

# A Contract-Based Approach for Ancillary Services in V2G Networks: Optimality and Learning

Yang Gao\*, Yan Chen\*, Chih-Yu Wang† and K. J. Ray Liu\*

\*Department of Electrical and Computer Engineering,  
University of Maryland, College Park, MD 20742, USA

† Graduate Institute of Communication Engineering,  
National Taiwan University, Taipei, Taiwan

E-mail: \*{yanggao, yan, kjrlui}@umd.edu; † tomkywang@gmail.com

**Abstract**—With the foreseeable large scale deployment of electric vehicles (EVs) and the development of vehicle-to-grid (V2G) technologies, it is possible to provide ancillary services to the power grid in a cost efficient way, i.e., through the bidirectional power flow of EVs. A key issue in such kind of schemes is how to stimulate a large number of EVs to act coordinately to achieve the service request. This is challenging since EVs are self-interested and generally have different preferences toward charging and discharging based on their own constraints. In this paper, we propose a contract-based mechanism to tackle this challenge. Through the design of an optimal contract, the aggregator can provide incentives for EVs to participate in ancillary services to power grid, match the aggregated energy rate with the service request and maximize its own profits. We prove that under mild conditions, the optimal contract-based mechanism takes a very simple form, i.e., the aggregator only needs to publish an optimal unit price to EVs, which is determined based on the statistical distribution of EVs' preferences. We then consider a more practical scenario where the aggregator has no prior knowledge regarding the statistical distribution and study how should the aggregator learn the optimal unit price from its interactions with EVs. Simulation results are shown to verify the effectiveness of the proposed contract-based mechanism.

## I. INTRODUCTION

Due to the need of reducing oil dependence and improving energy efficiency, many countries including United States have set up plans to support electric vehicles (EVs) [1]. It is predicated in [1] that by 2020, 25 percents of newly purchased light-duty vehicles should be grid-enabled EVs. Such a widespread adoption of EVs, together with the development of vehicle-to-grid (V2G) technologies [2], will open new opportunities for the power grid: using EVs' batteries as distributed electricity storage, it is possible to provide ancillary services to power grid, such as spinning reserve and regulation, in a cost efficient way. In current power grid, these services are accomplished primarily by turning large generators on and off or ramping them up and down, which are very costly, e.g., accounting for 5 – 10% of electric cost in the US [3]. Alternatively, in V2G networks, these services can be done by charging (or discharging) EVs' batteries when the generation is greater (or less) than the load in power grid. As the capacity of an individual EV is limited, the aggregator is introduced to coordinate a large group of EVs in providing ancillary services to the power grid [4].

The viability of EV-assisted ancillary services largely depends on the willingness of EVs to participate and to act coordinately. In practice, EVs are selfish in that they are only interested in maximizing their own utilities regardless of whether the ancillary services can be accomplished or not. Moreover, with the development of smart grid technologies [5], it is possible for EVs to make intelligent decisions representing their own interests. Therefore, it is no longer valid to assume that EVs will follow some controlling policies made by the aggregator unconditionally. Instead, proper incentive schemes must be designed to stimulate a large group of selfish and intelligent EVs to act coordinately to accomplish the ancillary service to the power grid. However, the design of effective incentive schemes is challenging due to the possible information asymmetry between the aggregator and EVs. In practice, since EVs generally face different practical constraints, such as arrival time, departure time, initial battery level and target battery level, they will have different preferences toward charging/discharging at different time. Nevertheless, such preferences are unknown to the aggregator, which makes the task of designing effective incentive schemes even more challenging.

To tackle this challenge, we propose to use contract theory, which studies, in the presence of asymmetric information, how the principal (the aggregator) delegates an action (charging/discharging at a certain rate) to intelligent and selfish agents (EVs) through a take-it-or-leave-it offer of a contract [6]. Through the optimal contract design, the aggregator not only can stimulate self-interested EVs to act coordinately to provide ancillary services to the power grid, but also maximize its own profits. We show theoretically that, under mild conditions, the optimal contract takes a very simple form where the aggregator only needs to publish the optimal unit price to EVs. Such an optimal contract-based mechanism has a distributed manner and can be implemented very efficiently with nearly no additional communication and control cost, compared with traditional pricing schemes.

To determine the optimal unit price explicitly, the aggregator needs to know the statistical distribution of EVs' preferences. We then extend our results to a more practical scenario where the aggregator has no prior knowledge regarding the statistical distributions and study how should the aggregator learn the

optimal unit price from its interactions with EVs. In such a case, the aggregator naturally faces an exploration-exploitation tradeoff between choosing the unit price with the best predicted performance to maximize immediate utility and trying different unit prices to obtain improved estimates. Inspired by the well-known UCB1 algorithm [7] in the machine learning literature, we propose an algorithm for the aggregator to learn the optimal unit price. To show the effectiveness of our algorithm, we compare it with the benchmark case where the aggregator has the prior knowledge and thus can choose the optimal unit price at every time slot. We prove theoretically that the total performance loss of our algorithm compared with the benchmark case over  $t$  time slots can be upper bounded uniformly by  $O(\log t)$ . In other words, the averaged performance loss will converge to 0 faster than  $O(\frac{\log t}{t})$  uniformly.

The rest of the paper is organized as follows. Section II presents the related work. In Section III, we introduce the system model and problem formulation. The optimal contract design is discussed in Section IV. Section V considers the scenario where the statistical distributions of EVs' preferences are unknown. Finally, we show simulation results in Section VI and draw conclusions in Section VII.

## II. RELATED WORK

Recently, a growing body of literature has investigated different charging control schemes for the aggregator. In [8], Xu and Wong proposed a coordinated charging control method that uses approximate dynamic programming to minimize the charging cost and reduce the power losses. Wu *et al.* proposed algorithms that help the aggregator to determine the purchase of energy in the day-ahead market and to distribute the purchased energy to EVs [9]. Among these works, many of them have studied the use of EVs for ancillary services. Frequency regulation has been considered in [10], where an optimal centralized control strategy was proposed. In [11], the authors studied the unidirectional V2G and developed an optimal algorithm for unidirectional regulation. In [12], Sortomme *et al.* proposed an optimal energy and ancillary service scheduling strategy that maximizes profits to the aggregator. Different from the works in [8] - [12] that assume implicitly that EVs are obedient to aggregator's policies, we consider selfish and intelligent EVs and focus on designing an incentive scheme to stimulate a large group of EVs to provide ancillary services to the power grid.

The incentive issue has been previously considered in work that focuses on demand response in general power grid networks [13]-[15]. In [16], the authors proposed a game theoretic demand response scheme that achieves the distributed load prediction through the participation of customers. Our work differs from the works in [13]-[16] primarily in that we focus on the scenario of providing ancillary services to the power grid through bidirectional power flows of EVs. In [17], Wu *et al.* studied the problem of coordinating a large group of selfish and intelligent EVs to provide frequency regulation to the power grid and proposed a pricing scheme to accomplish the service request at the equilibrium. However, one major drawback is that they assume a homogeneous setting without taking into

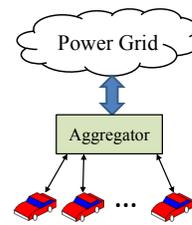


Fig. 1. The vehicle-to-grid system model considered in this paper.

account different preferences of EVs. In this work, we consider a heterogeneous setting and use contract theory to study the interaction between the aggregator and EVs under the presence of information asymmetry.

Another body of related literature is the work that studies the multi-armed bandit problem from the machine learning community. The multi-armed bandit problem is a simple yet fundamental example of the classic exploration versus exploitation tradeoff, which can be described as finding a balance between learning the profitable actions (i.e., exploration) and taking the empirically best action (i.e., exploitation). In their seminal work [18], Lai and Robbins studied this tradeoff and proved that under certain conditions the regret, i.e. the performance loss due to the fact that the policy does not always choose the best action, grows at least logarithmic in time. A learning policy was also proposed to achieve logarithmic regret asymptotically. Later in [7], Auer *et al.* proposed several index based policies which not only have a simpler form but also can achieve logarithmic regret uniformly over time. In this work, when the aggregator needs to learn the optimal unit price from the interactions with EVs, it naturally faces the exploration versus exploitation tradeoff. However, our case is different from the classic multi-armed bandit settings in that the optimal unit price depends on the service request and therefore is changing over time rather than static as in the classic multi-armed bandit settings.

## III. SYSTEM MODEL AND PROBLEM FORMULATION

### A. V2G System Model

Consider a V2G system as shown in Fig. 1. There are a group of  $N$  EVs interested in providing ancillary services to the power grid by charging/discharging their batteries. One aggregator is responsible for coordinating the participating EVs. However, the aggregator has no direct control over the charging/discharging behaviors of EVs, who are assumed to be self-interested and will act selfishly to maximize their own utilities.

We divide the daily operation of the power grid into multiple time slots, each of which corresponds to one service period. At each time slot, the grid sends a service request to the aggregator indicating the aggregated energy rate needed from the aggregator in order to accomplish the ancillary service. Denote by  $\Delta$  the service request sent to the aggregator. We study the nontrivial case that  $\Delta \neq 0$  in this paper. If  $\Delta > 0$ , the aggregator needs to consume power. If  $\Delta < 0$ , the aggregator

needs to inject power into the power grid. Such a service request is accomplished by the aggregator through coordinating the  $N$  associated EVs to charge/discharge their batteries. Moreover, similar to [17], we assume that the aggregator is equipped with a set of backup batteries to assure reaching the service request.

Since EVs generally have different constraints, such as arrival time, departure time, initial battery level and target battery level, they will have different preferences toward charging/discharging at each time slot, which we model by a willingness to pay (WTP) parameter  $\theta$  [19]. Let  $r$  denote the charging/discharging rate of an EV and  $p$  denote the price paid to the aggregator. Then the utility function of the EV with WTP parameter  $\theta$  can be written as

$$u_{\theta}(r, p) = \begin{cases} \theta r - p, & \text{if } r \geq 0, \\ (C + \theta)r - p, & \text{otherwise,} \end{cases} \quad (1)$$

where  $C > 0$  is the unit cost associated with discharging.

Note that both  $r$  and  $p$  can take either positive or negative values. In particular,  $r > 0$  means the EV charges its battery at current time slot while  $r < 0$  means discharging. Based on the utility function, the WTP parameter can be interpreted as the unit gain that an EV can receive by charging/discharging its battery. In this paper, we model the WTP parameter  $\theta \in \Theta = [\underline{\theta}, \bar{\theta}]$  as a random variable, which is drawn independently and identically for different EVs and for different time slots. We represent by  $F(\theta)$  the cumulative distribution function (CDF) and by  $f(\theta)$  the probability density function (PDF). Moreover, we assume that  $-C < \underline{\theta} < 0$  and  $\bar{\theta} > 0$ . The sign of  $\theta$  indicates whether the EV tends to charge or discharge: when  $\theta > 0$ , the EV prefers to charge; when  $\theta < 0$  the EV prefers to discharge and when  $\theta = 0$  the EV wants to remain idle. Moreover, the larger  $|\theta|$  is, the more an EV wants to charge/discharge its battery, respectively. As an independent decision-maker, each EV will act to maximize its own utility function in (1) without considering whether the aggregated load matches the service request or not. Therefore, an inherent conflict exists in terms of objectives between the aggregator and EVs.

We further assume that the WTP parameter is the private information of each EV, which implies that the aggregator has no access to the specific value of each EV's WTP parameter. We thus refer to the WTP parameter as each EV's type following the contract theory terminology. We first study the case where the aggregator is aware of the distribution of EV's type, i.e.,  $f(\theta)$  and  $F(\theta)$ . Then, in Section V, we extend our results to the scenario that the aggregator has no prior knowledge regarding  $f(\theta)$  and  $F(\theta)$ . In both cases, there exists an information asymmetry between the aggregator and EVs, which makes the coordination at the aggregator even harder.

### B. Contract-Theoretic Formulation

To resolve the conflicting objectives between the aggregator and EVs in the presence of asymmetric information, we propose to use a contract-theoretic approach. Through an optimal design of contract, the aggregator can not only stimulate self-interested EVs to act coordinately to accomplish the service request but

also maximize its own profits. In contract theory, a contract is a collection of contract items. Particularly, in our case, each contract item corresponds to a pair  $(r, p)$ , which specifies the EV's charging/discharging rate and the resulted payment to the aggregator. At each time slot, the aggregator will publish the contract to all participating EVs. Then each EV will choose one contract item that maximizes its utility defined in (1). According to the revelation principle [20], it is sufficient to consider the class of contracts that ensure each EV to truthfully choose the contract item designed for its type. Therefore, we can design our contract as a pair of functions as  $\phi = \{(r(\theta), p(\theta)), \theta \in \Theta\}$ . Throughout this paper, we restrict our attentions to functions that are differentiable, including generalized functions. For example, a unit step function is admissible since it yields a delta function as its derivative. To be a feasible contract,  $\phi$  needs to satisfy the incentive compatibility (IC) constraint and the individual rationality (IR) constraint, which we define as follows.

**Definition 1** (Incentive Compatibility). *A contract  $\phi = \{(r(\theta), p(\theta)), \theta \in \Theta\}$  satisfies the incentive compatibility constraint if it is the best response of each  $\theta$ -type EV to choose the contract item for type  $\theta$ , i.e.,*

$$u_{\theta}(r(\theta), p(\theta)) \geq u_{\theta}(r(\tilde{\theta}), p(\tilde{\theta})), \quad \forall \theta, \tilde{\theta} \in \Theta. \quad (2)$$

**Definition 2** (Individual Rationality). *A contract  $\phi = \{(r(\theta), p(\theta)), \theta \in \Theta\}$  satisfies the individual rationality constraint if each  $\theta$ -type EV receives a non-negative utility by accepting the contract item for type  $\theta$ , i.e.,*

$$u_{\theta}(r(\theta), p(\theta)) \geq 0, \quad \forall \theta \in \Theta. \quad (3)$$

A contract that satisfies the IR constraint will provide non-negative utilities to EVs of any types, and therefore ensures the participation of self-interested EVs.

In addition to the IC and IR constraints, the aggregator will design the contract such that the expected aggregated energy rate of all EVs meets the service request, i.e.,

$$N \int_{\underline{\theta}}^{\bar{\theta}} r(\theta) f(\theta) d\theta = \Delta. \quad (4)$$

Moreover, we place the following two constraints on the design of contracts

$$r(\theta)\Delta \geq 0, \quad \forall \theta \in \Theta \quad (5)$$

and

$$r(0) = 0. \quad (6)$$

The first constraint indicates that charging and discharging should not be performed simultaneously. Otherwise EVs cancel out each other's efforts in terms of providing ancillary services to the power grid. The second constraint indicates that we should not involve EVs that choose to remain idle into the ancillary service.

The rate function must also satisfy

$$r_{min} \leq r(\theta) \leq r_{max}, \quad \forall \theta \in \Theta, \quad (7)$$

where  $r_{max} > 0$  and  $r_{min} < 0$  are the maximum charging and discharging rates of EVs, respectively.

Denote by  $\Phi$  the set of contracts that satisfy all constraints in (2)-(7). Assume  $N$  is sufficiently large that  $\Phi$  is not empty. Among all contracts in  $\Phi$ , the aggregator will choose the optimal one, which maximizes its profit as

$$\phi^* = \arg \max_{\phi \in \Phi} \int_{\underline{\theta}}^{\bar{\theta}} p(\theta) f(\theta) d\theta. \quad (8)$$

The proposed contract-based mechanism in one time slot can be summarized in the following four steps.

- 1) The aggregator receives the service request from the power grid and calculates the optimal contract  $\phi^*$
- 2) The aggregator broadcasts  $\phi^*$  to all EVs
- 3) After receiving  $\phi^*$ , each EV selects one contract item that maximizes its utility and informs the aggregator its decision
- 4) The aggregator coordinates the ancillary service and records EVs' payments given the selected contract items

#### IV. OPTIMAL CONTRACT DESIGN

To find the optimal contract, we need to solve the optimization problem defined in (8), which is challenging because it optimizes over a class of functions specified by some complicated constraints. In this section, we first simplify the optimization problem to a certain extent by finding equivalent conditions to the IC and IR constraints. Then, by rewriting the above optimization problem into an equivalent but simpler form, we show that, under some mild conditions, the optimal contract takes a very simple form. In the rest of this paper, we assume without loss of generality that  $\Delta > 0$  in our analysis. The case with  $\Delta < 0$  can be analyzed similarly.

We present in the following two propositions that the IC and IR constraints can be simplified under our problem settings.

**Proposition 1.** *Suppose  $\Delta > 0$  and a contract  $\phi = \{(r(\theta), p(\theta)), \theta \in \Theta\}$  satisfies the efficiency constraint defined in (5). Then  $\phi$  satisfies the IC constraint if and only if,  $\forall \theta \in \Theta$ ,*

$$\dot{r}(\theta) \geq 0 \quad (9)$$

and

$$\theta \dot{r}(\theta) - \dot{p}(\theta) = 0. \quad (10)$$

*Proof:* From  $\Delta > 0$  and (5), we have  $r(\theta) \geq 0, \forall \theta \in \Theta$ , which implies  $u_\theta(r(\theta), p(\theta)) = \theta r(\theta) - p(\theta)$ .

To prove Proposition 1, we first show that the two conditions in (9) and (10) are necessary conditions for the IC constraint. From Definition 1, we have  $\forall \theta, \tilde{\theta} \in \Theta$ ,

$$\theta r(\theta) - p(\theta) \geq \theta r(\tilde{\theta}) - p(\tilde{\theta}),$$

and

$$\tilde{\theta} r(\tilde{\theta}) - p(\tilde{\theta}) \geq \tilde{\theta} r(\theta) - p(\theta).$$

Adding the above two inequalities, we have

$$(\theta - \tilde{\theta})(r(\theta) - r(\tilde{\theta})) \geq 0, \forall \theta, \tilde{\theta} \in \Theta.$$

Therefore, we can conclude that  $\dot{r}(\theta) \geq 0, \forall \theta \in \Theta$ .

Moreover, let

$$g_\theta(\tilde{\theta}) \triangleq \theta r(\tilde{\theta}) - p(\tilde{\theta}).$$

Then the IC constraint implies that

$$\theta \in \arg \max_{\tilde{\theta} \in \Theta} g_\theta(\tilde{\theta}), \quad \forall \theta \in \Theta.$$

Since  $g_\theta(\tilde{\theta})$  is differentiable, from the first-order optimality condition [21], we have

$$\left. \frac{\partial g_\theta(\tilde{\theta})}{\partial \tilde{\theta}} \right|_{\tilde{\theta}=\theta} = \theta \frac{d}{d\theta} r(\theta) - \frac{d}{d\theta} p(\theta) = 0, \quad \forall \theta \in (\underline{\theta}, \bar{\theta}).$$

Moreover, since boundary values of  $\dot{r}(\theta)$  and  $\dot{p}(\theta)$  will not affect our results, we can extend the above equality to the boundary points and establish (10).

Next, we prove conditions in (9) and (10) are also sufficient conditions for the IC constraint. We have  $\forall \theta, \tilde{\theta} \in \Theta$ ,

$$\begin{aligned} p(\theta) - p(\tilde{\theta}) &= \int_{\tilde{\theta}}^{\theta} \dot{p}(\tau) d\tau = \int_{\tilde{\theta}}^{\theta} \tau \dot{r}(\tau) d\tau \\ &= \theta r(\theta) - \tilde{\theta} r(\tilde{\theta}) - \int_{\tilde{\theta}}^{\theta} r(\tau) d\tau, \end{aligned}$$

where the second equality follows from (10) and the last equality is obtained through integration by parts.

After some manipulations, we have

$$\begin{aligned} \theta r(\theta) - p(\theta) &= \theta r(\tilde{\theta}) - p(\tilde{\theta}) + \int_{\tilde{\theta}}^{\theta} [r(\tau) - r(\tilde{\theta})] d\tau \\ &\geq \theta r(\tilde{\theta}) - p(\tilde{\theta}), \quad \forall \theta, \tilde{\theta} \in \Theta, \end{aligned}$$

where the inequality follows from (9). ■

**Proposition 2.** *Suppose  $\Delta > 0$  and a contract  $\phi = \{(r(\theta), p(\theta)), \theta \in \Theta\}$  satisfies the IC constraint and the efficiency constraint defined in (5). Then,  $\phi$  satisfies the IR constraint if and only if*

$$\underline{\theta} r(\underline{\theta}) - p(\underline{\theta}) \geq 0. \quad (11)$$

*Proof:* Let  $U(\theta) \triangleq \theta r(\theta) - p(\theta)$ . Since  $\phi$  satisfies the IC constraint, according to (10), we have

$$\dot{U}(\theta) = \frac{d}{d\theta} U(\theta) = r(\theta) + \theta \dot{r}(\theta) - \dot{p}(\theta) = r(\theta), \quad \forall \theta \in \Theta. \quad (12)$$

Since  $\Delta > 0$ , then (5) implies that  $r(\theta) \geq 0, \forall \theta \in \Theta$ . Therefore, we have

$$\underline{\theta} \in \arg \min_{\theta \in \Theta} U(\theta).$$

The IR constraint in (3) is thus equivalent to  $U(\underline{\theta}) \geq 0$ . ■

Based on the definition of  $U(\theta)$ , we can rewrite the objective function in (8) as

$$G(\phi) \triangleq \int_{\underline{\theta}}^{\bar{\theta}} \theta r(\theta) f(\theta) d\theta - \int_{\underline{\theta}}^{\bar{\theta}} U(\theta) f(\theta) d\theta.$$

The last term can be expressed in terms of  $r(\theta)$  as

$$\begin{aligned} \int_{\underline{\theta}}^{\bar{\theta}} U(\theta)f(\theta)d\theta &= \int_{\underline{\theta}}^{\bar{\theta}} f(\theta) \int_{\underline{\theta}}^{\theta} \dot{U}(\tau)d\tau d\theta + U(\underline{\theta}) \\ &= \int_{\underline{\theta}}^{\bar{\theta}} f(\theta) \int_{\underline{\theta}}^{\theta} r(\tau)d\tau d\theta + U(\underline{\theta}) \\ &= \int_{\underline{\theta}}^{\bar{\theta}} r(\theta)d\theta - \int_{\underline{\theta}}^{\bar{\theta}} r(\theta)F(\theta)d\theta + U(\underline{\theta}), \end{aligned}$$

where the second equality follows from (12) and the last equality is obtained through integration by parts.

Therefore, we have

$$G(\phi) = \int_{\underline{\theta}}^{\bar{\theta}} r(\theta)f(\theta) \left[ \theta - \frac{1-F(\theta)}{f(\theta)} \right] d\theta - U(\underline{\theta}).$$

We can simplify  $G(\phi)$  by the following two observations. First, since  $r(\theta) \geq 0, \forall \theta \in \Theta$ , we can conclude from (6) and (9) that  $r(\theta) = 0$  for  $\theta \in [\underline{\theta}, 0]$ . Second, we can maximize  $G(\phi)$  while satisfying the IR constraint by setting  $U(\underline{\theta}) = 0$ , i.e.  $p(\underline{\theta}) = 0$ .

Therefore, the optimal contract design problem in (8) can be simplified to a constrained optimization problem with respect to the rate function  $r(\theta)$  over  $[0, \bar{\theta}]$  as

$$\begin{aligned} \max_{r(\theta)} \quad & \int_0^{\bar{\theta}} r(\theta)f(\theta) \left[ \theta - \frac{1-F(\theta)}{f(\theta)} \right] d\theta \quad (13) \\ \text{subject to} \quad & \int_0^{\bar{\theta}} r(\theta)f(\theta)d\theta = \frac{\Delta}{N}, \\ & r(0) = 0, \\ & \dot{r}(\theta) \geq 0 \text{ and } r(\theta) \leq r_{max}, \quad \forall \theta \in [0, \bar{\theta}]. \end{aligned}$$

Once we have found the optimal rate function  $r^*(\theta)$ , we can determine the optimal pricing function  $p^*(\theta)$  as

$$p^*(\theta) = \int_{\underline{\theta}}^{\theta} \tau \dot{r}^*(\tau) d\tau. \quad (14)$$

To characterize the optimal contract analytically, we introduce below the concept of regular distribution.

**Definition 3** (Regular Distribution [22]). *We say that a distribution is regular if  $\left[ \theta - \frac{1-F(\theta)}{f(\theta)} \right]$  is non-decreasing.*

Regular distribution is an assumption widely adopted in mechanism design literature [20] [22], which comprises a large class of practical distributions, such as uniform, exponential and normal. We show in the following theorem that the optimal contract takes a very simple form under this mild condition.

**Theorem 1.** *For regular distributions, the optimal contract can be expressed as*

$$\begin{cases} r^*(\theta) = r_{max} \mathbf{1}(\theta \geq \alpha), \\ p^*(\theta) = \alpha r_{max} \mathbf{1}(\theta \geq \alpha), \end{cases} \quad (15)$$

where  $\alpha$  is the solution to

---

### Algorithm 1 : Implementation of The Optimal Contract

---

- 1: The aggregator receives the service request from the power grid and calculates the optimal unit price  $\alpha$ .
  - 2: The aggregator publishes  $\alpha$  to all EVs.
  - 3: Each EV decides whether to participate or not as well as the corresponding charging/discharging rate based on its own utility.
  - 4: The aggregator records the payments of participating EVs.
- 

$$r_{max} \int_{\alpha}^{\bar{\theta}} f(\theta)d\theta = \frac{\Delta}{N}. \quad (16)$$

*Proof:* It can be easily verified that  $r^*(\theta)$  and  $p^*(\theta)$  in (15) satisfy (14). Therefore, to prove Theorem 1, it suffices to show that  $r^*(\theta)$  in (15) is the solution to the optimization problem in (13).

We can check that  $r^*(\theta)$  satisfies the constraints in (13) and therefore is a valid candidate. To show its optimality, denote by  $\hat{r}(\theta)$  an arbitrary rate function that satisfies the constraints in (13). Let

$$\delta_r(\theta) = r^*(\theta) - \hat{r}(\theta). \quad (17)$$

Then we have  $\delta_r(\theta) \leq 0$  for  $\theta \in [0, \alpha]$ ,  $\delta_r(\theta) \geq 0$  for  $\theta \in [\alpha, \bar{\theta}]$  and

$$\int_0^{\alpha} \delta_r(\theta)f(\theta)d\theta + \int_{\alpha}^{\bar{\theta}} \delta_r(\theta)f(\theta)d\theta = 0. \quad (18)$$

Moreover, since the distribution is regular, we have,  $\forall \theta_1 \in [0, \alpha]$  and  $\forall \theta_2 \in [\alpha, \bar{\theta}]$ ,

$$\theta_1 - \frac{1-F(\theta_1)}{f(\theta_1)} \leq \alpha - \frac{1-F(\alpha)}{f(\alpha)} \leq \theta_2 - \frac{1-F(\theta_2)}{f(\theta_2)}. \quad (19)$$

Therefore, we have

$$\int_0^{\alpha} \delta_r(\theta)f(\theta) \left[ \theta - \frac{1-F(\theta)}{f(\theta)} \right] d\theta \geq \left[ \alpha - \frac{1-F(\alpha)}{f(\alpha)} \right] \int_0^{\alpha} \delta_r(\theta)f(\theta)d\theta,$$

and

$$\int_{\alpha}^{\bar{\theta}} \delta_r(\theta)f(\theta) \left[ \theta - \frac{1-F(\theta)}{f(\theta)} \right] d\theta \geq \left[ \alpha - \frac{1-F(\alpha)}{f(\alpha)} \right] \int_{\alpha}^{\bar{\theta}} \delta_r(\theta)f(\theta)d\theta.$$

Adding the above two inequalities, we derive

$$\int_0^{\bar{\theta}} \delta_r(\theta)f(\theta) \left[ \theta - \frac{1-F(\theta)}{f(\theta)} \right] d\theta \geq 0, \quad (20)$$

which implies that  $r^*(\theta)$  is the solution to the optimization problem in (13) and thus concludes the proof. ■

From Theorem 1, we can see that, under the assumption of regular distributions, it is optimal to let EVs with preferences higher than a certain threshold to charge with the maximum rate while keeping others idle. The threshold can also be interpreted as the optimal unit price which is determined by the service request as well as the distribution of EV's type. Such an optimal contract-based mechanism can be implemented very efficiently with nearly no additional communication and control cost, as demonstrated in the following corollary.

**Corollary 1.** *The optimal contract in Theorem 1 can be implemented through Algorithm 1.*

*Proof:* Algorithm 1 differs from a direct implementation of the contract-based mechanism in that it only requires the aggregator to publish the optimal unit price rather than the whole contract to EVs and let EVs decide their charging rates. Due to the rationality assumption of EVs, they will choose their charging rates so that their utilities can be maximized. Therefore, it suffices to prove that

$$r^*(\theta) = \max_{0 \leq r \leq r_{max}} u_\theta(r, \alpha r).$$

It can be easily verified that  $\max_{0 \leq r \leq r_{max}} u_\theta(r, \alpha r) = r_{max} \mathbf{1}(\theta \geq \alpha)$ , which is exactly the expression of  $r^*(\theta)$  in (15) and thus concludes the proof. ■

## V. LEARNING THE OPTIMAL UNIT PRICE WITHOUT PRIORS

In Section III, we have shown that the optimal contract-based mechanism for regular distributions takes a very simple form where the aggregator only needs to design and publish the optimal unit price. Nevertheless, such a simple scheme has one drawback in that it requires the distributional knowledge of EV's WTP parameter in order to calculate the optimal unit price. In practice, although it is reasonable to model EV's preference towards charging/discharging as a WTP parameter, sometimes it is hard for the aggregator to know the distribution of such a parameter. We tackle this challenge in this section. In particular, we will stick with the simple structure of the optimal contract and study how to learn the optimal unit price without the prior knowledge of  $f(\theta)$  and  $F(\theta)$ .

Consider a more practical setting where the unit price can only have discrete values. Let  $\Upsilon = \{\alpha_i | \alpha_i = \frac{i\bar{\theta}}{K}, i = 0, 1, \dots, K\}$  be the set of unit prices that the aggregator can choose from. Although the aggregator will suffer some loss by restricting the unit price to a set of discrete values, such a performance loss decreases as  $K$  increases. Moreover, since the achieved total energy rate at each time slot is just integer multiples of the maximum charging rate  $r_{max}$ , we assume that the service request takes value from the set  $\Omega = \{\Delta_j | \Delta_j = jr_{max}, j = 1, 2, \dots, M\}$  and the residue is handled by the backup batteries.

At time slot  $\tau$ , choosing unit price  $\alpha_i$  will lead to a total energy rate as

$$X_{i,\tau} = \sum_{n=1}^N \mathbf{1}(\theta_{n,\tau} \geq \alpha_i) r_{max}, \quad (21)$$

where  $\theta_{n,\tau}$  is the WTP parameter of EV  $n$  at time slot  $\tau$ . We would like to point out that independence holds for  $X_{i,\tau}$  in different time slots but does not hold across different unit prices, i.e.,  $X_{i,s}$  and  $X_{i,\tau}$  are independent while  $X_{i,\tau}$  and  $X_{j,\tau}$  are not for each  $0 \leq i < j \leq K$  and for each  $1 \leq s \leq \tau$ .

Assuming the service request at time slot  $\tau$  is  $\Delta_{j_\tau}$ , we can define a new random variable that represents the normalized

square of the difference between the total energy rate and the service request as

$$Y_{i,j_\tau} = \left( \frac{X_{i,\tau} - \Delta_{j_\tau}}{Mr_{max}} \right)^2. \quad (22)$$

The mean of  $Y_{i,j_\tau}$  is referred to as the normalized mean square residue and can be calculated as

$$\mu_{i,j_\tau} = \frac{N^2}{M^2} \left[ \frac{\beta_i(1 - \beta_i)}{N} + \left( \beta_i - \frac{j_\tau}{N} \right)^2 \right], \quad (23)$$

where

$$\beta_i = \int_{\alpha_i}^{\bar{\theta}} f(\theta) d\theta.$$

If the aggregator has the prior knowledge of the distribution, she would choose the optimal unit price  $\alpha_{i^*}$  at every time slot. Here, we adopt a slightly different yet more practical sense of optimality such that the normalized mean square residue is minimized, i.e.,

$$i_\tau^* \in \min_{0 \leq i \leq K} (\mu_{i,j_\tau}). \quad (24)$$

We denote  $\mu_{i_\tau^*,j_\tau}$  by  $\mu_{j_\tau}^*$  for notation simplicity.

Without the knowledge of  $f(\theta)$ , the aggregator needs to learn the optimal unit price from the interactions with EVs. During the learning procedure, the aggregator faces an exploration-exploitation tradeoff between choosing the unit price with the best predicted performance to maximize immediate utility and trying different unit prices to obtain improved estimates. Finding a learning algorithm that solves the exploration-exploitation tradeoff is traditionally formulated as a multi-armed bandit problem. However, results from multi-armed bandit literature cannot be directly applied here since they assume the optimal choice remains unchanged, whereas in our case, the optimal unit price depends on the service request and is changing over time.

Define by  $\sigma = \{\sigma_\tau\}$  the learning policy, where  $\sigma_\tau$  is a map from the observation history up to time slot  $\tau - 1$  to the index of unit price to be selected at time slot  $\tau$ . To evaluate the performance of  $\sigma$ , we adopt regret as our performance criterion [7] [18], which is the total performance loss with respect to the bench mark case of choosing the optimal unit price at every time slot. A formal definition of regret is given as follows.

**Definition 4 (Regret).** *The regret of policy  $\sigma$  after  $t$  time slots is defined by*

$$R_\sigma(t) = \mathbb{E} \left[ \sum_{\tau=1}^t (\mu_{\sigma_\tau, j_\tau} - \mu_{j_\tau}^*) \right], \quad (25)$$

where the expectation is taken over the possible randomness of the policy.

Our objective is to find a policy that yields low regret. We show the proposed policy in Algorithm 2, which modifies the UCB1 algorithm in [7] to tackle the case with time-variant optimal choices.

In Algorithm 2, we maintain two quantities for each unit price,  $\bar{y}_i$  and  $n_i$ , which represent the estimate of the normalized

**Algorithm 2 : Learning The Optimal Unit Price**


---

```

1: // Initialization
2: for  $t = 1$  to  $K + 1$  do
3:    $\sigma_t = t - 1$ 
4:   Observe and record  $x_{\sigma_t, t}$ 
5:    $n_{\sigma_t} \leftarrow 1$ 
6: end for
7: // Main Loop
8: while  $t \leq T$  do
9:   for  $i = 0$  to  $K$  do
10:     $\bar{y}_i \leftarrow \frac{1}{n_i} \sum_{\tau=1}^{t-1} \left( \frac{x_{i,\tau} - \Delta_{j_t}}{Mr_{max}} \right)^2 \mathbf{1}(\sigma_\tau = i)$ 
11:   end for
12:    $\sigma_t \leftarrow \arg \min_{0 \leq i \leq K} \bar{y}_i - \sqrt{\frac{2 \ln t}{n_i}}$ 
13:   Observe and record  $x_{\sigma_t, t}$ 
14:    $n_{\sigma_t} \leftarrow n_{\sigma_t} + 1$ 
15:    $t \leftarrow t + 1$ 
16: end while

```

---

mean square residue of unit price  $\alpha_i$  and the number of times  $\alpha_i$  has been chosen, respectively. The  $x_{i,\tau}$  is the realization of  $X_{i,\tau}$ , which can be observed if  $\alpha_i$  is chosen at time slot  $\tau$ . We record all observed  $x_{i,\tau}$  in the algorithm and use them to calculate  $\bar{y}_i$  at each time slot based on the service request  $\Delta_{j_t}$ . After initialization, the unit price is chosen simply according to an index policy that  $\sigma_\tau \in \arg \min_{0 \leq i \leq K} \bar{y}_i - \sqrt{\frac{2 \ln \tau}{n_i}}$ . Though simple, such an index policy well captures the exploration-exploitation tradeoff faced by the aggregator.

To show the effectiveness of the proposed policy, we prove in the following that its regret  $R_\sigma(t)$  is upper bounded uniformly by  $O(\log t)$ .

**Lemma 1.** Denote by  $T_i(t)$  the number of times that the unit price  $\alpha_i$  is chosen but does not have the optimal mean square residue after  $t$  rounds of the proposed policy, i.e.,

$$T_i(t) = \sum_{\tau=1}^t \mathbf{1}(\sigma_\tau = i, \mu_{i,j_\tau} > \mu_{j_\tau}^*). \quad (26)$$

Then, we can upper bound the expectation of  $T_i(t)$  by

$$\mathbb{E}[T_i(t)] \leq \frac{8 \ln t}{d_{min}^2} + 1 + \frac{\pi^2}{3}, \quad (27)$$

where

$$d_{min} = \min_{0 \leq i \leq K, 1 \leq j \leq M} (\mu_{i,j} - \mu_j^*), \text{ subject to } \mu_{i,j} \neq \mu_j^*.$$

*Proof:* See Appendix A for the proof. ■

**Theorem 2.** The regret  $R_\sigma(t)$  of the proposed policy  $\sigma$  can be upper bounded by

$$R_\sigma(t) \leq d_{max}(K+1) \left( \frac{8 \ln t}{d_{min}^2} + 1 + \frac{\pi^2}{3} \right),$$

where

$$d_{max} = \max_{0 \leq i \leq K, 1 \leq j \leq M} (\mu_{i,j} - \mu_j^*).$$

*Proof:* Following the definition of  $R_\sigma(t)$ , we have

$$\begin{aligned} R_\sigma(t) &= \mathbb{E} \left[ \sum_{\tau=1}^t (\mu_{\sigma_\tau, j_\tau} - \mu_{j_\tau}^*) \right] \\ &= \mathbb{E} \left[ \sum_{\tau=1}^t \sum_{i=0}^K (\mu_{i,j_\tau} - \mu_{j_\tau}^*) \mathbf{1}(\sigma_\tau = i, \mu_{i,j_\tau} > \mu_{j_\tau}^*) \right] \\ &\leq \mathbb{E} \left[ \sum_{\tau=1}^t \sum_{i=0}^K d_{max} \mathbf{1}(\sigma_\tau = i, \mu_{i,j_\tau} > \mu_{j_\tau}^*) \right] \\ &= \sum_{i=0}^K \left\{ d_{max} \mathbb{E} \left[ \sum_{\tau=1}^t \mathbf{1}(\sigma_\tau = i, \mu_{i,j_\tau} > \mu_{j_\tau}^*) \right] \right\} \\ &= \sum_{i=0}^K \{ d_{max} \mathbb{E}[T_i(t)] \} \\ &\leq \sum_{i=0}^K \left[ d_{max} \left( \frac{8 \ln t}{d_{min}^2} + 1 + \frac{\pi^2}{3} \right) \right] \\ &= d_{max}(K+1) \left( \frac{8 \ln t}{d_{min}^2} + 1 + \frac{\pi^2}{3} \right), \end{aligned}$$

which concludes the proof. ■

## VI. SIMULATION RESULTS

In this section, we conducted numerical simulations to evaluate the performance of the optimal contract-based mechanism. That is to say, the aggregator is assumed to know the distribution of EV's WTP parameter and can determine the optimal unit price explicitly in every time slot. A V2G system with  $N = 1000$  EVs is considered. We assume EV's WTP parameters are independent and identically distributed within  $[-1, 1]$  uniformly. The unit cost  $\mathcal{C}$  is assumed to be 1.1. Moreover, we set  $r_{max} = 19.2kW$  and  $r_{min} = -19.2kW$  according to the Level 2 charging standard in North America [24].

We consider the scenario of regulation and assume  $\Delta$  is updated every 20 seconds by the power grid. The value of  $\Delta$  is assumed to follow a Gaussian distribution with zero mean and the variance of  $4 \times 10^6 (MW^2)$ , as shown in Fig. 2(a). We show in Fig. 2(b) the difference between  $\Delta$  and the aggregated energy rate of all EVs by using the proposed mechanism. We can see that with the proposed mechanism, the aggregator can achieve over 95% of the service request. The differences are not zeros due to the randomness of EV's WTP parameter. Similar to [17], we assume that the residue is further handled by a set of backup batteries to assure reaching the service request.

We then compare the proposed mechanism with the pricing scheme in [17] in terms of the total payment received by the aggregator. In [17], to achieve the service request, the aggregator randomly selects a certain number of EVs to charge/discharge their batteries at a fixed rate. The aggregator will pay each selected EV a base price  $\omega$ , which is the same for all selected EVs, and charge them penalty prices if the service request can not be reached. Therefore, to avoid penalties, the selected EVs will follow the aggregator's instructions if they can receive non-negative utilities at the equilibrium. Otherwise they will simply

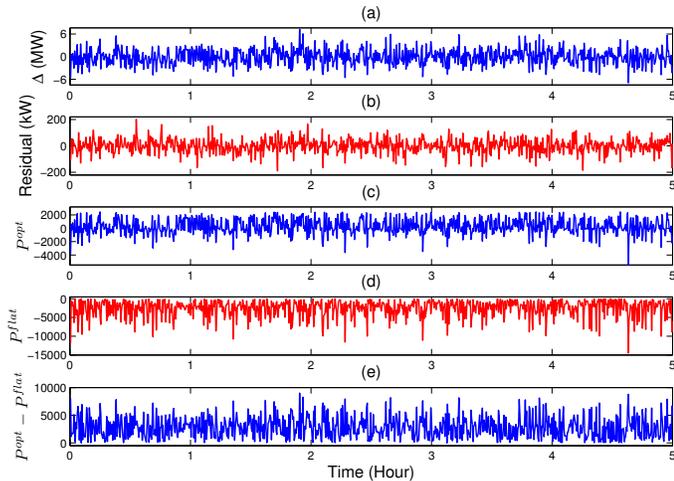


Fig. 2. Simulation results for frequency regulation: (a) The service request. (b) The difference between service request and the aggregated energy rate of all EVs. (c) The total payment received by the aggregator using the proposed mechanism. (d) The total payment received by the aggregator using the flat pricing scheme [17]. (e) Difference between the total payment to aggregator using the proposed mechanism and that using the flat pricing scheme.

choose not to participate. Since the aggregator does not know each EV's preference, the base price should be large enough so that every selected EV will have the incentive to participate. In the simulation, we set the fixed charging/discharging rate as  $r_{max}/r_{min}$ , respectively. To ensure participations, the base price is set as  $\omega = -\min_{\theta \in \Theta} \theta r_{max} = 19.2$  when  $\Delta > 0$  and  $\omega = -\min_{\theta \in \Theta} (\theta + C)r_{min} = 40.32$  when  $\Delta < 0$ . The total payment received by the aggregator using the optimal contract-based mechanism,  $P^{opt}$ , is shown in Fig. 2(c) and that using the flat pricing scheme in [17],  $P^{flat}$ , is shown in Fig. 2(d). We also show the difference between  $P^{opt}$  and  $P^{flat}$  in Fig. 2(e). From the simulation results, we can see that the optimal contract-based mechanism enables the aggregator to exploit different preferences of EVs and therefore to extract more profit while achieving the service request statistically. On the other hand, in the pricing scheme in [17], the aggregator always has to overpay the EVs, which results in a loss of profits to the aggregator.

## VII. CONCLUSIONS

In this paper, we study the problem of stimulating self-interested EVs in providing ancillary services to the power grid. A heterogeneous setting is considered where EVs have different preferences toward charging/discharging at different time. We formulate the interactions between the aggregator and EVs as an optimal contract design problem and characterize the optimal contract for regular distributions. The derived optimal contract takes a very simple form where the aggregator only needs to publish an optimal unit price to EVs and therefore can be implemented very efficiently. By using the optimal contract-based mechanism, the aggregator can maximize its profits while coordinating EVs to satisfy the service request. Although calculating the optimal unit price explicitly requires the statistical distribution of EVs' preferences, the case without

knowing such statistical distributions has also been investigated. In particular, we propose a learning algorithm for the aggregator to learn the optimal unit price through its interactions with EVs, which has a provably logarithmic upper bound on regret.

## APPENDIX A PROOF OF LEMMA 1

We introduce another random variable  $\hat{T}_i(t)$  to represent the number of times  $\alpha_i$  is chosen after  $t$  rounds of the proposed policy, i.e.,

$$\hat{T}_i(t) = \sum_{\tau=1}^t \mathbf{1}(\sigma_\tau = i). \quad (28)$$

Clearly, we have  $T_i(t) \leq \hat{T}_i(t)$  for every  $0 \leq i \leq K$  and every  $t \geq 1$ .

Recall that, for any service request  $\Delta_j$  and any unit price  $\alpha_i$ ,  $Y_{i,j}$  is a random variable with mean  $\mu_{i,j}$  which is independent over time. If unit price  $\alpha_i$  has been chosen  $s$  times, we can have  $s$  i.i.d. realizations of  $Y_{i,j}$  for every  $1 \leq j \leq N$ . Denote by  $\{Y_{i,j,k} | k = 1, \dots, s\}$  the sequence of  $s$  i.i.d. random variables corresponding to these realizations. Then we can write the sample mean as

$$\bar{Y}_{i,j,s} = \frac{1}{s} \sum_{k=1}^s Y_{i,j,k}. \quad (29)$$

Let  $h$  be an arbitrary positive integer. Then, for an arbitrary sequence of service requests  $\{\Delta_{j_\tau} | \tau = 1, \dots, t\}$ , We have

$$\begin{aligned} T_i(t) &\leq 1 + \sum_{\tau=K+2}^t \mathbf{1}(\sigma_\tau = i, \mu_{i,j_\tau} > \mu_{j_\tau}^*) \\ &\leq h + \sum_{\tau=K+2}^t \mathbf{1}(\sigma_\tau = i, \mu_{i,j_\tau} > \mu_{j_\tau}^*, T_i(\tau-1) \geq h) \end{aligned} \quad (30)$$

Based on (30) and after some manipulations, we can bound  $T_i(t)$  by the inequality in (31).

Notice that  $\bar{Y}_{i,j_{\tau+1},s_i} - \sqrt{\frac{2\ln(\tau)}{s_i}} \leq \bar{Y}_{i_{\tau+1},j_{\tau+1},s} - \sqrt{\frac{2\ln(\tau)}{s}}$  implies at least one of the following must hold

$$\bar{Y}_{i,j_{\tau+1},s_i} \leq \mu_{i,j_{\tau+1}} - \sqrt{\frac{2\ln(\tau)}{s_i}}, \quad (32)$$

$$\bar{Y}_{i_{\tau+1},j_{\tau+1},s} \geq \mu_{j_{\tau+1}}^* + \sqrt{\frac{2\ln(\tau)}{s}}, \quad (33)$$

$$\mu_{j_{\tau+1}}^* > \mu_{i,j_{\tau+1}} - 2\sqrt{\frac{2\ln(\tau)}{s_i}}. \quad (34)$$

We can bound the probability of events (32) and (33) using the Chernoff-Hoeffding bound [23] as

$$Pr \left( \bar{Y}_{i,j_{\tau+1},s_i} \leq \mu_{i,j_{\tau+1}} - \sqrt{\frac{2\ln(\tau)}{s_i}} \right) \leq e^{-4\ln \tau} = \tau^{-4},$$

and

$$Pr \left( \bar{Y}_{i_{\tau+1},j_{\tau+1},s} \geq \mu_{j_{\tau+1}}^* + \sqrt{\frac{2\ln(\tau)}{s}} \right) \leq e^{-4\ln \tau} = \tau^{-4}.$$

$$\begin{aligned}
T_i(t) &\leq h + \sum_{\tau=K+2}^t \mathbf{1}(\bar{Y}_{i,j_\tau,\hat{T}_i(\tau-1)} - \sqrt{\frac{2\ln(\tau-1)}{\hat{T}_i(\tau-1)}} \leq \bar{Y}_{i_\tau^*,j_\tau,\hat{T}_{i_\tau^*}(\tau-1)} - \sqrt{\frac{2\ln(\tau-1)}{\hat{T}_{i_\tau^*}(\tau-1)}}, \mu_{i,j_\tau} > \mu_{j_\tau}^*, T_i(\tau-1) \geq h) \\
&\leq h + \sum_{\tau=K+1}^t \mathbf{1}(\bar{Y}_{i,j_{\tau+1},\hat{T}_i(\tau)} - \sqrt{\frac{2\ln(\tau)}{\hat{T}_i(\tau)}} \leq \bar{Y}_{i_{\tau+1}^*,j_{\tau+1},\hat{T}_{i_{\tau+1}^*}(\tau)} - \sqrt{\frac{2\ln(\tau)}{\hat{T}_{i_{\tau+1}^*}(\tau)}}, \mu_{i,j_{\tau+1}} > \mu_{j_{\tau+1}}^*, \hat{T}_i(\tau) \geq h) \\
&\leq h + \sum_{\tau=K+1}^t \mathbf{1}(\min_{h \leq s_i < \tau} \bar{Y}_{i,j_{\tau+1},s_i} - \sqrt{\frac{2\ln(\tau)}{s_i}} \leq \max_{1 \leq s < \tau} \bar{Y}_{i_{\tau+1}^*,j_{\tau+1},s} - \sqrt{\frac{2\ln(\tau)}{s}}, \mu_{i,j_{\tau+1}} > \mu_{j_{\tau+1}}^*) \\
&\leq h + \sum_{\tau=K+1}^t \sum_{s=1}^{\tau} \sum_{s_i=h}^{\tau} \mathbf{1}(\bar{Y}_{i,j_{\tau+1},s_i} - \sqrt{\frac{2\ln(\tau)}{s_i}} \leq \bar{Y}_{i_{\tau+1}^*,j_{\tau+1},s} - \sqrt{\frac{2\ln(\tau)}{s}}, \mu_{i,j_{\tau+1}} > \mu_{j_{\tau+1}}^*). \tag{31}
\end{aligned}$$

$$\mathbb{E}[T_i(t)] \leq \left\lceil \frac{8\ln t}{d_{min}^2} \right\rceil + \sum_{\tau=K+1}^t \sum_{s=1}^{\tau} \sum_{s_i=\lceil \frac{8\ln t}{d_{min}^2} \rceil}^{\tau} \left[ Pr\left(\bar{Y}_{i,j_{\tau+1},s_i} \leq \mu_{i,j_{\tau+1}} - \sqrt{\frac{2\ln(\tau)}{s_i}}\right) + Pr\left(\bar{Y}_{i_{\tau+1}^*,j_{\tau+1},s} \geq \mu_{j_{\tau+1}}^* + \sqrt{\frac{2\ln(\tau)}{s}}\right) \right]. \tag{35}$$

Moreover, under the condition of  $\mu_{i,j_{\tau+1}} > \mu_{j_{\tau+1}}^*$ , we have

$$\mu_{j_{\tau+1}}^* - \mu_{i,j_{\tau+1}} + 2\sqrt{\frac{2\ln(\tau)}{s_i}} \leq \mu_{j_{\tau+1}}^* - \mu_{i,j_{\tau+1}} + d_{min} \leq 0$$

for  $s_i \geq (8\ln t)/d_{min}^2$ , which implies that we can make event (34) false by setting  $h = \lceil \frac{8\ln t}{d_{min}^2} \rceil$ . Therefore, we derive an upper bound for  $\mathbb{E}[T_i(t)]$  as shown in (35). Then, from (35), we have

$$\begin{aligned}
\mathbb{E}[T_i(t)] &\leq \left\lceil \frac{8\ln t}{d_{min}^2} \right\rceil + \sum_{\tau=K+1}^t \sum_{s=1}^{\tau} \sum_{s_i=\lceil \frac{8\ln t}{d_{min}^2} \rceil}^{\tau} 2\tau^{-4} \\
&\leq \frac{8\ln t}{d_{min}^2} + 1 + \sum_{\tau=1}^t \sum_{s=1}^{\tau} \sum_{s_i=1}^{\tau} 2\tau^{-4} \\
&\leq \frac{8\ln t}{d_{min}^2} + 1 + \frac{\pi^2}{3}.
\end{aligned}$$

## REFERENCES

- [1] "Electrification roadmap: Revolutionizing transportation and achieving energy security," Nov. 2009, [Online]. Available: <http://www.electrificationcoalition.org/reports/EC-Roadmap-screen.pdf>.
- [2] W. Kempton and J. Tomic, "Vehicle-to-grid power implementation: From stabilizing the grid to supporting large-scale renewable energy," *J. Power Sources*, vol. 144, no. 1, pp. 280-294, Jun. 2005.
- [3] W. Kempton and J. Tomic, "Vehicle-to-grid power fundamentals: Calculating capacity and net revenue," *J. Power Sources*, vol. 144, no. 1, pp. 268-279, 2005.
- [4] C. Guille and G. Gross, "A conceptual framework for the vehicle-to-grid (V2G) implementation," *Energy Policy*, vol. 37, no. 11, pp. 4379-4390, Nov. 2009.
- [5] A. Ipakchi and F. Albuyeh, "Grid of the future," *IEEE Power Energy Magazine*, vol. 7, no. 2, pp. 52-62, Mar. 2009.
- [6] J. Laffont and D. Martimort, *The Theory of Incentives: The Principal-Agent Model*, Princeton University Press, 2001.
- [7] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time Analysis of the Multiarmed Bandit Problem," *Machine Learning*, vol. 47, no. 2, pp. 235-256, 2002.
- [8] J. Xu and V. Wong, "An approximate dynamic programming approach for coordinated charging control at vehicle-to-grid aggregator," in *Proc. IEEE Int. Conf. Smart Grid Commun.*, Oct. 2011.
- [9] D. Wu, D. C. Aliprantis and L. Ying, "Load scheduling and dispatch for aggregators of plug-in electric vehicles," *IEEE Trans. Smart Grid*, vol. 3, no. 1, pp. 368-376, Mar. 2012.
- [10] S. Han, S. Han and K. Sezaki, "Development of an optimal vehicle-to-grid aggregator for frequency regulation," *IEEE Trans. Smart Grid*, vol. 1, no. 1, pp. 65-72, Jun. 2010.
- [11] E. Sortomme and M. A. El-Sharkawi, "Optimal charging strategies for unidirectional vehicle-to-grid," *IEEE Trans. Smart Grid*, vol. 2, no. 1, pp. 131-138, Mar. 2011.
- [12] E. Sortomme and M. A. El-Sharkawi, "Optimal scheduling of vehicle-to-grid energy and ancillary services," *IEEE Trans. Smart Grid*, vol. 3, no. 1, pp. 351-359, Mar. 2012.
- [13] C. Ibars, M. Navarro, and L. Giupponi, "Distributed demand management in smart grid with a congestion game," in *Proc. 1st IEEE Int. Conf. Smart Grid Commun.*, Oct. 2010.
- [14] P. Samadi, A. Mohsenian-Rad, R. Schober, V. Wong, and J. Jatskevich, "Optimal real-time pricing algorithm based on utility maximization for smart grid," in *Proc. 1st IEEE Int. Conf. Smart Grid Commun.*, Oct. 2010.
- [15] H. K. Nguyen, J. B. Song, and Z. Han, "Demand Side Management to Reduce Peak-to-Average Ratio using Game Theory in Smart Grid," *IEEE INFOCOM Workshop on Green Networking and Smart Grids*, March 2012.
- [16] Y. Chen, W.S. Lin, F. Han, Y.H. Yang, Z. Safar, and K.J.R. Liu, "A Cheat-Proof Game Theoretic Demand Response Scheme for Smart Grids", *Proc. IEEE International Conference on Communications (ICC)*, Ottawa, June 2012.
- [17] C. Wu, H. Mohsenian-Rad and J. Huang, "Vehicle-to-aggregator interaction game," *IEEE Trans. Smart Grid*, vol. 3, no. 1, pp. 434-442, Mar. 2012.
- [18] T. Lai and H. Robbins, "Asymptotically efficient adaptive allocation rules," *Advances in Applied Mathematics*, vol. 6, no. 1, pp. 4-22, 1985.
- [19] Z. Fan, "A distributed demand response algorithm and its application to PHEV charging in Smart Grids", to appear, *IEEE Trans. Smart Grid*.
- [20] D. Fudenberg and J. Tirole, *Game Theory*, The MIT Press, 1991.
- [21] S. Boyd, L. Vandenberghe, *Convex Optimization*, Cambridge Univ. Press, Cambridge, U.K., 2004.
- [22] R. Myerson, "Optimal auction design," *Math. Oper. Res.*, vol. 6, 1981.
- [23] W. Hoeffding, "Probability Inequalities for Sums of Bounded Random Variables," *American Statistical Association Journal*, 1963, pp. 13-30.
- [24] *SAE Electric Vehicle and Plug-in Hybrid Electric Vehicle Conductive Charge Coupler*, SAE J1772, Jan. 2010.