# A Multiarmed Bandit Based Incentive Mechanism for a Subset Selection of Customers for Demand Response in Smart Grids

**Shweta Jain,**[1] **Sujit Gujar**[2]
[1]Indian Institute of Technology, Ropar, [2]International Institute of Information Technology, Hyderabad.
shwetajain@iitrpr.ac.in, sujit.gujar@iiit.ac.in

## Abstract

Demand response is a crucial tool to maintain the stability of the smart grids. With the upcoming research trends in the area of electricity markets, it has become a possibility to design a dynamic pricing system, and consumers are made aware of what they are going to pay. Though the dynamic pricing system (pricing based on the total demand a distributor company is facing) seems to be one possible solution, the current dynamic pricing approaches are either too complex for a consumer to understand or are too naive leading to inefficiencies in the system (either consumer side or distributor side). Due to these limitations, the recent literature is focusing on the approach to provide incentives to the consumers to reduce the electricity, especially in peak hours. For each round, the goal is to select a subset of consumers to whom the distributor should offer incentives so as to minimize the loss which comprises of cost of buying the electricity from the market, uncertainties at consumer end, and cost incurred to the consumers to reduce the electricity which is a private information to the consumers. Due to the uncertainties in the loss function (arising from renewable energy resources as well as consumption needs), traditional auction theory-based incentives face manipulation challenges. Towards this, we propose a novel combinatorial multi-armed bandit (MAB) algorithm, which we refer to as GLS-MAB to learn the uncertainties along with an auction to elicit true costs incurred by the consumers. We prove that our mechanism is regret optimal and is incentive compatible. We further demonstrate efficacy of our algorithms via simulations.

## Introduction

A *Smart Grid* is an electricity network that can intelligently adapt according to the behaviour and actions of all the users connected to it (def 2010). The connected users can be categorized into distributors, producers, consumers, and prosumers (that do both, producer as well as consumer). Due to intelligent communication monitoring, the smart grid enables two-way information and power exchange between different connected users. The smart grid allows the integration of new communication (Farhangi 2010) that induces increased participation of prosumers to make informed decisions on consuming/producing the electricity appropriately. This participation will allow consumers to be more responsive to time-varying grid conditions, including peak demand and renewable excess or shortfall. Typically the consumers interact with the smart grids through software agents with capabilities of taking decisions on the behalf of the consumers (or prosumers). Thus, unlike the traditional electricity networks, the smart grid offers a promising future towards providing more reliable service to its users. The primary question we ask in this paper is, can we manage our electricity smartly? We answer this through designing a *demand response program*.



Demand response is a critical part of the smart grids, which refers to the change in consumers' energy usage behavior with respect to the pricing signals from the distributor company. An important point to note is that when it comes to resources like water or electricity, it is the government's responsibility to supply the required electricity; thus, the option left to the government is either to install more expensive generators to match the demand and supply or to make the consumers voluntarily optimize their electricity load by providing some monetary incentives. When taking the second approach, the major challenge is that different consumers may behave differently with the given incentives. Thus, in order to increase participation from the consumers, it becomes crucial to learn their reaction towards these incentives. Learning the customers' behavior is challenging due to uncertainty and randomness involved with the integrated renewable energy resources on the smart grid and on prosumer side. We intend to learn these uncertainties over time using multi-armed bandit approach.

We have the following demand response problem. There are different consumers integrated into the smart grid network. Henceforth, we call consumers/prosumers as well as the software agents acting on behalf of them *agents*. As each agent derives a certain value from a unit (KWh) of electricity, it incurs a cost per unit reduction (CPR) when asked to reduce a unit of electricity. Hence the distributor company must ensure that when giving an offer to the agent, it matches its CPR. Since the CPRs are private, a simple mechanism to ask CPRs from the agents may not work as the strategic agents may report higher CPR in order to receive higher monetary offers. In-spite of monetary

incentives higher than its CPR, an agent may not be able to reduce the consumption due to stochasticity involved in production/consumption of the electricity at its end. This stochasticity may be due to uncertainty in renewable energy resources or some other factors like the arrival of sudden guests at residential house or sudden workload in industry. We model these uncertainties via acceptance rates (AR) of the agents. At each round, when the distributor company faces a shortage, we wish to select a subset of agents whom the distributor company asks to reduce the consumption of electricity in order to minimize its losses. The following are our main contributions:

### Contributions

- We show that the problem of maximizing the efficiency maps to a subset selection combinatorial optimization problem to minimise a non-monotone supermodular loss function which in general can be a hard problem.
- We propose a polynomial-time algorithm, namely GLS that leads to local optimal solution for the proposed demand response (Theorem 2).
- We employ the techniques from Multi-Armed Bandit (MAB) literature to learn ARs (GLS-MAB). We prove that GLS-MAB is regret optimal with respect to the local optima (Theorem 3).
- With the help of GLS-MAB, we design a truthful mechanism to elicit CPRs from the agents (Theorem 6).
- We show via simulations that we attain sub-linear regret with respect to a global optima as well.

## Related Work

The existing demand response methods like time-of-day tariff (Akasiadis et al. 2015; Ramchurn et al. 2011; Robu et al. 2016; Jain et al. 2013), real-time pricing (Chao 2012), critical peak pricing (Zhang, Wang, and Fu 2009; Park et al. 2015), direct load control (Hsu and Su 1991), demand bidding (Anderson and Fuloria 2010) are either too complex or not very efficient. To solve these issues researchers have explored giving incentives to the consumers to reduce their electricity consumption.

A similar MAB mechanism for demand response is considered in (Jain, Narayanaswamy, and Narahari 2014), where the agents who share the same parameters are clustered together. At each time, one cluster is chosen for demand response and all the consumers in that cluster are asked to reduce the consumption by one unit. This is a naive approach as it may be optimal to select more than one clusters thus leading to a combinatorial problem. A subset selection problem for demand response is considered in (Ma et al. 2016; Ma, Parkes, and Robu 2017) where reward (penalty) is imposed on the agents if they accept the offer and reduce (not reduce) the consumption by eliciting the ARs. Along similar lines, (Methenitis, Kaisers, and La Poutré 2019) proposed mechanisms where the probability of accepting the offer (acceptance rate, AR) as well the costs are elicited. In practical settings, ARs may not be known to the agents and should be learnt over a period of time. A combinatorial MAB approach considered in (Li, Hu, and Li 2018) learns ARs but does not capture the CPRs and the strategic behaviour of the agents. In summary, none of the existing works have considered learning ARs as well as eliciting true CPRs in a combinatorial MAB framework. In addition, none of the above papers model the loss due to not meeting the demand reduction explicitly together with the cost of incentives in the case of high demand reduction.

## Mathematical Model

There are $N = \{1, 2, \ldots, n\}$ *agents*, that can prepare for a demand response when given the right incentives. Each agent incurs a fixed cost for reducing one unit of electricity. We call this fixed cost as CPR (cost per unit reduction) denoted by $c_i$ for agent $i$ which is the private information to the agent. The CPRs depend on consumers' priorities, for example, industrial consumers may have more CPR as compared to domestic consumers. Even if there are slight variations in the CPR based on the time slots, the consumers can choose to bid their mean CPR of all the time slots. Even if an agent commits to the reduction, the agent may fail to do so due to some external uncertain events like some emergency situation or failing to generate expected electricity due to renewable resources at their end. We model these uncertainties as Acceptance Rate (AR), $p_i$, which denotes the probability of accepting the offer by an agent $i$. At each time, which we denote as round $t$, the distributor company faces a shortage of electricity denoted by $D_t$ which may be positive, negative, or zero. The goal of the distributor company is to select the subset of the agents to whom it asks to reduce electricity and offer them the proper incentives so that it is in their best interest to reduce the electricity. If the distributor company fails to satisfy the demand $D_t$, then it would have to buy from the market. Buying electricity generally leads to quadratic loss (Li, Hu, and Li 2018) which is given as: $L_t(S_t) = \left(\sum_{i \in S_t} X_{t,i} - D_t\right)^2$. Here, $X_{t,i}$ is a binomial random variable with parameter $p_i$ which is 1, when the agent $i$ reduces one unit of electricity at round $t$ and is 0 otherwise. For every round $t$, our demand response problem is to select a subset that minimizes the expected loss and the expected cost from the agents.

$$\mathbb{E}L(S_t) = \mathbb{E}\left[CL_t(S_t) + \sum_{i \in S_t} X_{t,i}c_i\right] = C\mathbb{E}L_t(S_t) + \sum_{i \in S_t} p_i c_i$$
$$= C\left(\sum_{i \in S_t} p_i - D_t\right)^2 + C\sum_{i \in S_t} p_i(1 - p_i) + \sum_{i \in S_t} p_i c_i$$

Here $C$ is the cost that the distributor company incurs for the loss to buy electricity from the market. We minimize expected cost of the agents as oppose to maximizing rewards as we consider the problem of maximizing social welfare due to the following reasons:

- Electricity is a social good
- Revenue of the efficiency maximizing mechanism with $k + 1$ bidders is no less than the revenue of the revenue-maximizing mechanism with $k$ bidders (Bulow and Klemperer 1996).

- Social welfare maximizing mechanisms are simple to understand by the consumers.

### The Demand Response Mechanism

We propose a two phase demand response mechanism that goes for $T$ rounds. At each round $t$, the following happens:
**Phase 1:**

- Agents choose to report $c_i$ with bids $\hat{c}_i^t$. We assume that true CPRs of agents do not change across rounds, however their bids may as the bids depend on offer at each round.
- The distributor company observes the shortage $D_t$.
- With $\hat{c}_i^t$s, $D_t$, and the learnt acceptance rates from the past history, the distributor company prepares a demand response comprising of offer price $r_i^t$.
- Based on $r_i^t$s, agents decide whether to accept the offer to reduce the consumption by one unit or not.

**Phase 2:**

- Each agent who accepted the offer, observes their random variable $X_{t,i}$ generated based on $p_i$. If $X_{t,i}$ is 1 then they reduce the consumption by one unit.
- For each of selected agents, distributor companies pays $r_i^t$ if it reduces the consumption and 0 if it fails to reduce.

Note that unlike previous works, we do not wish to impose the penalty on the agents if they fail to reduce the consumption. Instead, we would like to propose $r_i^t$ in such a way that it is in best interest for the agent to reduce the electricity each round if it can do so. Our goal is to design a learning algorithm to optimize $\mathbb{E}L(S_t)$, a supermodular function, and design a mechanism to determine payments $r_i^t$ to ensure truthful reports of CPRs.

### Challenges

- ARs need to be learnt across rounds.
- Proper incentives need to be given to elicit true CPRs.
- Even when ARs and CPRs are known, this problem is hard due to combinatorial nature.

Towards solving these challenges, we start with perfect information setting where ARs and CPRs are known and design a polynomial time algorithm GLS for this setting. We next move to unknown ARs model where we extend GLS to GLS-MAB using the MAB techniques. When CPRs are not known, we prove that our algorithm GLS-MAB is stochastic monotone and hence the rewards can be constructed so as to get incentive compatible and individual rational mechanism.

## Perfect Information Setting

We first show that the given problem of obtaining global optima $S_t^{**}$ at round $t$ when ARs and CPRs are known is a minimization of a non-monotone supermodular function. Therefore it may not be possible to even approximate the minimum of expected loss function within any factor.

**Lemma 1** *$\mathbb{E}L(S_t)$ is non-monotone and supermodular.*

**Proof:** Consider any sets $A \subset B \subset [n]$ and $j \notin B$ then:

$$\mathbb{E}L(A \cup \{j\}) - \mathbb{E}L(A) = p_j(2C\sum_{i\in A} p_i + c_j + C - 2CD_t)$$

$$< p_j(2C\sum_{i\in B} p_i + c_j + C - 2CD_t) = \mathbb{E}L(B \cup \{j\}) - \mathbb{E}L(B)$$

Non-monotonicity is easy to verify as $\sum_{i\in A} p_i - D_t$ can be positive or negative. □

**Theorem 1** *(Mittal and Schulz 2013) Let $f : 2^S \to \mathbb{Z}_+$ be a supermodular function defined over the subsets of $S$. Then it is not possible to approximate the minimum of $f$ to within any factor, unless P = NP.*

Though the above result does not directly imply that our problem is also inapproximable as it is real valued function, it is indicative that ours problem might be difficult to approximate. Due to the above result, we look for a local optimal solution $S_t^*$ as oppose to $S_t^{**}$. Our first Lemma characterizes the agents that should be present in $S_t^*$

**Lemma 2** *1.* $\frac{c_j}{2} < C(-\sum_{i\in S_t^*\setminus\{j\}} p_i + D_t - 1/2)\ \forall j \in S_t^*$

*2.* $\frac{c_j}{2} > C(-\sum_{i\in S_t^*} p_i + D_t - \frac{1}{2})\ \forall j \in [n] \setminus S_t^*$

**Proof:**

1. $\mathbb{E}L(S_t^*) - \mathbb{E}L(S_t^* \setminus \{j\}) < 0$
   $\implies p_j(2C\sum_{i\in S_t^*} p_i - 2CD_t - 2Cp_j + C + c_j) < 0$
   $\implies c_j/2 < C(-\sum_{i\in S_t^*\setminus\{j\}} p_i + D_t - 1/2)$
2. Suppose $\exists j \in [n]\setminus S_t^*$ such that $c_j/2 < C(-\sum_{i\in S_t^*} p_i + D_t - 1/2)$, then: $\mathbb{E}L(S_t^* \cup \{j\}) - \mathbb{E}L(S_t^*) = p_j(2C\sum_{i\in S_t^*} p_i + c_j + C - 2CD_t) < 0$ which leads to the contradiction that $S_t^*$ is local optima. □

**Corollary 1** *If $\frac{c_j}{2} > C(D_t - 1/2)$ then $j \notin S_t^*$.*

### Greedy Local Search (GLS) Algorithm

With the help of the above characterization result, we now present a polynomial time algorithm to produce the set $S_t^*$ for each round $t$ in Algorithm 1. As the algorithm searches for a local optimal greedily, we refer to it as *Greedy Local Search* (GLS). GLS arranges and renumber the agents in the decreasing order of $Cp_i - c_i/2$ and keep adding an agent $j$ if it satisfies Lemma 2. The next set of results proves the correctness of the algorithm. The following Lemma proves that when more agents are added in the set, Lemma 2 will be still satisfied for the agents that are already added and thus, they would remain in $S_t^*$.

**Lemma 3** *If $c_j/2 < C(-\sum_{i<j, i\in S_t^*} p_i + D_t - 1/2)$ then $c_i/2 < C(-\sum_{k<j, k\in S_t^*\setminus\{i\}} p_k + D_t - 1/2 - p_j),\ \forall i < j$*

**Proof:** We prove via induction: We have $c_1/2 < C(D_t - 1/2)$ (if such an agent is not present then optimal set is $\emptyset$ due to Corollary 1). Now let us try to add second agent such that $c_2/2 < C(-p_1 + D_t - 1/2)$. $Cp_1 - c_1/2 > Cp_2 - c_2/2 \implies c_1/2 < C(-p_2 + D_t - 1/2)$. For induction hypothesis, assume: $c_i/2 < C(-\sum_{k\le j, k\in S_t^*\setminus\{i\}} p_k + D_t - 1/2)$. Now, consider agent $j+1$ such that: $c_{j+1}/2 < C(-\sum_{k\le j, k\in S_t^*} p_k + D_t - 1/2)$ and $Cp_i - c_i/2 > Cp_{j+1} - c_{j+1}/2\ \forall i < j, i \in S_t^*$.

**Algorithm 1:** Greedy Local Search (GLS) Algorithm

**Input:** CPRs, $\{c_1, c_2, \dots c_n\}$, ARs, $\{p_1, p_2, \dots p_n\}$, Shortage at round $t$, $D_t$
**Output:** Optimal subset at round $t$, $S_t^*$

$S_t^* = \emptyset$;
Eliminate agents with high cost: $c_j/2 > C(D_t - 1/2)$;
Arrange remaining agents in the order $Cp_i - \frac{c_i}{2}$. Let the agents be numbered as $1, 2, 3, \dots, n_1$;
**for** $i \leftarrow 1$ ***to*** $n_1$ **do**
  **if** $c_i/2 < -C\sum_{j \in S_t^*} p_j + CD_t - C/2$ **then**
    $S_t^* = S_t^* \cup \{i\}$

**Proof of Induction:** Pick any agent $i < j, i \in S_t^*$: $Cp_i - c_i/2 > Cp_{j+1} - c_{j+1}/2 \implies c_i/2 < -Cp_{j+1} - C\sum_{k \le j, k \in S_t^* \setminus \{i\}} p_k + CD_t - C/2$. □

We next prove that if at any round, we have skipped an agent, it can never become a part of $S_t^*$.

**Lemma 4** *If* $c_k/2 > C(-\sum_{i:i<k \text{ AND } i \in S_t^*} p_i + D_t - 1/2)$ *then* $c_k/2 > C(-\sum_{i \in S_t^*} p_i + D_t - 1/2)$.

**Proof:** $c_k/2 > C(-\sum_{i:i<k \text{ AND } i \in S_t^*} p_i + D_t - 1/2) \implies c_k/2 > C(-\sum_{i:i<k \text{AND } i \in S_t^*} p_i + D_t - 1/2 - \sum_{i \in S'} p_i)\ \forall S' \subset [n]$ □

From the above results, we have the following Theorem:

**Theorem 2** *The set $S_t^*$ obtained from GLS is a local optima.*

## Imperfect Information Setting: Unknown ARs

Motivated by UCB algorithm for MAB, we present GLS-MAB (Algorithm 2) to learn the ARs of the agents. Let us denote $\hat{p}_i^+ = \hat{p}_i + \sqrt{\frac{2\ln t}{n_i(t)}}$ and $\hat{p}_i^- = \hat{p}_i - \sqrt{\frac{2\ln t}{n_i(t)}}$ as upper confidence and lower confidence bound on $p_i$ at round $t$. $\hat{p}_i$ denotes the learnt probability which is given as $\hat{p}_i = \frac{\sum_{t'=1}^{t} \mathbb{1}(X_{t',i}=1)}{n_i(t)}$, where $n_i(t)$ denotes the number of times the agent is given the offer till round $t$. The agents are then arranged in the order of $C\hat{p}_i^+ - \frac{c_i}{2}$ that ensures that the agents are given the chance optimistically.

### Regret Analysis for GLS-MAB

Regret of a learning algorithm is the difference in the loss achieved by the algorithm and the loss incurred by the optimal subset if the probabilities were known. In our setting the optimal subset may differ every round. We derive the regret of our algorithm with respect to GLS that finds a local optima $S_t^*$. In simulations section, we empirically show that the regret with respect to $S_t^{**}$ is not much. Typically the regret at round $t$ w.r.t. a global optimal solution is: $R_t^G = \mathbb{E}L(S_t) - \mathbb{E}L(S_t^{**})$. However, as finding $S_t^{**}$ is computationally hard, we define regret as $R_t = \mathbb{E}L(S_t) - \mathbb{E}L(S_t^*)$, where $S_t^*$ is the solution returned by GLS with known ARs. The overall regret of the algorithm is given as: $R(T) = \sum_{t=1}^{T} R_t$ ($R^G(T) = \sum_{t=1}^{T} R_t^G$). We look for the algorithms which gives sub-linear regret with respect to the total number of rounds $T$. Let us denote $\Delta_{ij} = |Cp_i - c_i/2 - (Cp_j - c_j/2)|$ and let $\Delta = \min_i \Delta_{ij}$.

**Algorithm 2:** GLS-MAB: GLS For unknown ARs

**Input:** CPRs, $\{c_1, c_2, \dots c_n\}$, Total number of rounds $T$
**Output:** Sequence of allocations $S_1, S_2, \dots S_T$

$S_1 = [n]$ i.e. make offer to everybody to get certain estimates on AR, $n_i(1) = 1\ \forall i$;
**for** $t \leftarrow 2$ ***to*** $T$ **do**
  Observe $X_{i,t-1}\ \forall i$, Shortage $D_t$;
  Update Estimated ARs, upper confidence bounds on ARs and lower confidence bounds on ARs as follows: $\hat{p}_i = \frac{\sum_{t'=1}^{t-1} X_{i,t'}}{n_i(t-1)}$, $\hat{p}_i^+ = \hat{p}_i + \sqrt{\frac{2\ln t}{n_i(t-1)}}$, and $\hat{p}_i^- = \hat{p}_i - \sqrt{\frac{2\ln t}{n_i(t-1)}}$ respectively.;
  Eliminate agents $j$ s.t. $c_j/2 > C(D_t - 1/2)$;
  Out of remaining agents, renumber the agents in the order $C\hat{p}_i^+ - \frac{c_i}{2}$;
  $S_t = \emptyset$;
  **for** $i \leftarrow 1$ ***to*** $n'$ **do**
    **if** $c_i/2 < -C\sum_{j \in S_t} \hat{p}_j^- + CD_t - C/2$ **then**
      $S_t = S_t \cup \{i\}$

**Theorem 3** *The regret of GLS-MAB is $O(\sqrt{T})$.*

**Proof:** Let the agents are numbered in the order of $p_i - c_i/2$ i.e. $Cp_1 - c_1/2 \ge Cp_2 - c_2/2 \ge Cp_3 - c_3/2 \ge \dots Cp_n - c_n/2$. We prove the theorem through a series of lemmas:

**Lemma 5** *If* $n_i(T) \ge \frac{8C^2 \ln T}{\Delta^2}\ \forall i \in [n]$, *then* $\forall i, j \in [n]$: $Cp_i - c_i/2 > Cp_j - c_j/2 \implies C\hat{p}_i^+ - \frac{c_i}{2} > C\hat{p}_j^+ + \frac{c_j}{2}$ *and* $C\hat{p}_i^- - \frac{c_i}{2} > C\hat{p}_j^- - \frac{c_j}{2}$ *with high probability.*

**Proof:** Suppose not, then we have:

$$C\hat{p}_i + C\sqrt{\frac{2\ln t}{n_i(t)}} - \frac{c_i}{2} < C\hat{p}_j + C\sqrt{\frac{2\ln t}{n_j(t)}} - \frac{c_j}{2}$$

$$\text{or } C\hat{p}_i - C\sqrt{\frac{2\ln t}{n_i(t)}} - \frac{c_i}{2} < C\hat{p}_j - C\sqrt{\frac{2\ln t}{n_j(t)}} - \frac{c_j}{2}$$

$$\implies Cp_i - \frac{c_i}{2} < Cp_j + 2C\sqrt{\frac{2\ln t}{n_j(t)}} - \frac{c_j}{2}$$

$$\text{or } Cp_i - 2C\sqrt{\frac{2\ln t}{n_i(t)}} - \frac{c_i}{2} < Cp_j - \frac{c_j}{2}$$

which is a contradiction since $n_k(t) \ge \frac{8C^2}{\Delta^2} \ln T\ \forall k$. □

**Corollary 2** *After each agent has been selected for $\frac{8C^2}{\Delta^2} \ln T$ rounds, we will have correct ordering on the agents with respect to UCB bounds.*

**Lemma 6** *If* $c_j/2 < -C\sum_{i<j, i \in S_t} \hat{p}_i^- + CD_t - C/2$ *and* $n_j(t) > \frac{8C^2}{\Delta^2} \ln T\ \forall j \in [n]$ *then* $c_i/2 < -C\sum_{k<j, k \in S_t \setminus \{i\}} \hat{p}_k^- + CD_t - C/2 - C\hat{p}_j^-,\ \forall i < j$

**Proof:** The proof follows similar steps as in Lemma 3 with the use of Lemma 5. □

For ease of notation, let $D' = D_t - 1/2$ and let us denote $X_i^t = 1$ if $i \in S_t^*$ and $Y_i^t = 1$ if $i \in S_t$.

**Lemma 7** *If $i_1$ is the first index s.t. $\forall i < i_1$ $X_i^t = Y_i^t$ and $X_{i_1}^t \neq Y_{i_1}^t$ then either $X_{i_1}^t = 0$ i.e. $i_1 \in S_t$ or $\exists j \in [n]$ such that $n_j(t) < \frac{8C^2}{\Delta^2} \ln T$*

**Proof:** If $X_{i_1}^t = 1$ and $\forall j \in [n]$, $n_j(t) \geq \frac{8C^2}{\Delta^2} \ln T$ then $c_{i_1}/2 \leq C(-\sum_{j:j<i_1 \text{AND} j \in S_t^*} p_j + D') \leq C(-\sum_{j:j<i_1 \text{AND} j \in S_t} \hat{p}_j^- + D')$. However then, $Y_{i_1}^t = 1$ which leads to the contradiction to $X_{i_1}^t \neq Y_{i_1}^t$. □

**Lemma 8** *$\forall i > i_1$, if $Y_i^t = 0$ and $n_j(t) > \frac{8C^2}{\Delta^2} \ln T \; \forall j \in [n]$ then $X_i^t = 0$ i.e. the algorithm will not miss any optimal agents after selecting $i_1$ in $S_t$.*

**Proof:** $Y_{i_1}^t = 1 \implies c_{i_1}/2 \leq C(-\sum_{j \in S_t} \hat{p}_j^- + D')$
$X_{i_1}^t = 0 \implies c_{i_1}/2 > C(-\sum_{j<i_1 AND j \in S_t^*} p_j + D')$
$Y_i^t = 0 \implies c_i/2 > C(-\sum_{j \in S_t} \hat{p}_j^- + D') > C(-\sum_{j<i_1 AND j \in S_t^*} p_j + D') \implies X_i^t = 0 \; (i > i_1)$.

Thus, $S_t^* \subset S_t$ after all the agents are selected for at-least $\frac{8C^2 \ln(T)}{\Delta^2}$ rounds. We next bound the number of sub-optimal agents present in $S_t$.

**Lemma 9** *For any $i_1 < i$, if $n_j(t) > \frac{8n^2C^2}{\Delta^2} \ln T \; \forall j \in S_t^*$ and $n_j(t) > \frac{8C^2}{\Delta^2} \ln(T) \; \forall j \in [n]$ and $i \in S_t$ then $i \in S_t^*$ i.e. we will not add more sub-optimal agents in our selected set.*

**Proof:** Consider the first agent $i > i_1$ such that $i \in S_t$ but $i \notin S_t^*$. Thus, till agent $i$, we have $S_t = S_t^* \cup i_1$. If $i \notin S_t^*$ then $c_i/2 > CD' - C(\sum_{j<i \text{And} j \in S_t^*} p_j)$ Also, $c_i/2 < CD' - C(\sum_{j<i \text{And} j \in S_t^*} \hat{p}_j^-) - C\hat{p}_{i_1}^- < CD' - C(\sum_{j<i \text{And} j \in S_t^*} p_j) + \Delta - C\hat{p}_{i_1}^- \implies \hat{p}_{i_1}^- < \frac{\Delta}{C}$

The last condition is not possible if $\Delta$ is small enough and $p_{i_1}$ is large enough. Thus, we will get regret only due to the agent $i_1$. If $i_1 \in S_t$ but $i_1 \notin S_t^*$ implies that $c_i/2 > C(D - 1/2 - \sum_{j<i_1} p_j)$ but $c_i/2 < C(D - 1/2 - \sum_{j<i_1} \hat{p}_j^-)$. Then, the regret contributed due to agent $i_1$ is $\mathbb{E}L(S_t^* \cup i_1) - \mathbb{E}L(S_t^*) = 2p_{i_1}(C\sum_{i \in S_t^*} p_i + c_{i_1}/2 + CD') < 2p_{i_1}(C\sum_{i \in S_t^*} p_i - C\sum_{i \in S_t^*} \hat{p}_i^-) < 2np_{i_1}\sqrt{\frac{2\ln t}{n_i(t)}} \; i \in S_t^*$.

**Proof of Theorem 3:** An agent $i \neq i_1$ contributes to regret if:

1. $i \in S_t$ and $i \notin S_t^*$: This happens only when $\exists j \in S_t^*$ s.t. $Cp_i - c_i/2 < Cp_j - c_j/2$ but $C\hat{p}_i^+ - c_i/2 > C\hat{p}_j^+ - c_j/2$. This event has low probability if $n_i(t) \geq \frac{8C^2}{\Delta^2} \ln(T)$.
2. $i \in S_t^*$ and $i \notin S_t$: This can only happen if $\exists j \in S_t$ s.t. $Cp_i - c_i/2 > Cp_j - c_j/2$ but $C\hat{p}_i^+ - c_i/2 < C\hat{p}_j^+ - c_j/2$. Again this will not happen after $n_j(t) \geq \frac{8C^2}{\Delta^2} \ln(T)$.

Thus, after $n\frac{8C^2}{\Delta^2} \ln(T)$ above cases will not occur. Let $T' = n^2 \frac{8C^2}{\Delta^2} \ln(T)$. The total regret is given as:

$$\max_{i \in S_t} T' + 2np_{i_1} \sum_{t>T'}^{T} \sqrt{\frac{2\ln t}{n_i(t)}} \leq T' + p_{i_1}\Delta \sum_{t=T'}^{T} \sqrt{\frac{1}{t}}$$

$$\leq T' + p_{i_1}\Delta\sqrt{T} = O(\sqrt{T}) \square$$

## Imperfect Information Setting: Unknown CPRs

While eliciting the unknown CPRs from the agents, the agents may be strategic and may not reveal their true CPRs in order to maximize their own utilities. We first formalize the game theoretic notions in this section. Let us denote the CPR bid profile by $\hat{c} = \{\hat{c}_1, \hat{c}_2, \ldots, \hat{c}_n\}$. We wish to design a mechanism $\mathcal{M} = (\mathcal{S}(\hat{c}), \mathcal{R}(\hat{c}))$ which consists of an allocation rule, $\mathcal{S} = \{S_1, S_2, \ldots, S_T\}$ where each $S_t$ is the selected subset of agents at round $t$, and a reward rule $\mathcal{R} = \{\mathcal{R}_1, \mathcal{R}_2, \ldots, \mathcal{R}_T\}$ where each $\mathcal{R}_t = \{r_1^t, r_2^t, \ldots, r_n^t\}$ represents the offer price. Note that the selected subset at any round $t$ will depend on the bid profile, but we will not explicitly mention this dependence every time. The expected cost per unit reduction of any agent $i$ is $p_i c_i$. Thus given the reward $r_i^t$, the utility of a agent at round $t$ is given as: $\mathbb{E}[u_i(S_t(\hat{c}_i, \hat{c}_{-i}), c_i)] = p_i(-\mathbb{1}(i \in S_t)c_i + r_i^t)$ , where $\hat{c}_{-i}$ is the bid profile of all the agents other than $i$. The utility for the distributor company is given as: $\mathbb{E}[u_C(S_t(\hat{c}_i, \hat{c}_{-i}), c_i)] = -C(D_t - \sum_{i \in S_t} p_i)^2 - C\sum_{i \in S_t} p_i(1 - p_i) - \sum_{i \in S_t} p_i r_i^t$. The expected social welfare (sum of the expected valuations) is $W(\mathcal{S}) = -\sum_{t=1}^{T} C(D_t - \sum_{i \in S_t} p_i)^2 - \sum_{t=1}^{T} C\sum_{i \in S_t} p_i(1-p_i) - \sum_{t=1}^{T}\sum_{i \in S_t} p_i c_i$. We now define some of the desirable game theoretic properties that we would like mechanism $\mathcal{M} = (\mathcal{S}, \mathcal{R})$ to satisfy:

**Definition 1** *Allocative Efficiency(AE): An allocation rule $\mathcal{S}$ is allocatively efficient if it maximizes the social welfare.*

It is easy to see that allocation rule produced by GLS is allocative efficient. Further, when ARs are not known, the regret in social welfare due to GLS-MAB is given in Theorem 3. We now provide some definitions that are required to elicit the CPRs truthfully from the agents.

**Definition 2** *Individual Rationality (IR): Mechanism $\mathcal{M}$ is IR for an agent if truthful bidding results in positive utility.*

If the mechanism is IR than it is always in the best response for the agent to reduce the electricity if he can as failing to reduce the consumption will lead to utility of zero.

**Definition 3** *Dominant Strategy Incentive Compatible (DSIC): Mechanism $\mathcal{M}$ is called DSIC if for each agent bidding its true cost maximizes its utility irrespective of the bids of other agents for every round.*

A mechanism is DSIC iff the allocation rule is monotone and the payment satisfies certain property (Myerson 1981). For now, we focus on the monotone allocation rule. The monotonicity is defined as:

**Definition 4 (Monotone Allocation Rule)** *Consider two bids, $\hat{c}_i$ and $\hat{c}_i^-$ for agent $i$ with $\hat{c}_i \geq \hat{c}_i^-$. An allocation rule $S$ is monotone if $i \in S^t(\hat{c}_i, \hat{c}_{-i}) \implies i \in S^t(\hat{c}_i^-, \hat{c}_{-i}) \; \forall i \; \forall t$.*

The next result shows that the allocation returned by GLS results in the monotone allocation rule.

**Theorem 4** *Let $S(p,c)$ denote the subset returned by the GLS with ARs $p = \{p_1, p_2, \ldots, p_n\}$ and CPRs $c = \{c_1, c_2, \ldots, c_n\}$. Let $c^{-(i)} = \{c_1, c_2, \ldots, c_i - \epsilon, \ldots, c_n\}$ be another CPR profile, where CPR of other agents remain same but the CPR of $i$ is reduced. Then $i \in S^t(p,c) \implies i \in S^t(p, c^{-(i)})\ \forall t$.*

**Proof:**
Assume that with $c$, agent $i$ was ranked $k$ in the ordering with respect to $p_j - c_j/2$ and with cost profile $c^{-(i)}$, he is ranked $k^-$. Then, it is easy to see that $k^- < k$. If $i \notin S^t(p, c^{-(i)})$ then
$\frac{c_i - \epsilon}{2} > C(-\sum_{j<k^- \text{AND} j \in S^t(p,c^{-(i)})} p_j + D_t - 1/2)$
$\implies \frac{c_i}{2} > C(-\sum_{j<k^- \text{AND} j \in S^t(p,c^{-(i)})} p_j + D_t - 1/2)$
$> C(-\sum_{j<k \text{AND} j \in S^t(p,c)} p_j + D_t - 1/2)$. Thus, $i \notin S^t(p,c)\}$. □

Achieving DSIC is a very strong condition in learning environment and thus we look for ex-post monotonicity or stochastic monotonicity. We will show that GLS-MAB is stochastic monotone which is monotonicity in expectation, where expectation is taken over the randomness of the acceptance rate of the agents.

**Definition 5 (Stochastic Monotone Allocation Rule)** *Consider any two bids for agent $i$, $\hat{c}_i$ and $\hat{c}_i^-$ such that $\hat{c}_i \geq \hat{c}_i^-$. An allocation rule $S$ is called stochastic monotone if for every agent $i$, we have if $\mathbb{E}[\sum_{t=1}^{T} \mathbb{1}(i \in S^t(\hat{c}_i, \hat{c}_{-i}))] \leq \mathbb{E}[\sum_{t=1}^{T} \mathbb{1}(i \in S^t(\hat{c}_i^-, \hat{c}_{-i}))]$.*

In order to prove that the allocation is stochastic monotone, we use the property of *Independence of Irrelevant Alternatives (IIA)* which we define below.

**Definition 6** *Independence of Irrelevant Alternatives (IIA): At any round $t$, if the estimates on AR of all the agents are fixed other than $i$, than change in the estimates on AR of $i$ should not transfer allocation from agent $j$ to agent $l$.*

**Lemma 10** *GLS-MAB satisfies IIA property.*

**Proof:** Let $j$ and $l$ be the two agents such that $j \in S^t(\hat{p}_i, \hat{p}_{-i})$ and $l \notin S^t(\hat{p}_i, \hat{p}_{-i})$ with the estimate of agent $i$ as $\hat{p}_i$ till round $t$. This means the following should hold: $c_j/2 \leq -C\sum_{k \in S^t(\hat{p}_i, \hat{p}_{-i})} \hat{p}_k^+ + CD_t - C/2$ and
$c_l/2 > -C\sum_{k \in S^t(\hat{p}_i, \hat{p}_{-i})} \hat{p}_k^- + CD_t - C/2$.

Let us denote the changed estimates of agent $i$ by $\hat{p'}_i$. For ease of notation, denote $S^t = S^t(\hat{p}_i, \hat{p}_{-i})$ $S^{t'} = S^t(\hat{p'}_i, \hat{p}_{-i})$. Then we will prove: if $j \notin S^{t'}$ then $l \notin S^{t'}$.

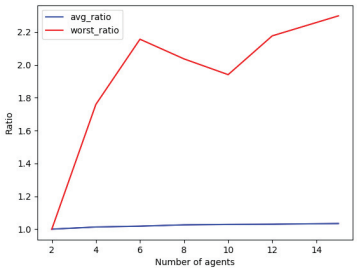
$$j \notin S^{t'} \implies c_j/2 > C(-\sum_{k \in S^{t'}} \hat{p}_k^+ + D_t - 1/2)$$
$$\implies -C\sum_{k \in S^{t'}} \hat{p}_k^+ + CD_t - C/2 < -C\sum_{k \in S^t} \hat{p}_k^+ + CD_t - C/2$$
$$< -C\sum_{k \in S^t} \hat{p}_k^- + CD_t - C/2 < c_l/2$$

Thus, agent $l \notin S^t(\hat{p'}_i, \hat{p}_{-i})$. We now prove the main result of this section:

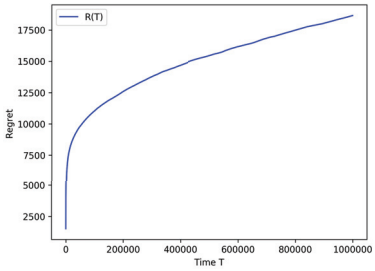
**Theorem 5** *Allocation rule returned by GLS-MAB is stochastic monotone.*

**Proof:** We want to prove that in expectation, with bid $\hat{c}_i^-$, agent $i$ receives more offers as compared to with bid $\hat{c}_i$. Once IIA property is in place (Lemma 10), the proof of stochastic monotonicity is similar to the proof in (Babaioff, Kleinberg, and Slivkins 2010). However, the proof in (Babaioff, Kleinberg, and Slivkins 2010) works only for the case when single agent needs to be selected at each round. Let us represent a table $n \times T$ where $(i,t)$ represents whether agent $i$ accepted the offer or not when he was given the offer at round $t$. Note that we might not be able to observe each and every entry of this table, as an agent may not be given the offer for all $T$ rounds. For the sake of the proof we fix the table and we will prove that the allocation rule by GLS-MAB is monotone with respect to any given table and hence has to be stochastic monotone. Let us now proceed by fixing an instance of the table and use induction on round $t$.
When $t = 1$, the offer is given to everyone irrespective of the bids and hence we have $i \in S^1(\hat{c}_i, \hat{c}_{-i})$ and $i \in S^1(\hat{c}_i^-, \hat{c}_{-i})$. Thus, from induction hypothesis we have: $\sum_{t'=1}^{t} \mathbb{1}(i \in S^{t'}(\hat{c}_i^-, \hat{c}_{-i})) \geq \sum_{t'=1}^{t} \mathbb{1}(i \in S^{t'}(\hat{c}_i, \hat{c}_{-i}))$. We now have to prove that the above condition is true for round $t+1$ as well. Note that we just have to worry about the case when the above condition holds with equality, otherwise for $t+1$ round, the condition holds trivially. Thus, let us assume: $\sum_{t'=1}^{t} \mathbb{1}(i \in S^{t'}(\hat{c}_i^-, \hat{c}_{-i})) = \sum_{t'=1}^{t} \mathbb{1}(i \in S^{t'}(\hat{c}_i, \hat{c}_{-i}))$. When the number of allocations to agent $i$ are equal in both the cases, the estimates of agent $i$ are fixed, since we have fixed the table. However, it may so happen that the estimates of other agents are changed since $i$ may get totally different rounds with bids $\hat{c}_i$ and $\hat{c}_i^-$. We will prove that if this is the case then estimates of other agents will also not change due to IIA property. Let us denote $a^t(\hat{c}_i^-) = t - \sum_{t'=1}^{t} \mathbb{1}(i \in S^{t'}(\hat{c}_i^-, \hat{c}_{-i}))$ and $a^t(\hat{c}_i^-) = t - \sum_{t'=1}^{t} \mathbb{1}(i \in S^{t'}(\hat{c}_i^-, \hat{c}_{-i}))$ as the number of instances where agent $i$ was not selected till round $t$. We will prove that for any rounds $t, s$ if $a^t(\hat{c}_i^-) = a^s(\hat{c}_i)$ then $\sum_{t'=1}^{t} \mathbb{1}(j \in S^{t'}(\hat{c}_i^-, \hat{c}_{-i})) = \sum_{t'=1}^{s} \mathbb{1}(j \in S^{t'}(\hat{c}_i^-, \hat{c}_{-i}))\ \forall j \neq i$.

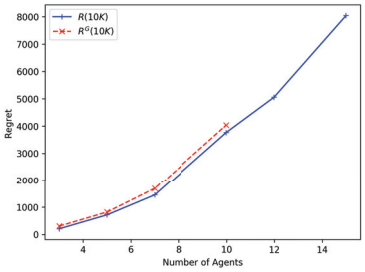
Let us prove this using induction, when $a^t(\hat{c}_i^-) = a^s(\hat{c}_i) = 0$ i.e. when $i$ is selected for all rounds. In this case the allocation of any agent depends only on the estimates of agent $i$ and his own bid. It does not depend on the bid of agent $i$. Since, the agent $i$ is selected for all the rounds, the estimates at any round for agent $i$ will be the same in both the case and hence the same subset of agents will be selected with bid $\hat{c}_i^-$ and $\hat{c}_i$. Thus, we have $t - \sum_{t'=1}^{t} \mathbb{1}(j \in S^{t'}(\hat{c}_i^-, \hat{c}_{-i})) = s - \sum_{t'=1}^{s} \mathbb{1}(j \in S^{t'}(\hat{c}_i^-, \hat{c}_{-i}))\ \forall j \neq i$. By induction hypothesis when $a^t(\hat{c}_i^-) = a^s(\hat{c}_i) = a$, then $t - \sum_{t'=1}^{t} \mathbb{1}(j \in S^{t'}(\hat{c}_i^-, \hat{c}_{-i})) = s - \sum_{t'=1}^{s} \mathbb{1}(j \in S^{t'}(\hat{c}_i^-, \hat{c}_{-i}))\ \forall j \neq i$. We need to prove for the case when $a^t(\hat{c}_i^-) = a^s(\hat{c}_i) = a+1$. Let $t'$ and $s'$ be the last rounds such that $a^{t'}(\hat{c}_i^-) = a^{s'}(\hat{c}_i) = a$. Thus, from $t'+1$ to $t-1$ agent $i$ is selected with bid $\hat{c}_i^-$ and similarly from $s'+1$ to $s-1$ agent $i$ is selected with bid $\hat{c}_i$. From induction hypothesis: $t' - \sum_{t=1}^{t'} \mathbb{1}(j \in S^t(\hat{c}_i^-, \hat{c}_{-i})) = s' - \sum_{t=1}^{s'} \mathbb{1}(j \in S^t(\hat{c}_i^-, \hat{c}_{-i}))\ \forall j \neq i$. Since, the estimates of all the agents




(a) Worst-case and average case ratio of local optima and global optima vs $n$




(b) Expected regret ($R_T$) vs $T$, $n = 15$




(c) Expected regret ($R_T$ and $R_T^G$) vs $n$

Figure 1: Experimental Results

are same till round $t'$ and $s'$, and from $t'+1$ to $t$ agent $i$ is selected, from IIA property, it will not influence allocation of other agents and same subset will be selected for all the rounds, the estimates till round $t-1$ and $s-1$ will be the same. Thus, the subset at round $t$ and $s$ will be the same and hence the proof follows. Now, at $t+1$, the estimates are same. Thus, due to the monotonicity Theorem 4, we have: $\sum_{t'=1}^{t+1} \mathbb{1}(i \in S^{t'}(\hat{c}_i^-, \hat{c}_{-i})) \geq \sum_{t'=1}^{t+1} \mathbb{1}(i \in S^{t'}(\hat{c}_i, \hat{c}_{-i}))$.

One of the major hurdle to obtain a truthful mechanism is to come up with a monotone allocation rule. Once we have the monotone allocation rule, one can easily obtain the payment using the black box technique provided in (Babaioff, Kleinberg, and Slivkins 2010). Though the mechanism is provided for single agent selection at each round, the technique can easily be extended to multiple agent selection and has been done in (Jain et al. 2018). With this we have the final result as follows:

**Theorem 6** *GLS-MAB produces incentive compatible mechanism in expectation where expectation is taken over the randomness of the acceptance rate.*

## Simulation Results

We now present some simulation results to demonstrate the efficacy of GLS and GLS-MAB and validate the proposed theoretical bounds. For the simulation purposes, we have fixed the cost of buying the electricity $C = 3$ and maximum CPR of the agents to be 1. For each round, we generated the demand shortage $\sim U[1, \frac{n}{4}]$, $n$ being the number of agents.

In Fig. 1(a), we compare the solution obtained by GLS and the optimal solution. We compute an optimal solution via the brute-force technique by considering all possible subsets. We compute the ratio of the loss incurred by GLS to the loss incurred by the optimal algorithm. We plot the average ratio and worst-case ratio of over 5000 samples by varying ARs and CPRs of the agents in each sample. As can be seen from the figure that the average ratio is very close to one, i.e., most of the times, the solution obtained was very close to the global optima. Even in the worst-case scenario, the ratio is tightly bounded and remains close to two.

We next study the growth of regret with $t$ in Fig. 1(b). We vary rounds from $t = 1 \rightarrow 10^6$ and do it across randomly generated 40 samples. The graph shown is the average of the regret of 40 samples. As can be seen from the figure, the regret is sub-linear, close to a logarithmic function of $t$. It is better than theoretically obtained bound of $\sqrt{T}$. We study the effect of the number of agents $n$ on regret in Fig. 1(c). We consider both the regret of GLS-MAB w.r.t. GLS ($R(T)$) as well as w.r.t. a global optima ($R^G(T)$). For $R(T)$, we consider up to $n = 15$ whereas $R^G(T)$ we consider only up to $n = 10$. We see that the regret grows quadratically with the number of agents as established in our theoretical analysis. We can further see from Fig. 1(c) that $R_T$ and $R_T^G$ are very close to each other. The simulation results establish that although the proposed algorithm provides us a local optimum, it provides us excellent performance in terms of the regret in practice.

## Conclusion and Future Work

When a distribution company wants to reduce peak energy cost, it is best to incentivize the consumers to reduce their consumption as opposed to buying high-cost electricity from the secondary market. It is called a demand response mechanism. However, designing such monetary offers in demand response mechanisms are challenging due to the high uncertainty in the smart grids arising from more and more renewable integration. To resolve these uncertainties, we proposed a GLS algorithm to select a subset of consumers to offer incentives. We designed GLS-MAB to learn these uncertainties and transformed it into a truthful combinatorial MAB mechanism to design monetary offers. We then analyzed our mechanism and showed that it is incentive compatible and achieves optimal regret in terms of social welfare.

Though we are solving the problem of minimizing a non-monotone supermodular function which in general can be a hard problem, we have not proved that our particular problem is NP-hard or not. In future, we would like to either prove that the problem is NP-hard or come up with a polynomial-time algorithm that gives global optima. This work can be extended to a setting where each consumer can reduce more than one unit, which can again be a private information to the consumer. In a more compelling future work, one may look for a more complicated distribution function (for example Gaussian) over ARs. One can also look at other types of mechanism like posted price mech-

anism where consumers come online, and the distributor company has to decide whether to ask the consumer to reduce the consumption or not.

## References


Akasiadis, C.; Panagidi, K.; Panagiotou, N.; Sernani, P.; Morton, A.; Vetsikas, I. A.; Mavrouli, L.; and Goutsias, K. 2015. Incentives for rescheduling residential electricity consumption to promote renewable energy usage. In *2015 SAI Intelligent Systems Conference (IntelliSys)*, 328–337.

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

Babaioff, M.; Kleinberg, R. D.; and Slivkins, A. 2010. Truthful mechanisms with implicit payment computation. In *Eleventh ACM Conference on Electronic Commerce EC'10*, 43–52. ACM.

Bulow, J., and Klemperer, P. 1996. Auctions versus negotiations. *American Economic Review* 86(1):180–194.

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

2010. EU Commission Task Force for Smart Grids. http://www.ieadsm.org/wp/files/Tasks/Task17-IntegrationofDemandSideManagement,EnergyEfficiency, DistributedGenerationandRenewableEnergySources/ Backgroundmaterial/Eg1documentv_24sep2010conf.pdf.

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

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.

Jain, S.; Balakrishnan, N.; Narahari, Y.; Hussain, S. A.; and Voo, N. Y. 2013. Constrained tâtonnement for fast and incentive compatible distributed demand management in smart grids. In *Proceedings of the fourth international conference on Future energy systems*, 125–136. ACM.

Jain, S.; Gujar, S.; Bhat, S.; Zoeter, O.; and Narahari, Y. 2018. A quality assuring, cost optimal multi-armed bandit mechanism for expertsourcing. *Artificial Intelligence* 254:44 – 63.

Jain, S.; Narayanaswamy, B.; and Narahari, Y. 2014. A multiarmed bandit incentive mechanism for crowdsourcing demand response in smart grids.

Li, Y.; Hu, Q.; and Li, N. 2018. Learning and selecting the right customers for reliability: A multi-armed bandit approach. In *2018 IEEE Conference on Decision and Control (CDC)*, 4869–4874. IEEE.

Ma, H.; Robu, V.; Li, N. L.; and Parkes, D. C. 2016. Incentivizing reliability in demand-side response. In *the proceedings of The 25th International Joint Conference on Artificial Intelligence (IJCAI'16)*, 352–358.

Ma, H.; Parkes, D. C.; and Robu, V. 2017. Generalizing demand response through reward bidding. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*, AAMAS '17, 60–68. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems.

Methenitis, G.; Kaisers, M.; and La Poutré, H. 2019. Forecast-based mechanisms for demand response. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*, 1600–1608. International Foundation for Autonomous Agents and Multiagent Systems.

Mittal, S., and Schulz, A. S. 2013. An fptas for optimizing a class of low-rank functions over a polytope. *Mathematical Programming* 141(1-2):103–120.

Myerson, R. B. 1981. Optimal auction design. *Mathematics of Operations Research* 6(1):58–73.

Park, S.; Jin, Y.; Song, H.; and Yoon, Y. 2015. Designing a critical peak pricing scheme for the profit maximization objective considering price responsiveness of customers. *Energy* 83:521–531.

Ramchurn, S.; Vytelingum, P.; Rogers, A.; and Jennings, N. 2011. Agent-based control for decentralised demand side management in the smart grid. In *AAMAS*, 5–12.

Robu, V.; Chalkiadakis, G.; Kota, R.; Rogers, A.; and Jennings, N. R. 2016. Rewarding cooperative virtual power plant formation using scoring rules. *Energy* 117:19 – 28.

Zhang, Q.; Wang, X.; and Fu, M. 2009. Optimal implementation strategies for critical peak pricing. In *2009 6th International Conference on the European Energy Market*, 1–6. IEEE.