A Multiarmed Bandit Incentive Mechanism for Crowdsourcing Demand Response in Smart Grids

Shweta Jain
Indian Institute of Science,
Bangalore

Balakrishnan Narayanaswamy
Department of CSE,
UCSD

Y. Narahari
Indian Institute of Science,
Bangalore

Abstract

Demand response is a critical part of renewable integration and energy cost reduction goals across the world. Motivated by the need to reduce costs arising from electricity shortage and renewable energy fluctuations, we propose a novel multi-armed bandit mechanism for demand response (MAB-MDR) which makes monetary offers to strategic consumers who have unknown response characteristics, to incetivize reduction in demand. Our work is inspired by a novel connection we make to crowdsourcing mechanisms. The proposed mechanism incorporates realistic features of the demand response problem including time varying and quadratic cost function. The mechanism marries auctions, that allow users to report their preferences, with online algorithms, that allow distribution companies to learn user-specific parameters. We show that MAB-MDR is dominant strategy incentive compatible, individually rational, and achieves sublinear regret. Such mechanisms can be effectively deployed in smart grids using new information and control architecture innovations and lead to welcome savings in energy costs.

Introduction

Peak demand and supply-demand imbalance are two critical problems faced by electricity generation and distribution companies. "In order to supply demand that varies daily and seasonally, and given that demand is largely uncontrollable and interruptions very costly, installed generation capacity must be able to meet peak demand" Strbac (2008). For example, during the peak hours of the California electricity crisis in 2000/2001, it is estimated that a 5% lowering of demand would have resulted in a 50% price reduction International Energy Agency (2003).

Distributed generation, using renewable sources, is gaining prominence and is perceived as vital to achieving cost and carbon reduction goals Ackermann, Andersson, and Soder (2001). However, an increase in electricity supply from renewable sources results in larger fluctuations due to their unreliable nature. New communication Sood et al. (2009) and actuation Farhangi (2010) infrastructure in the electricity grid will allow increased prosumer participation in grid markets. This participation raises the hope that it may be possible to extract value from a time varying and intermittent renewable energy resource through intelligent optimization of markets Bitar et al. (2011), generation Bhuvaneswari et al. (2009), storage Zhu et al. (2011) and loads Kowli and Meyn (2011). This serves as the primary motivation for this work.

We propose a novel mechanism which seeks to incentivize strategic users to reduce demand when there is a shortfall due to grid failures or time varying supply. Our specific contributions are that we,

- establish a natural parallel between demand response (DR) in energy systems and crowdsourcing, creating many opportunities for well-studied algorithms in crowdsourcing to be applied to important DR problems.
- emphasize the difference between DR and crowdsourcing in terms of both the time varying and non-linear cost function and highlight the need for melding learning with mechanism design. We also show how, while ideas from crowdsourcing can be applied to DR, the specific algorithms and proof techniques needed to obtain guaranteed performance require non-trivial modifications.
- propose a multiarmed bandit mechanism for demand response (MAB-MDR) that satisfies crucial game theoretic properties including dominant strategy incentive compatibility, individual rationality, and sublinear regret.
- demonstrate the efficacy of the proposed mechanism through extensive simulations, based on real-world energy supply and consumption data.

Demand Response in Electricity Markets

Demand response (DR) refers to the change in users' behavior (i.e. their electricity consumption) in response to signals from the utilities Albadi and El-Saadany (2007). One particular signal is dynamic pricing, which can result in cost savings, market-wide financial benefits, increased reliability, and market performance improvement QDR (2006). However, it can also lead to customer confusion due to uncertain supply, volatile prices and lack of information Chao (2012). Optimizing prices is a complex problem which has received much attention in the literature. Pricing based on the demand curve Joo and Ilic (2010) requires that users know and reveal their preferences, which is unlikely in practice and can lead to privacy and security concerns Anderson and Fuloria (2010). Contract based or Direct Load Control (DLC) approaches Hsu and Su (1991), where consumer appliances are contracted to respond directly to signals from the utility,

require that consumers are willing to give up control of their consumption.

An important point to note is that legacy entails that providers have an 'obligation to serve', i.e. providers cannot deny to users access to electricity, but can only incentivize them to voluntarily reduce or defer consumption. In simpler terms, it is preferable if providers use carrots as opposed to sticks, and this is the philosophy we take in our work. A recent body of work at the intersection of computer science and economics has established that general incentives (as opposed to just pricing) are a powerful way to allocate scarce resources. In our work, we build on these insights and propose a multiarmed bandit mechanism for the problem of making offers to the consumers to reduce their consumption so as to compensate for shortage of electricity when renewable generation does not meet expectations. These mechanisms do not have the drawbacks discussed above. In particular, we show that our mechanism is dominant strategy incentive compatible (i.e. encourages consumers to report their true cost per unit reduction), individually rational (i.e. gives each consumer positive utility for accepting the offer), and achieves a sublinear regret of $O(T^{2/3})$ where T is the number of time steps for which the mechanism runs.

Modeling Offers for Demand Response

We consider a model where offers (monetary rewards) are made to the consumers in exchange for reducing the consumption. For simplicity of exposition, we assume consumers are asked to reduce the same amount of consumption. Offers are made at the starting of each time slot. T denotes total number of slots.

Let $\bar{d} = \{\bar{d}^1, \bar{d}^2, \ldots, \bar{d}^T\}$ represent the expected total demand profile of all the consumers in each time slot. This is learned over time by the distribution company based on historical consumption. To meet demand, renewable energy sources are available to the distribution company in addition to the option of buying electricity from the market. Renewable supply is stochastic and depends on the environmental conditions. Let $\bar{r} = \{\bar{r}^1, \bar{r}^2, \ldots, \bar{r}^T\}$ represent the expected amount of energy generated from renewable resources, which is a known quantity. The distribution company then plans to buy $\bar{d} - \bar{r}$ units of electricity in the day ahead market Harris (2006).

The actual amount of energy generated by the renewable resource (r^t) is revealed only at the start of a slot t. The shortage of electricity at any slot t is thus given by $e^t = \bar{r}^t - r^t$ which has to be generated by expensive fast ramping generators or at expensive spot market prices. The question we ask in this work is how can the distribution company reduce costs by incentivizing the consumers to reduce their consumption, while learning about their uncertain response characteristics ?

We model consumers using n clusters, where all n_i consumers of a cluster i have the same utilities for electricity and the same response characteristics to offers (described more carefully below). These clusters correspond, for example, to industrial consumers of a certain industry, or may be the result of other forms of consumer segmentation. For more detailed justification see Ardakanian, Keshav, and Rosenberg (2011). For simplicity we assume that making an offer to a certain cluster i implies giving monetary incentives to all n_i consumers that belong to the cluster. Note that this assumption on its own is not limiting, since cluster size can be as small as required, though we shall see that many small clusters will result in worse bounds. We associate three quantities with each cluster, the

- price at which consumers are willing to reduce consumption by one unit (cost per unit reduction (CPR)). CPR is private information of the consumers that depends on their utility function.
- probability of accepting the offer (acceptance rate (AR)) if the offer price is greater than the CPR. Consumer may not accept the offer even if the offer price exceeds the CPR. The reasons for this include (a) an inflexible all or nothing requirement of the consumer; (b) sudden lack of requirement for the power; (c) availability of the consumer (is he at home?). Consumers have no control on these factors and thus, AR is independent of the offer prices and other consumers. AR is unknown to the consumer as well as the distribution company.
- number of consumers, which is publicly known.

These correspond to private information that require incentives for truthful revelation, stochastic information that must be learned and common knowledge. We are interested in the design of MAB mechanisms that account for all three, and we see DR as a specific instantiate of this general problem (for more details see the Related Work on MAB problems).

Clusters of consumer types are formed such that all the consumers in same cluster share the same CPR and AR, based on past bids and response behavior. For each cluster i, denote its CPR by $c_i \in [c_{min}, c_{max}]$, AR by $\lambda_i \in [0, 1]$ and number of consumers by n_i. The main challenge in computing the optimal offers involves learning ARs efficiently and eliciting CPRs truthfully. CPRs can be elicited by either

- *Posted Offer Mechanisms:* Offer the same price to every consumer and set the price based on the user behavior.
- *Auction Mechanisms:* Design a mechanism to elicit true CPR from the consumers by asking for bids.

Posted offer mechanisms are easier for the consumer as he does not need to compute his entire utility function and only needs to evaluate offered price to decide whether or not to accept the offer. Auctions are easier for the distribution company. To illustrate the power of crowdsourcing mechanisms, we use auctions to elicit CPR truthfully, while learning ARs to design price offers optimally.

Our auction mechanism proceeds as follows: The distribution company asks each cluster to report their CPR bids. Based on the learned ARs, and CPR bids, offer price to a particular cluster i is decided. If the offer price is greater than true CPR of the consumer, then he accepts the offer with probability λ_i (since he gets positive utility) and distribution company pays him the offer price. ARs and CPRs are considered static and thus do not change with time.

There is a close connection between making offers to the consumers in demand response and giving tasks to the workers in crowdsourcing. In the next section, we expand on and exploit this parallel between and show how algorithms from

Crowdsourcing	Demand Response (DR)
Workers	Consumers
Requester	Distribution company
Tasks	Offers to reduce consumption
Price of a task	Offer Price
Capability or Reliability	Acceptance rate (AR)
Willingness to do task	Cost per unit reduction (CPR)
Linear reward function	*Quadratic reward function*
Homogeneous tasks	*Time varying offers*

Table 1: Parallel between Crowdsourcing and Demand Response

crowdsourcing can be modified to design efficient mechanisms for demand response.

A Parallel with Crowdsourcing

Jeff Howe offers the following definition of crowdsourcing: "Crowdsourcing is the act of taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an undefined, generally large group of people in the form of an open call" Howe. The designated agent is often called the requester, while people working on the task are called crowd workers. The requester needs to assign the task to these workers with unknown capability and willingness to complete the task based on the goals to achieve. Table 1 precisely demonstrates this parallel between crowdsourcing and demand response. This opens up the possibility of using intuition from techniques for crowdsourcing for DR. There is a vast body of literature available in crowdsourcing with various models, we point out some of the more relevant ones to show how objectives in crowdsourcing and demand response relate. Tran-Thanh et al. Tran-Thanh et al. (2012) present a model to assign the task to the workers with highest capability while having a budget constraint on the total cost of giving tasks. In DR, this involves making offers to the consumers having highest AR where the distribution company has a certain budget. In other work, tasks are given adaptively to the workers until a certain accuracy is achieved Abraham et al. (2013). In DR, one can make the offers to the consumers by asking sequentially to reduce their consumption until a certain amount of reduction is achieved. Another line of work involves designing a posted price mechanism where workers having homogeneous quality arrive online Babaioff et al. (2012); Singla and Krause (2013). In our work here, we only look at auction mechanisms and do not explore this direction, but leave it for (very relevant) future work.

While the above parallels suggest that one can directly borrow ideas from crowdsourcing, there are certain unique features of the DR problem that need to be accounted for. Making offers to consumers having the highest AR is not optimal as offers should also depend on the (time varying) shortage of electricity distribution company faces. In addition, the objectives in crowdsourcing (maximizing accuracy) are very different from DR. While in crowdsourcing a linear reward is usually considered, in DR, the reward function is generally quadratic and thus requires non-trivial extensions to some of the existing MAB algorithms. Moreover, while our work considers learning ARs and eliciting CPRs simultaneously, no existing work in crowdsourcing has considered both learning qualities and eliciting true costs in a setting such as ours. Thus, while borrowing intuition from the crowdsourcing literature, we present new algorithms and bounds that extend the state-of-the-art in MAB mechanisms.

Notation	Description
$\mathcal{N}$	Set of all clusters
n	Number of clusters
T	Total number of rounds
$c_i, \hat{c}_i$	Actual and reported CPR of consumer i
λ_i	AR of all consumers in cluster i
n_i	Number of consumers in cluster i
n_+, n_-	$\max_i n_i$ and $\min_i n_i$
e^t	Shortage of electricity at time t
$a_i^t \in \{0,1\}$	Allocation of cluster i at time t
p_i^t	Payment for a consumer in cluster i
τ	Number of exploration rounds

Table 2: Notation used

Utility Model

The notation table is given in Table 2. Denote the CPR bid profile by $\hat{c} = \{\hat{c}_1, \hat{c}_2, \ldots, \hat{c}_n\}$. The mechanism $\mathcal{M} = (\mathcal{A}(\hat{c}), \mathcal{P}(\hat{c}))$ consists of (1) an allocation rule, $\mathcal{A} = \{\mathcal{A}^1, \mathcal{A}^2, \ldots, \mathcal{A}^T\}$ where each $\mathcal{A}^t = \{a_1^t, a_2^t, \ldots, a_n^t\}$ is the vector of zeros and ones (deterministic mechanism), and (2) a payment rule $\mathcal{P} = \{\mathcal{P}^1, \mathcal{P}^2, \ldots, \mathcal{P}^T\}$ where each $\mathcal{P}^t = \{p_1^t, p_2^t, \ldots, p_n^t\}$ represents the offer price. If offer is not made to the cluster i i.e. $a_i^t = 0$ then the payment p_i^t is zero. If $a_i^t = 1$ then all the consumers in type i are given the offers at slot t, whereas, $\sum_i a_i^t = 0$ means offer is not given at all. At any time slot, offers are made to a single type of consumers, i.e., $\forall t, \sum_i a_i^t \leq 1$. Note that the allocation rule and the payment rule corresponding to any time slot always depend on the bid profile $\hat{c}$, but for brevity of notation, we do not explicitly mention this dependence. Let $\mathcal{N} = \{1, 2, \ldots, n\}$ denote the set of all types of consumers. The expected value of each consumer in type i is given by $-\lambda_i c_i$ as an offer is accepted with probability λ_i. The expected utility of any consumer in type i at any slot t from mechanism $(\mathcal{A}, \mathcal{P})$ is given by:

$$\mathbb{E}[u_i(\mathcal{A}^t(\hat{c}_i, \hat{c}_{-i}), c_i)] = \lambda_i(-a_i^t c_i + p_i^t)$$

$\hat{c}_{-i}$ is the bid profile of consumers in types other than i.

For the distribution company, the expected valuation at time t is given by $R - (e^t - \sum_i a_i^t n_i \lambda_i)^2$, where R is the base utility for the distribution company, e^t is the shortage and $n_i \lambda_i$ is the expected number of reduction units from consumer type i. Thus, the second term denotes the cost of procuring electricity which is usually quadratic Harris (2006). The net utility of the distribution company is:

$$\mathbb{E}[u_C(\mathcal{A}^t(\hat{c}_i, \hat{c}_{-i}), c_i)] = R - (e^t - \sum_i a_i^t n_i \lambda_i)^2 - \sum_i n_i \lambda_i p_i^t$$

We emphasize that the generality of our mechanism does allow more complex models. The expected social welfare (sum of the expected valuations) from mechanism $(\mathcal{A}, \mathcal{P})$ is:

$$W(\mathcal{A}) = R - \sum_{t=1}^{T} (e^t - \sum_i a_i^t n_i \lambda_i)^2 - \sum_{t=1}^{T} \sum_i a_i^t \lambda_i n_i c_i \quad (1)$$

We now define some of the desirable properties that a mechanism $\mathcal{M} = (\mathcal{A}, \mathcal{P})$ should satisfy:

Definition 1 *Allocative Efficiency (AE): An allocation rule $\mathcal{A}$ is said to be allocatively efficient if it maximizes the social welfare. Formally, $\mathcal{A}$ is allocatively efficient if:*

$$\mathcal{A} \in \arg\max_{\mathcal{A}} W(\mathcal{A})$$

Definition 2 *Dominant Strategy Incentive Compatibility (DSIC): Mechanism $\mathcal{M}$ is said to be DSIC if truth telling is a dominant strategy for all the types of consumers. Formally, $\forall i \in \mathcal{N},\ \hat{c}_{-i}$ and $\hat{c}_i$, we have,*

$$\sum_{t=1}^{T} p_i^t(c_i, \hat{c}_{-i}) - c_i a_i^t(c_i, \hat{c}_{-i}) \geq \sum_{t=1}^{T} p_i^t(\hat{c}_i, \hat{c}_{-i}) - c_i a_i^t(\hat{c}_i, \hat{c}_{-i})$$

Definition 3 *Individual Rationality (IR): Mechanism $\mathcal{M}$ is said to be IR for a consumer if participating in mechanism always gives him positive utility. Formally,*
$-\hat{c}_i a_i^t(\hat{c}_i, c_{-i}) + p_i^t(\hat{c}_i, c_{-i}) \geq 0\ \forall t\ \forall i \in \mathcal{N}, \forall \hat{c}_i$ *and* $\forall c_{-i}$

At each slot t, distribution company observes the shortage e^t, and makes an offer to a consumer type which maximizes the social welfare given by Equation (1) and thus involves solving the following optimization problem:

$$\min(\min_i \lambda_i n_i(\lambda_i n_i - 2e^t + c_i), 0)$$

The outer minimum ensures that offers are made only if there is an increase in social welfare. Since the response behavior of the consumers is unknown, the optimization problem cannot be solved directly, our proposed strategy involves learning ARs and eliciting CPRs. The distribution company will have to both explore (make offers to consumers so as to learn their AR) and exploit (make offers to the optimal consumer with best capability so far). Multiarmed bandit (MAB) mechanisms offer a natural solution.

Related Work on MAB Problems

The mechanism design problem we study can be considered an extension of multiarmed bandit problems where rewards are stochastic. A recent survey by Bubeck and Bianchi Bubeck and Cesa-Bianchi (2012) compiles several variations. The most popular algorithm, Upper Confidence Bound (UCB) was proposed by Peter Auer et al. Auer, Cesa-Bianchi, and Fischer (2002). Our mechanism is exploration separated, and as a result, is also related to work on mechanisms in the probably approximately correct (PAC) setting proposed by Even-Dar et al. Even-Dar, Mannor, and Mansour (2006) and Kalyanakrishnan et al. Kalyanakrishnan and Stone (2010). These algorithms are specific to linear reward functions and do not consider quadratic reward functions which is the focus of our work. Our work is also related to contextual multiarmed bandits when electricity shortage is considered to be the context. However, existing algorithms Langford and Zhang (2007); Lu, Pál, and Pál (2010) cannot be applied directly due to strategic nature of the consumers.

MAB mechanisms constitute a more recent direction of research, and have been explored mostly for sponsored search auctions Babaioff, Kleinberg, and Slivkins (2010); Babaioff, Sharma, and Slivkins (2009); Devanur and Kakade (2009); Gatti, Lazaric, and Trovò (2012) where the reward function is again linear and is a forward auction setting. We extend MAB mechanisms to the reverse setting while accounting for the quadratic reward.

We consider social welfare maximization since electricity is a socially important good and it is natural to design mechanisms that maximize a measure of social utility. However, maximizing the welfare of the company or its revenue can be similarly studied.

Social Welfare Regret

The Social welfare regret of a mechanism is defined as the difference between the maximum social welfare achieved if ARs are known by any mechanism and the expected social welfare achieved when ARs are unknown. Formally, social welfare regret of a mechanism $(\mathcal{A}, \mathcal{P})$ is:

$$R(\mathcal{A}) = \sum_{t=1}^{T} a_i^t n_i \lambda_i(\lambda_i n_i + c_i - 2e^t) - \sum_{t=1}^{T} \min(\min_i n_i \lambda_i(\lambda_i n_i + c_i - 2e^t), 0)$$

Our goal is to design an algorithm that achieves sublinear regret (i.e. regret that goes asymptotically to 0 as T goes to infinity) and satisfies the properties in Definitions 1, 2 and 3.

MAB-MDR: A Novel Exploration-Separated Mechanism

Since ARs are unknown, we cannot apply a Vickery-Clarke-Groves (VCG) mechanism Narahari (2014) to elicit CPRs truthfully Babaioff, Sharma, and Slivkins (2009); Devanur and Kakade (2009). Multiarmed Bandit mechanisms provide a powerful alternative. Existing MAB algorithms require reward functions to be monotone in terms of unknown qualities Auer et al. (2003); Chen, Wang, and Yuan (2013) but monotonicity condition is violated in our quadratic case. Thus, traditional MAB algorithms and analysis do not apply. In this section, we propose a novel exploration separated mechanism provided in Algorithm 1 that satisfies all the desired properties mentioned in Definitions 1, 2 and 3 with unknown ARs and strategically revealed CPRs.

The proposed mechanism is exploration separated where, the mechanism learns the ARs of all the consumer types by giving the maximum offer price for τ (given in Corollary 1) number of rounds in round robin fashion. Each consumer of type i, accepts the offer with probability λ_i, independently of other consumers in the same type. Thus, n_i independent samples are generated by making offers to i^{th} types and these can be used to learn the AR λ_i. Let $\hat{\lambda}_i$ represent the sample mean of λ_i after exploration phase. After τ rounds are completed, upper confidence bound $\hat{\lambda}_i^+$ and lower confidence bound $\hat{\lambda}_i^-$ are computed using Hoeffding's inequality. These are bounds for the true ARs.

For this mechanism, Theorem 2 shows that the mechanism is DSIC. As a result, any consumer in a particular cluster can be asked to bid his CPR (which will be truthful and hence same for all the consumers). After the exploration phase, ARs are not updated in the exploitation phase. This ensures incentive compatibility. In each round the type that maximizes the social welfare with respect to the bounds on ARs is chosen.

Algorithm 1: MAB-MDR

Input: CPR bids, $\{\hat{c}_1, \hat{c}_2, \ldots \hat{c}_n\}$, Number of exploration steps τ, confidence parameter μ

Output: Sequence of allocations $\mathcal{A}^1, \mathcal{A}^2, \ldots \mathcal{A}^T$ and payments $\mathcal{P}^1, \mathcal{P}^2, \ldots, \mathcal{P}^T$

1 (Exploration Rounds) **for** $t \leftarrow 1$ ***to*** τ **do**

2 Make offers to all consumers in type $i = ((t-1) \mod n) + 1$, $a_i^t = 1$, $a_j^t = 0 \; \forall j \in \mathcal{N} \setminus i$

3 Make a payment $p_i^t = c_{max}$ to consumer in type i if he accepts the offer.

4 $\forall i$, Update $\hat{\lambda}_i$, $\hat{\lambda}_i^+ = \min(\hat{\lambda}_i + \sqrt{\frac{n}{2\tau n_i} ln(\frac{2}{\mu})}, 1)$, $\hat{\lambda}_i^- = \max(\hat{\lambda}_i - \sqrt{\frac{n}{2\tau n_i} ln(\frac{2}{\mu})}, 0)$

5 (Exploitation rounds) **for** $t \leftarrow \tau + 1$ ***to*** T **do**

6 $i = \arg\min_i n_i \hat{\lambda}_i^+ (n_i \hat{\lambda}_i^- + \hat{c}_i - 2e^t)$

7 **if** $(n_i \hat{\lambda}_i^- + \hat{c}_i - 2e^t) \leq 0$ **then**

8 $a_i^t = 1$ (Make offers to all consumers in type i), $a_j^t = 0 \; \forall j \in \mathcal{N} \setminus i$

9 $j = \arg\min_{j \in \mathcal{N} \setminus i} n_j \hat{\lambda}_j^+ (n_j \hat{\lambda}_j^- + \hat{c}_j - 2e^t)$

10 **if** $(n_j \hat{\lambda}_j^- + \hat{c}_j - 2e^t) \leq 0$ **then**

11 $p_i^t = \min(\frac{n_j \hat{\lambda}_j^+ (n_j \hat{\lambda}_j^- + \hat{c}_j - 2e^t) - n_i \hat{\lambda}_i^+ (n_i \hat{\lambda}_i^- - 2e^t)}{\hat{\lambda}_i^+ n_i}, c_{max})$

12 **else**

13 $p_i^t = \min(-(n_i \hat{\lambda}_i^- - 2e^t), c_{max})$

14 Make payment p_i^t to consumer in type i if he accepts the offer.

15 **else**

16 $a_j^t = 0 \; \forall j \in \mathcal{N}$ (Make no offer at all)

In the simpler linear cost model, consumers with the highest upper confidence bound are selected, but since the costs are quadratic in terms of AR in our model, the consumers have to be selected more carefully. In particular, our mechanism uses *lower* confidence bounds. Offers are only made if consumers in a type generate higher expected social welfare compared to the case where no offer is made at all (ensured by step 7). Finally, the offer price is paid to each consumer if he accepts the offer. Offer prices are set such that it is truthful for every type of consumers to report true CPRs; at the same time, the mechanism results in positive utility to every consumer.

Note: CPR and AR of all the consumers are assumed to be same for all the slots. Otherwise, multiple instances of the mechanism can be run for each slot j and (for example) for weekdays and weekends.

Analysis of MAB-MDR

In this section, we provide theoretical guarantees for MAB-MDR. Our analysis is very different from that of traditional MAB setting, as picking a cluster with highest upper confidence bound (UCB) is non-trivial with quadratic costs. Lemma 1 shows that MAB-MDR offers a price that is greater than the CPR bid of the consumer to whom offer is made and thus he derives positive utility (Theorem 1). Theorem 2 shows that it is always in the best interest of consumers to bid their true CPRs. We then provide bounds on the expected regret in Theorem 3.

Lemma 1 *The offer price is greater than CPR bid for the consumer to whom offer is made.*

Proof: W.L.O.G, $p_i^t(\hat{c}_i, \hat{c}_{-i}) < c_{max}$. $\forall j \neq i$, we have, $n_i \hat{\lambda}_i^+ (n_i \hat{\lambda}_i^- + \hat{c}_i - 2e^t) \leq \min(n_j \hat{\lambda}_j^+ (n_j \hat{\lambda}_j^- + \hat{c}_j - 2e^t), 0)$ $\implies \hat{c}_i \leq p_i^t(\hat{c}_i, \hat{c}_{-i})$ (Rearranging Terms) □

Theorem 1 *MAB-MDR is IR for all the consumers.*

Proof: Follows from Lemma 1 □

Theorem 2 *MAB-MDR is DSIC.*

Proof: We will show that $\forall t$, $\forall i \in \mathcal{N}$, $\forall \hat{c}_{-i}$ and $\forall \hat{c}_i$,

$$p_i^t(c_i, \hat{c}_{-i}) - c_i a_i^t(c_i, \hat{c}_{-i}) \geq p_i^t(\hat{c}_i, \hat{c}_{-i}) - c_i a_i^t(\hat{c}_i, \hat{c}_{-i})$$

In exploration steps, fixed offers (c_{max}) are made irrespective of the bids. Thus, the condition is trivially satisfied. For any exploitation round t,

- Case 1: $c_i > \hat{c}_i \implies a_i^t(c_i, \hat{c}_{-i}) \leq a_i^t(\hat{c}_i, \hat{c}_{-i})$
  If $a_i^t(c_i, \hat{c}_{-i}) = a_i^t(\hat{c}_i, \hat{c}_{-i})$, then above condition is trivially satisfied as the payments does not depend on the bid of i^{th} consumer. W.L.O.G., $0 = a_i^t(c_i, \hat{c}_{-i}) < a_i^t(\hat{c}_i, \hat{c}_{-i}) = 1$ This implies $\forall j \neq i$

  $$n_i \hat{\lambda}_i^+ (n_i \hat{\lambda}_i^- + c_i - 2e^t) \geq \min(n_j \hat{\lambda}_j^+ (n_j \hat{\lambda}_j^- + \hat{c}_j - 2e^t), 0)$$
  $$n_i \hat{\lambda}_i^+ (n_i \hat{\lambda}_i^- + \hat{c}_i - 2e^t) \leq \min(n_j \hat{\lambda}_j^+ (n_j \hat{\lambda}_j^- + \hat{c}_j - 2e^t), 0)$$
  $$\implies -c_i + p_i^t(\hat{c}_i, \hat{c}_{-i}) \leq 0$$

- Case 2: $c_i < \hat{c}_i$

$a_i^t(c_i, \hat{c}_{-i}) \geq a_i^t(\hat{c}_i, \hat{c}_{-i}) \implies 1 = a_i^t(c_i, \hat{c}_{-i}) > a_i^t(\hat{c}_i, \hat{c}_{-i}) = 0$. Thus, $-c_i + p_i^t(c_i, \hat{c}_{-i}) \geq 0$ (Lemma 1) □

Let $\Delta^t = \max_i \lambda_i n_i (\lambda_i n_i - 2e^t + c_i) - \min_i \lambda_i n_i (\lambda_i n_i - 2e^t + c_i)$, $n_+ = \max_i n_i, n_- = \min_i n_i$ and $\epsilon = \sqrt{\frac{n}{2\tau n_-} ln(\frac{2}{\mu})}$.

Lemma 2 *Expected social welfare regret ($\mathbb{E}[r^t]$) at any time slot t in the exploitation round is bounded by:*

$$(1 - 2\mu) n_+ \epsilon (\max((n_+ + \frac{2e^t}{n_+}), 2n_+)) + 2\mu\Delta^t$$

Proof: The social welfare regret at time t if type $i \neq i^*$ is selected where i^* is the optimal type is given by:

$$r^t = \lambda_i n_i (\lambda_i n_i - 2e^t + c_i) - \lambda_{i^*} n_{i^*} (\lambda_{i^*} n_{i^*} - 2e^t + c_{i^*})$$

By Hoeffding's inequality, with probability at least $(1 - 2\mu)$,

$$\hat{\lambda}_i^- \leq \lambda_i \leq \hat{\lambda}_i^+ \text{ and } \hat{\lambda}_{i^*}^- \leq \lambda_{i^*} \leq \hat{\lambda}_{i^*}^+ \quad (2)$$
$$\text{also, } \hat{\lambda}_i^+ - \lambda_i \leq 2\epsilon \text{ and } \lambda_i - \hat{\lambda}_i^- \leq 2\epsilon \quad (3)$$

- Case 1: $i^* \neq 0 \implies \lambda_{i^*} n_{i^*} - 2e^t + c_{i^*} \leq 0$

$$n_{i^*} \hat{\lambda}_{i^*}^+ (\hat{\lambda}_{i^*}^- n_{i^*} - 2e^t + c_{i^*}) \geq n_i \hat{\lambda}_i^+ (\hat{\lambda}_i^- n_i - 2e^t + c_i) \quad (4)$$

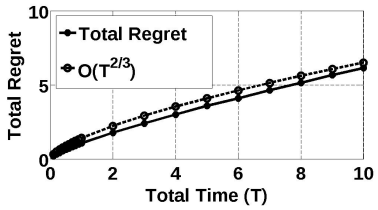

Figure 1: Effect of time on total regret. The regret follows the $O(T^{2/3})$ bound, validating our analysis (Theorem 3).

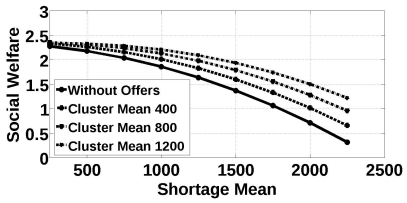

Figure 2: Effect of shortage on social welfare with/without offers with different cluster sizes. The social welfare increases when incentives are provided to larger number of consumers

Thus, with probability at least $(1-2\mu)$,

$$\begin{aligned} r^t &\leq \lambda_i n_i(\lambda_i n_i - 2e^t + c_i) - n_{i^*}\hat{\lambda}^+_{i^*}(\hat{\lambda}^-_{i^*} n_{i^*} - 2e^t + c_{i^*}) \\ &\leq \lambda_i n_i(\lambda_i n_i - 2e^t + c_i) - n_i\hat{\lambda}^+_i(\hat{\lambda}^-_i n_i - 2e^t + c_i) \\ &\leq n_i^2(\lambda_i^2 - \hat{\lambda}^+_i\hat{\lambda}^-_i) + 2e^t n_i(\hat{\lambda}^+_i - \lambda_i) - n_i c_i(\hat{\lambda}^+_i - \lambda_i) \\ &\leq n_i^2\hat{\lambda}^+_i(\lambda_i - \hat{\lambda}^-_i) + 2e^t n_i(\hat{\lambda}^+_i - \lambda_i) \\ &\leq n_+ 2\epsilon(\hat{\lambda}^+_i n_+ + \frac{2e^t}{n_+}) \leq n_+ 2\epsilon(n_+ + \frac{2e^t}{n_+}) \end{aligned}$$

where the first inequality arises from Equation (2), second from Equation (4), and the third from Equation (3).

- Case 2: $i^* = 0$ and $\hat{\lambda}^-_i n_i - 2e^t + c_i \leq 0$, Then,

$$\begin{aligned} r^t &= \lambda_i n_i(\lambda_i n_i - 2e^t + c_i) \leq \lambda_i n_i(\hat{\lambda}^-_i n_i + 2\epsilon n_i - 2e^t + c_i) \\ &\leq n_i(2\epsilon n_i) \leq n_+(2\epsilon n_+) \text{ with probability } (1-\mu) \end{aligned}$$

where, the first inequality comes from Equation (3). Thus, $r^t \leq n_+\epsilon(\max((n_+ + \frac{2e^t}{n_+}), 2n_+))$, with probability $(1-2\mu)$ and Δ^t (maximum regret) with probability 2μ. □

Theorem 3 *The expected welfare regret is bounded by:*

$$\sum_{t=1}^{\tau}\Delta^t + \sum_{t=\tau+1}^{T}(2\mu\Delta^t + (1-2\mu)\epsilon n_+(\max\left(n_+ + \frac{2e^t}{n_+}\right), 2n_+))$$

Proof: Follows from Lemma 2. □

Corollary 1 *Expected welfare regret is* $O(T^{2/3}n^{1/3}n_+^{1/3}(\frac{n_+}{n_-})^{1/3})$ *and thus the average regret goes to* 0 *as* T *goes to infinity.*

Proof: Substituting $\mu = \frac{1}{T}$ and $\tau = T^{2/3}n^{1/3}n_+^{1/3}(\frac{n_+}{n_-})^{1/3}$ which minimizes the regret expression in Theorem 3. □

Simulation Experiments

In this section, we present some interesting simulation results that were obtained using real-world data. We use hourly demand traces from the New England ISO and wind speeds from a set of wind farms in the US. We converted these wind speeds into wind power outputs through a cubic transformation that models wind turbines Hau (2000).

We considered 10 clusters for our experiments, with each cluster having a randomly chosen number of consumers with a mean of 100. The total demand (load) is distributed among these clusters and CPRs are computed for a utility function of type $\alpha\log(x)$, where x represents the demand and α's are uniform. ARs are uniformly distributed between 0 and 1. Mean wind energy $\bar{r}$ is considered as the average of wind energy available at all the slots. The shortage for each time slot t is then computed by taking the difference between observed wind energy r^t and mean wind energy $\bar{r}$. This is essentially the shortfall using a persistence forecast, which is very hard to improve upon in practice Monteiro et al. (2009). The base utility R is computed by taking the average of squared consumed electricity. The experiments are repeated for 100 different samples to ensure sufficient confidence in the simulation results.

Figure 1 depicts the growth of regret with total time T. By Corollary 1, the total regret grows as a function of $T^{2/3}$, and this is validated by our simulations. In Figure 2, we depict the relation between social welfare and shortfall. The shortfall is varied by adding a random noise with variable mean. Figure 2 shows that social welfare can be improved by an impressive factor of 2 by making the offers to large number of consumers when the variance of the shortfall increases. The figure shows that difference in social welfare with and without offers increases with increase in mean cluster size. This is to be expected since the expected number of consumers accepting the offer increases. Thus, the social welfare improves with increased participation and doesn't show saturation effects at reasonable levels. Note that the benefits of our mechanism is more significant when the shortage is higher; incentives should therefore be provided to a larger number of consumers if the shortage is larger.

Conclusion and Future Work

Demand response is critical for renewable integration and energy cost reduction. We proposed a multiarmed bandit mechanism (MAB-MDR) to design incentive offers to consumers who have unknown response characteristics, to reduce costs due to electricity shortage and renewable energy fluctuations. We showed that intuition from crowdsourcing mechanisms is useful, however requires a non-trivial extension due to time varying and quadratic cost function. To ensure truthfulness, MAB-MDR is exploration separated and thus may result in high cost of exploration. Compared to traditional historical data based approaches, which are susceptible to strategic play, the exploration cost is justified but a detailed analysis of the tradeoff between cost and the regret is needed. Analysis of MAB-MDR showed that it is DSIC, IR, and achieves sublinear regret of $O(T^{2/3})$.

This work offers a first step in modeling incentive offers in smart grids to make consumers participate in achieving better demand response thus leading to savings in energy

costs and opens up many challenging directions. Designing a more general mechanism to choose a subset of the members of a cluster or a subset of clusters is the immediate future work. The work can be further extended to many different settings, e.g. where different consumers are allowed to reduce the consumption by different amounts which again can be private information to the consumers. The other interesting direction is situations with online arrival of consumers.

References

Abraham, I.; Alonso, O.; Kandylas, V.; and Slivkins, A. 2013. Adaptive crowdsourcing algorithms for the bandit survey problem. In *COLT*, volume 30 of *JMLR Proceedings*, 882–910. JMLR.org.

Ackermann, T.; Andersson, G.; and Soder, L. 2001. Distributed generation: a definition. *Electric Power Systems Research* 57(3):195 – 204.

Albadi, M., and El-Saadany, E. 2007. Demand response in electricity markets: An overview. In *IEEE Power Engineering Society General Meeting*.

Anderson, R., and Fuloria, S. 2010. On the security economics of electricity metering. *Proceedings of the WEIS*.

Ardakanian, O.; Keshav, S.; and Rosenberg, C. 2011. Markovian models for home electricity consumption. In *Proc. ACM SIGCOMM Green Networking Workshop*, 31–36.

Auer, P.; Cesa-Bianchi, N.; Freund, Y.; and Schapire, R. 2003. The nonstochastic multiarmed bandit problem. *SIAM Journal on Computing* 32(1):48–77.

Auer, P.; Cesa-Bianchi, N.; and Fischer, P. 2002. Finite-time analysis of the multiarmed bandit problem. *Mach. Learn.* 47(2-3):235–256.

Babaioff, M.; Dughmi, S.; Kleinberg, R.; and Slivkins, A. 2012. Dynamic pricing with limited supply. In *EC-2012*, 74–91. ACM.

Babaioff, M.; Kleinberg, R. D.; and Slivkins, A. 2010. Truthful mechanisms with implicit payment computation. In *EC-2010*, 43–52. ACM.

Babaioff, M.; Sharma, Y.; and Slivkins, A. 2009. Characterizing truthful multi-armed bandit mechanisms. In *EC-2009*, 79–88. ACM.

Bhuvaneswari, R.; Edrington, C. S.; Cartes, D. A.; and Subramanian, S. 2009. Online economic environmental optimization of a microgrid using an improved fast evolutionary programming technique. In *North American Power Symposium (NAPS), 2009*, 1 –6.

Bitar, E.; Rajagopal, R.; Khargonekar, P.; Poolla, K.; and Varaiya, P. 2011. Bringing wind energy to market. *IEEE Transactions on Power Systems* 1225 – 1235.

Bubeck, S., and Cesa-Bianchi, N. 2012. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in ML* 5(1):1–122.

Chao, H. 2012. Competitive electricity markets with consumer subscription service in a smart grid. *Journal of Regulatory Economics* 1–26.

Chen, W.; Wang, Y.; and Yuan, Y. 2013. Combinatorial multi-armed bandit: General framework and applications. In *ICML-2013*, volume 28, 151–159.

Devanur, N. R., and Kakade, S. M. 2009. The price of truthfulness for pay-per-click auctions. In *EC-2009*, 99–106.

Even-Dar, E.; Mannor, S.; and Mansour, Y. 2006. Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. *JMLR* 7:1079–1105.

Farhangi, H. 2010. The path of the smart grid. *IEEE Power and Energy Magazine* 8(1):18–28.

Gatti, N.; Lazaric, A.; and Trovò, F. 2012. A truthful learning mechanism for contextual multi-slot sponsored search auctions with externalities. In *EC-2012*, 605–622.

Harris, C. 2006. *Electricity Markets: Pricing, Structures and Economics*. Sussex, England: Wiley.

Hau, E. 2000. *Windturbines*. Springer Berlin etc.

Howe, J. Crowdsourcing: A definition. *http://crowdsourcing.typepad.com/*.

Hsu, Y.-Y., and Su, C.-C. 1991. Dispatch of direct load control using dynamic programming. *Power Systems, IEEE Transactions on* 6(3):1056–1061.

International Energy Agency. 2003. The power to choose - enhancing demand response in liberalised electricity markets findings of IEA demand response project.

Joo, J.-Y., and Ilic, M. D. 2010. A multi-layered adaptive load management (alm) system: Information exchange between market participants for efficient and reliable energy use. In *Transmission and Distribution Conference and Exposition, 2010 IEEE PES*, 1–7. IEEE.

Kalyanakrishnan, S., and Stone, P. 2010. Efficient selection of multiple bandit arms: Theory and practice. In *ICML-2010*.

Kowli, A., and Meyn, S. 2011. Supporting wind generation deployment with demand response. In *IEEE Power and Energy Society General Meeting*, 1 –8.

Langford, J., and Zhang, T. 2007. The epoch-greedy algorithm for multi-armed bandits with side information. In *Advances in neural information processing systems*, 817–824.

Lu, T.; Pál, D.; and Pál, M. 2010. Contextual multi-armed bandits. In *International Conference on Artificial Intelligence and Statistics*, 485–492.

Monteiro, C.; Bessa, R.; Miranda, V.; Botterud, A.; Wang, J.; Conzelmann, G.; et al. 2009. Wind power forecasting: state-of-the-art 2009. Technical report, Argonne National Laboratory (ANL).

Narahari, Y. 2014. *Game Theory and Mechanism Design*. IISc Press and World Scientific. chapter 18: Vickrey-Clarke-Groves (VCG) Mechanisms.

QDR, Q. 2006. Benefits of demand response in electricity markets and recommendations for achieving them.

Singla, A., and Krause, A. 2013. Truthful incentives in crowdsourcing tasks using regret minimization mechanisms. In *WWW-2013*, 1167–1178.

Sood, V.; Fischer, D.; Eklund, J.; and Brown, T. 2009. Developing a communication infrastructure for the smart grid. In *Electrical Power & Energy Conference (EPEC), 2009 IEEE*, 1–7. Ieee.

Strbac, G. 2008. Demand side management: Benefits and challenges. *Energy Policy* 36(12):4419 – 4426.

Tran-Thanh, L.; Chapman, A. C.; Rogers, A.; and Jennings, N. R. 2012. Knapsack based optimal policies for budget-limited multi-armed bandits. In *AAAI*.

Zhu, T.; Mishra, A.; Irwin, D.; Sharma, N.; Shenoy, P.; and Towsley, D. 2011. The case for efficient renewable energy management for smart homes. *Proceedings of the Third Workshop on Embedded Sensing Systems for Energy-efficiency in Buildings (BuildSys)*.