Scalable and Privacy-Preserving Admission Control for Smart Grids

Giovanni Neglia [1], Giuseppe Di Bella [2], Laura Giarré [3], Ilenia Tinnirello [3]

***Abstract*— Energy demand and production need to be constantly matched in the power grid. The traditional paradigm to continuously adapt the production to the demand is challenged by the increasing penetration of more variable and less predictable energy sources, like solar photovoltaics and wind power. An alternative approach is the so called *direct control* of some inherently flexible electric loads to shape the demand. Direct control of deferrable loads presents analogies with flow admission control in telecommunication networks: a request for network resources (bandwidth or energy) can be delayed on the basis of the current network status in order to guarantee some performance metrics. In this paper we go beyond such an analogy, showing that usual teletraffic tools can be effectively used to control energy loads. In particular we propose a family of control schemes which can be easily tuned to achieve the desired trade-off among resources usage, control overhead and privacy leakage.**

I. Introduction

Direct Load Control (DLC) allows energy utilities to control electric loads at the customers' premises. In the past DLC was used in critical situations to prevent blackouts by shutting down these loads. More recently, an extensive use of DLC has been advocated as a way to shape energy demand peaks or provide other ancillary services, often by controlling thermostatic loads [1], [2], such as air conditioners and heating systems, because they allow a fine-tuning regulation of power demand. Alternative approaches consider electric vehicles or other battery-empowered appliances [3], which, beside acting as adaptive loads, can also reinject energy in the grid. Multiple load typologies, including interruptile or deferrable loads such as pool pumps, have been also considered for responding to different frequency components of the regulation signal [4].

In this paper we consider a scenario where DLC functionalities are deployed at a large set of small deferrable energy loads, like appliances at residential users and we propose simple control approaches rooted in teletraffic engineering which intrinsically work better for large scale systems. Indeed, loads on the electrical grid are multiplexed at different aggregation levels (distribution transformers, primary station, prize zone), similarly to traffic from data sources multiplexed at routers of different hierarchical levels. In particular, we propose a simple control mechanism that guarantees that the instantaneous power demand exceeds a given bound with probability smaller than ϵ. Only a stochastic characterization of the power demand for each class of appliances is required. This mechanism combines two different operation paradigms. In the first one, appliances need to ask a controller the permission to start, and the controller will limit the number of simultaneously active appliances to $n(t)$. In the second one, an activation probability function—$p(t)$—is broadcast periodically to all the appliances; appliances do not notify the controller but they start with probability $p(t)$ and postpone their decision to the time $t+T$ with probability $1-p(t)$. The first operation paradigm requires more communication exchanges between the appliances and the controller and reveals more information about the customers' habits. The second paradigm works in an open-loop fashion and, then, does not disclose any private information. At the same time the lack of an exact knowledge of the current number of active appliances causes a lower average utilization of the resources in order to satisfy the constraint. We simply combine the two paradigms by means of a probability q: when an appliance wants to start operating, it will ask the permission to the controller with probability q, and it will decide autonomously using the function $p(t)$ with probability $1-q$. The parameter q can then be chosen in order to achieve the wished trade-off among resources' usage, control overhead and privacy leakage. In fact, by increasing q we gradually i) reduce privacy by exposing more the energy profile of each user, ii) increase control overhead because the controller needs to directly interact with a larger number of appliances, iii) increase the efficiency. Fig. 1 qualitatively depicts the effect of the parameter q. Our analysis (Sec. III) and experiments (Sec. V) allow to quantify these trade-offs.

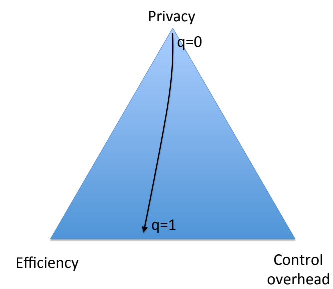

Fig. 1. Different trade-offs among efficient resources' usage, control overhead and privacy leakage achievable by tuning the parameter q.

The detailed description of our system is in Sec. II. In our mechanism, the control policy is determined by the two functions $p(t)$ and $n(t)$. In Sec. III, we show how techniques developed for flow admission control in packet networks can be used to determine a stationary control policy (i.e. $p(t)=p$ and $n(t)=n$), when the appliances' activation rate is assumed to be time-invariant. Our paper contributes then to

[1] G. Neglia is with Inria Sophia-Antipolis Méditerranée, Email: giovanni.neglia@inria.fr.

[2] G. Di Bella is with Swarm Joint Open Lab Telecom Italia, Email: giuseppe.dibella@telecomitalia.it. Part of this work has been carried out while he was a PhD student at University of Palermo.

[3] L. Giarré and I. Tinnirello are with Università degli Studi di Palermo, Email: {firstname.lastname}@unipa.it.

show how teletraffic engineering tools can be advantageously used also in the context of future smart power grids. In Sec. IV, we derive time-variant control policies for the more realistic case when the appliance activation rate i) is time-varying and ii) needs to be estimated. Some results are in the companion technical report [5].

Our main goal in the present paper is not to design a full-fledged solution for DLC, but rather to show how teletraffic tools (such as those used for Call Admission Control in telecommunication networks) can be effectively used to achieve a large scale deployment of DLC to shape energy demand peaks. The use of teletraffic tools is not completely novel. Network calculus has been recently exploited to size energy batteries [6] and transformers [7]. Queuing theory is also used in [8] for sizing the population of customers subscribing a DLC program under a given maximum tolerable delay for activating the controlled appliances.

II. System description

We consider the problem of peak shaving for an energy supplier, which wants to reduce its customers' consumption during the time of the day when energy costs are higher. To achieve these goals, the energy suppliers may interact with the distribution system operators (DSO), and/or with novel intermediate figures called load aggregators. Our solution can operate in both the scenarios, so we will talk generically about the load controller to denote the entity which drives the appliances.

The energy supplier specifies a high-level constraint for the controller in terms of the maximum tolerable probability to exceed a given power demand in the controlled area. The load controller is responsible to meet this requirement by deciding about the activation of deferrable loads at households, such as dish-washers and laundry-machines. On the basis of the demand prediction, the requirements are mapped into a control signal that is broadcast to all the controlled households for deciding about the admission or the deferral of load activation requests. Each household is equipped with a gateway able to receive the control signal from the controller and to interact with the domestic appliances by means of local area technologies (such as WiFi, ZigBee or PLC). Smart appliances can be natively equipped with a programmable interface able to communicate with the gateway, while dumb appliances can be controlled by means of smart plugs to be inserted between the appliances' plugs and the power sockets.

In our control system, the signal broadcast by the load controller is the probability $p(t)$ to accept or defer a load activation request at time t (a similar control mechanism was considered in [9]). At each household, the gateway can autonomously process a novel activation request from an appliance on the basis of this function: the activation request is accepted with probability $p(t)$ and deferred of a fixed time interval T with probability $1 - p(t)$. This operation applies to a fraction $1-q$ of all the requests. The others are directly forwarded to the load controller and the gateway waits for an ACK from it in order to accept the request. The controller

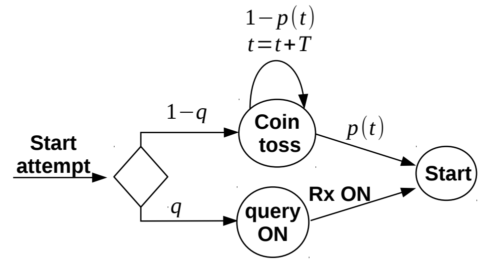

Fig. 2. Household control model in terms of activation probability $p(t)$ and ratio of forwarded queries q.

maintains a cap $n(t)$ on the number of appliances it allows to be active at a given time instant. The ACK can then be delayed until some previously activated appliance does not stop operating. This mechanism allows the load controller to actually estimate the time-varying power demand, but also to have a tighter control on the aggregate power demand and then to achieve a more effective resource utilization in the controlled area.

Fig. 2 summarizes the actuator model at the household: different control operations can be programmed by the load controller by specifying q and $p(t)$. For example, when q is set to 1, all the activation requests are forwarded to the load controller by means of unicast transmissions. This implies a better control on the aggregated power demand, because the number of active appliances is known and the only source of randomness is due to appliances' power consumption. On the contrary, when q is very small, most of the communications are unidirectional (the signal $p(t)$ transmitted from the load controller to the households), while decisions can be taken locally with minimal delays. However, the aggregate power consumption is affected by two sources of randomness (the one related to the number of active appliances, and the other one related to appliances' consumption) which require a lower admission rate for not exceeding the power constraint.

III. Control in a stationary setting

Our goal is to design a direct load control system able to impose a certain threshold on the overall power absorbed from a set of appliances with known statistical properties. In the following, we will consider that the instantaneous power consumption of an appliance is a stationary stochastic process $X(t)$.

Let $m(t)$ be the number of active appliances at time t, then the total absorbed power is $P(t) = \sum_{i=1}^{m(t)} X_i$. We consider that the energy supplier wants to guarantee that $P(t)$ does not exceed the threshold $\bar{P}$ with probability larger than ϵ,[1] i.e.:

$$\Pr\left(\sum_{i=1}^{m(t)} X_i > \bar{P}\right) < \epsilon. \quad (1)$$

[1]The interested reader can refer to [5] for a broad dissertation on how similar results can be derived for a different type of constraints on $P(t)$, dealing as well with teletraffic tools inspired by the token bucket traditionally used to shape data traffic.

A. Classic large deviation results for $q = 1$

If $q = 1$, i.e. if all the queries are forwarded to the controller, our problem is to determine the maximum value n, such that $\Pr\left(\sum_{i=1}^{n} X_i > \bar{P}\right) < \epsilon$. The same problem has been considered in the context of call admission control in telecommunication networks [10]. The interpretation of the quantities is different: the purpose is to determine the maximum number n of homogeneous data flows, each with instantaneous rate X_i in order to guarantee that the aggregate rate on a link exceeds the value $\bar{P}$ (e.g. the link capacity) with probability at most ϵ.

Different approaches have been proposed to determine n. Here, we introduce a simple one based on basic large deviation results (see e.g. [11, Ch. 6]), but more sophisticated ones could also be applied.

Let $M_X(s) = \ln \mathbb{E}[e^{sX}]$ be the cumulant generating function of $X(t)$. When n is large, Cramer's theorem can be used to approximate the probability that the sum of n independent random variables exceeds a bound ($\bar{P}$ in our case) as follows:

$$\ln \Pr(X_1 + ... + X_n > \bar{P}) \approx \inf_{s \geq 0} [n M_X(s) - s\bar{P}],$$

and then n can be determined as the largest integer such that $\left(\inf_{s\geq 0}[n M_X(s) - s\bar{P}]\right) < \ln \epsilon$. The set of values $\{0, 1, \dots, n\}$ is called the *acceptance region* for the admission controller. The approach can be easily generalized to a finite number of appliances' classes.

B. Extension to $q < 1$

In this section we show how the results above can be extended to the scenario where a subset of the appliances do not perform any query to the controller, but start autonomously their operation. For the moment we assume that appliances would like to activate according to a Poisson process with constant rate λ and we ignore retrials.Then, there is a request process with rate $q\lambda$ to the query-response system, and a spontaneous activation process of appliances with rate $\lambda_c = p(1-q)\lambda$. Both processes are Poisson ones. In the following Sec. IV we take into account the exogeneous time-variant activation process as well as the retrial mechanism. In order to keep the equations simple, we assume that all the appliances have the same activation time equal to D, but the results can be easily extended to the case when D is a random variable.

Let N be the random number of appliances starting autonomously. Under the assumptions indicated above, N is distributed as a Poisson random variable with parameter $\mathrm{E}[N] = \Lambda_c = \lambda_c D$. The condition in Eq. (1), then becomes

$$\Pr\left(\sum_{i=1}^{n} X_i + \sum_{j=n+1}^{n+N} X_j > \bar{P}\right) < \epsilon, \tag{2}$$

where $(X_i)_{i\in\mathbb{N}}$ is a sequence of independent and identically distributed random variables.

Large deviation results hold for large systems, for example when both the number of random variables and the threshold to be exceeded diverge. In Eq. (2) we need to let both the addends scale in the same way, otherwise one of them would become negligible in comparison to the other. We assume then that the expected number of appliances starting autonomously (Λ_c) scales linearly with n: $\Lambda_c = n\Lambda_c^0 = n(1-q)pD\lambda_c^0$, where we also took into account λ_c's dependence on p and $1-q$. We proved the following result:

Proposition 1: Let $(X_i)_{i\in\mathbb{N}}$ be a sequence of independent and identically distributed random variables with cumulant generating function $M_X(s)$, and N be a Poisson random variable with mean $n\Lambda_c^0$. We assume that $M_X(s)$ is defined for any s and $\Pr(X > c) > 0$. It holds:

$$\lim_{n\to\infty} \frac{1}{n} \ln \Pr\left(\sum_{i=1}^{n} X_i + \sum_{j=n+1}^{n+N} X_j > nc\right) = \inf_{s\geq 0}\left[M_X(s) + \Lambda_c^0\left(e^{M_X(s)} - 1\right) - sc\right].$$

In a finite-size system, this result is used to approximate the probability that the total power consumption exceeds the bound $\bar{P}$ as follows:

$$\ln \Pr\left(\sum_{i=1}^{n} X_i + \sum_{j=n+1}^{n+N} X_j > \bar{P}\right) \approx \inf_{s\geq 0}\left[n M_X(s) - \lambda_c D\left(e^{M_X(s)} - 1\right) - s\bar{P}\right].$$

and then the acceptance region is made by all the pairs of values (p, n) such that:

$$\inf_{s\geq 0}\left[n M_X(s) + \lambda_c D\left(e^{M_X(s)} - 1\right) - s\bar{P}\right] \leq \ln \epsilon. \tag{3}$$

C. A numerical example

In order to illustrate the admission control rule determined by the results above, we consider a toy example where there is a single class of appliances, which would like to activate according to a Poisson process with rate $\lambda = 12$ appliances per minute. The appliance is active for $D = 90$ minutes and its consumption profile can be modeled as the two-state Markov process: in the high consumption state it consumes 1.5kW for 10 minutes (on average), in the low consumption 500W for 20 minutes.

For this appliance, the cumulant generating function can be easily calculated as $M_X(s) = \ln\left(1/3e^{s1.5} + 2/3e^{s0.5}\right)$. Then inequality (3) can be used to determine the acceptable pairs (p, n) with $\lambda_c = p(1-q)\lambda$.

The term $e^{M_X(s)}$ grows very fast for $s \geq 0$ and because of this the point of minimum for the left-hand side of Eq. (3) is close to 0, specially for larger $p(1-q)$. The frontier can then be approximated as follows:

$$\inf_{s\geq 0}\left[\left(n + (1-q)pD\lambda\right) M_X(s) - s\bar{P}\right] = \ln \epsilon,$$

then, for a given value of q, the frontier is made by pairs (p, n) such that $n + (1-q)pD\lambda = \text{constant}$. To completely characterize this linear relation, it is enough to observe that for $p = 0$ the maximum value for n (let us denote it as $n^*_{p=0}$) does not depend on q, i.e. all the frontiers pass by

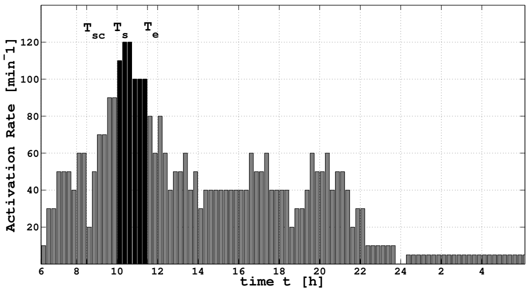

Fig. 3. Aggregate activation rate for 30000 washing machines.

the point $(0, n^*_{p=0})$ and they are described by the following linear equation parameterized in q:

$$n + (1-q)pD\lambda = n^*_{p=0}. \tag{4}$$

In particular, if the constraint $\bar{P}$ is set equal to 80% of the expected power consumption in absence of any control, it results $n^*_{p=0} = 800$. If there are n appliances activated through the query-response mechanism, the expected number of appliances in the system is $n+(1-q)pD\lambda$ and then the expected power consumption is $\mathrm{E}[P] = (n+(1-q)pD\lambda)\mathrm{E}[X]$. This means that, for a given q value, all the points in the frontier described by Eq. (4) have the same resources' usage $\mathrm{E}[P] = n^*_{p=0}\mathrm{E}[X]$ while satisfying the constraint. Moreover, for any value of q we can achieve the same efficiency by selecting (n, p) on the corresponding frontier. Although Eq. (4) stems from a linear approximation, it suggests that the efficiency (in terms of average utilization) of our control mechanism is not very sensitive to the parameter q. This is confirmed by our numerical example: we observed less than 5% reduction of the maximum expected load changing q from $q = 1$ to $q = 0$ despite the fact that for $q = 0$ the control needs to deal also with the variability of the number of appliances admitted. However, the importance of direct queries in practical applications is not only due to efficiency reasons, but also to the possibility to accurately estimate the appliance activation rate as discussed in Sec. IV. In [5] we further discuss the relation between q and average utilization.

IV. Control in a time-variant setting

In this section, we show how the analytical results derived above can be practically used when the appliances' activation process is not stationary, its rate is unknown, and probabilistically controlled appliances retry to activate some time later.

Indeed, the usage of electric appliances exhibits a strong time-of-the-day effect. For example Fig. 3 shows the activation rate of washing machines over 15 minutes intervals, as derived from data in [12]. We need then to take into account the effect of a time variant $\lambda(t)$, which is generally unknown even if historical data may be available.

Because of the time variability of the process to control, the control policy needs to be time-varying too, i.e. in general we will have $n(t)$ and $p(t)$. There are different possible choices on how to jointly adapt the two control actions. They lead to different performance in terms of resources' usage (see [5] for the effect of n and p on the efficiency), communication requirements (e.g. if p is constant, it does not need to be transmitted periodically) and fairness between the two groups of appliances—those controlled probabilistically and those controlled through query-response—(e.g. in terms of delay before the activation). In this paper, we decided to consider $n(t) = \text{const} = n$ and then to compensate for process changes by dynamically tuning $p(t)$. The control is time-slotted with time intervals of length T_c. Without loss of generality we consider that the control starts at time $t = 0$, and we denote by p_k the value of the control action during the k-th time slot, i.e. $p(t) = p_k$ for $t \in [(k-1)T_c, kT_c)$. For the sake of simplicity we will also assume that the retrial delay T is equal to T_c and the activation time of the appliance is a multiple of T_c ($D = dT_c$), even if the three parameters are in general independent.

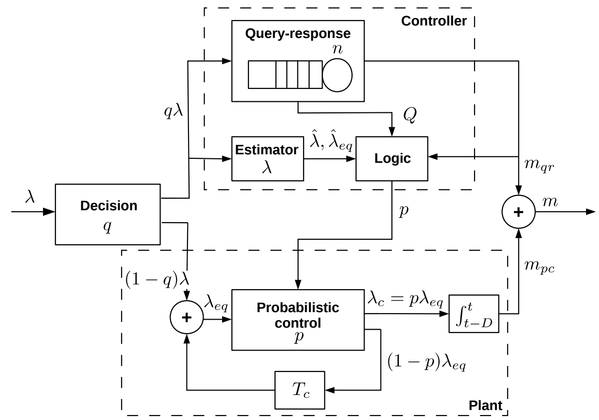

Fig. 4. System block diagram.

Fig. 4 shows the model of the whole system we are going to describe in the following sections. As it is usual in control theory, we call plant the combination of the process under control and the control actuator.

The plant is shown in the lower part of Fig. 4. The input is the spontaneous activation process with rate λ_k during the k-th time slot, i.e. the process of the activation instants in absence of any form of control (we omit time dependence in the figure). The control system assumes $\lambda(t)$ to be constant during a control slot, while this is not necessarily the case. In Sec. V we evaluate the effects of such an approximation. We keep assuming that the point process of all the activation time instants can be correctly modeled by a (non-homogeneous) Poisson process with rate $\lambda(t)$, because it originates from the superposition of many independent individual choices (users deciding to turn on their appliances). The initial requests are randomly split in two independent Poisson processes with rate respectively $q\lambda(t)$ and $(1-q)\lambda(t)$. The rate of appliances that will consider to activate autonomously in slot k is $\lambda_{eq,k}$. It holds:

$$\lambda_{eq,k} = (1-q)\lambda_k + \lambda_{eq,k-1}(1-p_{k-1}), \tag{5}$$

where the first addend is due to all the appliances that are considering to activate themselves for the first time during the k-th time slot, and the second one is due to those that have already considered this decision in the $(k-1)$-th slot and have postponed it to the current one. The actual activation rate of probabilistically controlled appliances[2] is $\lambda_{c,k} = p_k \lambda_{eq,k}$. Finally, the actual number of active appliances $m_{pc,k}$ admitted through the probabilistic control and active at time kT_c is equal to those arrived in the interval $[kT_c - D, kT_c]$. This is a Poisson random variable with expected value $\mathrm{E}[m_{pc,k}] = \sum_{h=k-d+1}^{k} \lambda_{c,h}$.

The upper part of Fig. 4 describes the controller. This directly receives the activation queries with rate $q\lambda_k$ and it manages the activations as in a $M/D/n$ queue by guaranteeing that the number of appliances active at a given time does not exceed n. Let $m_{qr,k}$ denote the number of appliances controlled through the query-response mechanism and active at time kT_c.

Estimators. The controller does not know the state of the plant (e.g. how many appliances are taking the decision to activate autonomously). It needs then to estimate the rates λ and λ_{eq}. An estimate for quantity x is denoted as $\hat{x}$. In particular two different set of estimates will be useful: $\hat{\lambda}_k^p$ and $\hat{\lambda}_{eq,k}^p$ will estimate the sequences until the current slot k, while $\hat{\lambda}_{k+1}^f$ and $\hat{\lambda}_{eq,k+1}^f$ will be used as a prediction for the slot $k+1$ in order to determine p_{k+1}. In this paper we consider simple estimators for these quantities, but we show that they work reasonably well in Sec. V. Given $N_{r,k}$ the number of queries received during the k-th slot, the Maximum Likelihood Estimator for λ_k is simply $\hat{\lambda}_k^p = N_{r,k}/(qT_c)$. Clearly, this estimate could be improved if some a-priori is available (e.g. from historical data). The prediction for slot $k+1$ is $\hat{\lambda}_{k+1}^f = \hat{\lambda}_k^p$.

The estimates $\hat{\lambda}_{eq,k}^p$ and $\hat{\lambda}_{eq,k+1}^f$ are obtained as a function respectively of $(\hat{\lambda}_k^p, \hat{\lambda}_{eq,k-1}^p)$ and of $(\hat{\lambda}_{k+1}^f, \hat{\lambda}_{eq,k}^p)$ using Eq. (5) (see [5] for more details).

Control logic. The controller determines p_{k+1} on the basis of the acceptance region derived as described in Sec. III. For simplicity we consider a linearized frontier:

$$n + \Lambda_{k+1} = \text{const}, \tag{6}$$

where Λ_{k+1} is the expected number of active probabilistically controlled appliances at the end of the $(k+1)$-th slot. The companion report [5] describes how we have selected the constant on the right hand side and the value of n. In the stationary case without retrials, it was simply $\Lambda_{k+1} = \lambda_c D = \lambda(1-q)pD$. Here, we can express Λ_{k+1} as the sum of two terms, one (denoted as $\Lambda_{hist,k+1}$) due to all the probabilistically controlled appliances already in the system at the begin of slot $k+1$, the other due to the estimated number of appliances which will activate during slot $k+1$. We can estimate $\Lambda_{hist,k+1}$ as follows $\hat{\Lambda}_{hist,k+1} = \sum_{h=0}^{d-2} p_{k-h} \hat{\lambda}_{eq,k-h}^p T_c$, and then

$$\hat{\Lambda}_{k+1} = \hat{\Lambda}_{hist,k+1} + \hat{\lambda}_{eq,k+1}^f T_c p_{k+1}. \tag{7}$$

Finally p_{k+1} can be iteratively derived from Eq. (6) and Eq. (7) This may lead to a too prudential strategy, because the number $m_{qr,k+1}$ of active appliances in the query-response

[2] It is possible to show that also the point process of such activations is a Poisson process.

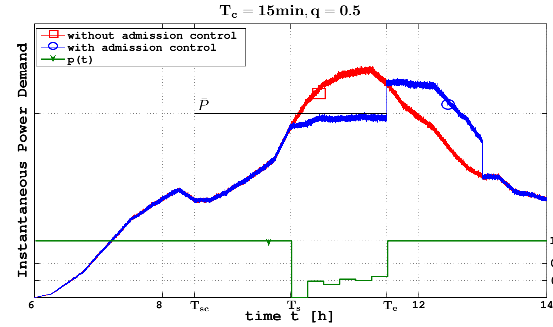

Fig. 5. Instantaneous absorbed power (with and without the applied control), and the activation probability $p(t)$.

queue can be significantly smaller than n, specially at the begin of the control period (the queue fills initially with rate λq) and then the configuration above would lead to a severe underutilization. A solution is to predict the number of active appliances in the query-response queue during the residual control period (i.e. $\hat{m}_{qr,h}$ for $h > k$). Then, p_{k+1} can be calculated replacing n in Eq. (6) with the most pessimistic forecast until the end of the control period, i.e. with $\max\{\hat{m}_{qr,h}, h > k\}$ (see [5] for more details). This is the approach we adopted.

V. Numerical Results

In this section we show the performance of the system described in Sec. IV in a realistic setting. In particular we consider 30000 washing machines under control. The instantaneous power consumption of a washing machine is assumed to follow the simple model in Sec. III-C and the activation time is $D = 90$ minutes. The spontaneous activation rate $\lambda(t)$ is derived from experimental data in [12] and it is shown in [5]. The largest expected power demand $P_{\max}$ is at around time 11.00am. We consider that the energy supplier sets the constraint $\Pr(P > \bar{P} = 0.8P_{\max}) < 0.1$ in the interval $[T_s, T_e] = [10.00\text{am}, 11.30\text{am}]$. We assume that already active appliances cannot be turned off, then the control may need to start at $T_{sc} = T_s - D = 8.30\text{am}$ in order to satisfy the constraint at $T_s = 10.00\text{am}$.

Fig. 5 plots the evolution of the power demand with and without control for $q = 0.5$ and $T_c = 15$ minutes together with the probability signal $p(t)$. We observe that the controller does not actually affect the system ($p(t) = 1$) until $t = 10.00\text{am}$ and that it actually manages to maintain the absorbed power below $\bar{P}$ for the whole duration of the control interval. Observe also how power consumption significantly increases after T_e. This is due to the fact that a severe constraint has been imposed for a long time interval. The power increase can be made smooth by gradually increasing $\bar{P}$ after T_e.

Fig. 6 shows the probability to exceed the bound at time $t = 10.26\text{am}$ estimated over 1000 simulations for different values of $q \in [0, 1]$ and for $T_c = 1, 5, 15, 30$ minutes. The time instant of observation falls in the interval where there is the largest activation rate after a period when the rate has been almost constantly increasing. It is then a particularly

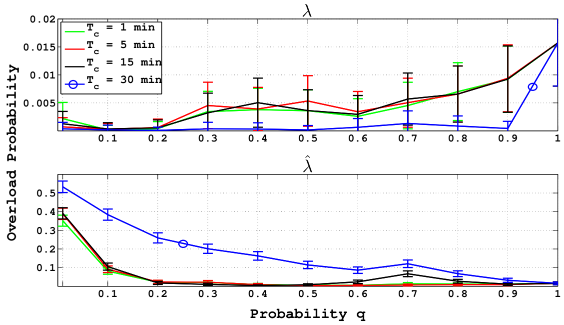

Fig. 6. Probability to exceed $\bar{P}$ at $t = 10.26$am.

critical instant for the control system. All confidence intervals in the figures have 95% confidence level. The upper plot corresponds to the case when the system has perfect estimation of the average request rate in the next timeslot, i.e. $\hat{\lambda}^f_{k+1} = \lambda_{k+1}$. We observe that the probability values are well below $\epsilon = 0.1$. The curves almost overlap for all the values of T_c but $T_c = 30$ minutes. This is due to the fact that the actual activation request rate is constant over 15 minutes time intervals, then for $T_c = 1, 5, 15$ minutes, the knowledge of the average rate in the next slot corresponds to the knowledge of the actual rate. For $T_c = 30$ minutes, the average arrival rate is a bad predictor for the actual arrival rate. The lower plot in Fig. 6 shows the same metrics when the simple estimators described in Sec. IV are used. In this case, we expect an increase of utilization, due to the fact that the controller will usually underestimate λ_{k+1} for $t < 10.00$am (because the arrival rate keeps increasing), so it selects a too high probability p_{k+1} allowing the activation of a number of appliances larger than the correct value. This error has larger consequences for small q, when a larger percentage of appliances is activated through the probabilistic control. In particular for $q < 0.1$ the constraint is no more satisfied. For $T_c = 30$ minutes, the controller uses the average rate measured in $[9.30\text{am}, 10.00\text{am}]$ to estimate the arrival rate during the interval $[10.00\text{am}, 10.30\text{am}]$ with about a 20% of relative error. This justifies the bad performance achieved with this setting. We observe that one could simply counteract the estimation errors, for example by reducing p_{k+1} by a given factor corresponding to the maximum variability of the arrival rate from a control slot to the following one.

In the technical report [5], there are also the curves for the time-average utilization and overload probability over the whole control interval which show similar behavior. Moreover, we evaluate there how the communication overhead at the controller depends on the request arrival rate, the control interval $[T_{sc}, T_e]$ and the parameters q and T_c.

VI. Conclusion

In this paper we propose a DLC scheme for smart grids that can work with a large number of dumb appliances. The main idea is to control a class of electric appliances by combining a centralized query-response system with a probabilistic system (periodically programmed by the central controller) able to take local decisions. The combination of the two approaches can be configured for providing the desired trade-off between resource utilization, communication overhead and privacy. The control policy has been designed by using well-established teletraffic tools for admission control, which work well for large scale systems, in order to provide probabilistic guarantees on the power demand in the controlled zone.

VII. Acknowledgements

We thank the anonymous reviewer who provided us a simpler proof of Proposition 1. This research was partly sponsored by the European Commission within the framework of the CONGAS project FP7-ICT-2011-8-317672 and by the Italian PON iNext.

References

[1] L. Mathieu, S. Koch, and D. Callaway, "State estimation and control of electric loads to manage real-time energy imbalance," *Power Systems, IEEE Transactions on*, vol. 28, no. 1, p. 430 440, 2013.

[2] H. Hao, Y. Lin, A. Kowli, P. Barooah, and S. Meyn, "Ancillary Service to the Grid through Control of Fans in Commercial Building HVAC Systems," *IEEE Transactions on Smart Grid*, vol. 5, no. 4, 2014.

[3] W. Kempton and S. E. Letendre, "Electric vehicles as a new power source for electric utilities," *Transportation Research Part D: Transport and Environment*, vol. 2, no. 3, pp. 157 – 175, 1997. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1361920997000011

[4] S. Meyn, P. Barooah, A. Busic, and J. Ehren, "Ancillary service to the grid from deferrable loads: The case for intelligent pool pumps in Florida," in *Decision and Control (CDC), 2013 IEEE 52nd Annual Conference on*, Dec 2013, pp. 6946–6953.

[5] G. Neglia, G. Di Bella, L. Giarré, and I. Tinnirello, "Scalable and Privacy-Preserving Admission Control for Smart Grids," Inria, Tech. Rep. RR-8769, 2015.

[6] Y. Ghiassi-Farrokhfal, S. Keshav, and C. Rosenberg, "Toward a realistic performance analysis of storage systems in smart grids," *Smart Grid, IEEE Transactions on*, vol. 6, no. 1, pp. 402–410, January 2015.

[7] O. Ardakanian, S. Keshav, and C. Rosenberg, "On the use of teletraffic theory in power distribution systems," in *Future Energy Systems: Where Energy, Computing and Communication Meet (e-Energy), 2012 Third International Conference on*, May 2012, pp. 1–10.

[8] G. Di Bella, L. Giarré, M. Ippolito, A. Jean-Marie, G. Neglia, and I. Tinnirello, "Modeling Energy Demand Aggregator for Residential Users," in *Decision and Control (CDC), 2013 IEEE 52nd Annual Conference on*, 2013, pp. 6280–6285.

[9] G. Neglia, G. Di Bella, L. Giarré, and I. Tinnirello, "Unidirectional Probabilistic Direct Control for Deferrable Loads," in *IEEE INFOCOM Workshop on Communications and Control for Smart Energy Systems (CCSES)*, 2014.

[10] E. Knightly and N. Shroff, "Admission control for statistical qos: theory and practice," *Network, IEEE*, vol. 13, no. 2, pp. 20–29, Mar 1999.

[11] F. Kelly and E. Yudovina, *Stochastic Networks*. Cambridge University Press, 2014.

[12] R. Miceli, "Sustainable development and energy saving laboratory." DIEET - University of Palermo, Tech. Rep., 2007.