure 5(b)), QBAP reduces peak consumption by 30.2% (vs. 21.3% for MCAP). Note that the overall energy output

(i.e., the integral over time of the consumed power) is the same for the three policies. Moreover, for $t_0 = 3$h we obtain significant reductions in peak load, while PEV batteries are fully charged by 3am.

VIII. CONCLUSION

We have introduced the concept of service curves for electricity consumption, inspired from load smoothing mechanisms proposed for the Internet. Service curves provide the systems operator with the ability to alleviate the load on the grid at critical peak-hours through demand-response mechanisms. Such mechanisms come at the cost of small discomfort on the consumer side in the form of service reductions over limited periods of time. This approach has the advantage of decoupling the consumer and provider problems, while avoiding to expose consumers to electricity price volatility. We have formally characterized such service curves using max-plus calculus and provided efficient methods to provably implement the constraints imposed thereby. We have shown that on the consumer side, discomfort can be reduced by employing simple policies which take into consideration the parameters of the service curves. Finally, we have provided simple heuristics that can be implemented by the DSO in order to reduce peak load. In the setting of bursty arrivals of heterogeneous consumers, we have shown that QBAP, a heuristic which controls the load via binary power signals, outperforms MCAP, which sends smooth signals and is provably optimal in a stationary setting.

REFERENCES

[1] S. Borenstein, M. Jaske, and A. Rosenfeld. Dynamic pricing, advanced metering, and demand response in electricity markets. 2002.
[2] S. Keshav and C. Rosenberg. Direct adaptive control of electricity demand. 2010.
[3] J.-Y. Le Boudec and P. Thiran. *Network Calculus*. Springer Verlag Lecture Notes in Computer Science volume 2050, July 2001.
[4] S. Meyn, M. Negrete-Pincetic, G. Wang, A. Kowli, and E. Shafieepoor-fard. The value of volatile resources in electricity markets. In *Proc. of the 49th Conf. on Dec. and Control*, 2010.
[5] OECD. *OECD Economic Surveys: France 2011*. OECDiLibrary, March 2011.