Eliciting Truthful Reports with Partial Signals in Repeated Games

Yutong Wu[1]($\boxtimes$), Ali Khodabakhsh[1], Bo Li[2], Evdokia Nikolova[1],
and Emmanouil Pountourakis[3]

[1] The University of Texas at Austin, Austin, TX 78712, USA
`{yutong.wu,ali.kh}@utexas.edu, nikolova@austin.utexas.edu`
[2] The Hong Kong Polytechnic University, Hung Hom, Hong Kong, China
`comp-bo.li@polyu.edu.hk`
[3] Drexel University, Philadelphia, PA 19104, USA
`manolis@drexel.edu`

Abstract. We consider a repeated game where a player self-reports her usage of a service and is charged a payment accordingly by a center. The center observes a partial signal, representing part of the player's true consumption, which is generated from a publicly known distribution. The player can report any value that does not contradict the signal and the center issues a payment based on the reported information. Such problems find application in net metering billing in the electricity market, where a customer's actual consumption of the electricity network is masked and complete verification is impractical. When the underlying true value is relatively constant, we propose a penalty mechanism that elicits truthful self-reports. Namely, besides charging the player the reported value, the mechanism charges a penalty proportional to her inconsistent reports. We show how fear of uncertainty in the future incentivizes the player to be truthful today. For Bernoulli distributions, we give the complete analysis and optimal strategies given any penalty. Since complete truthfulness is not possible for continuous distributions, we give approximate truthful results by a reduction from Bernoulli distributions. We also extend our mechanism to a multi-player cost-sharing setting and give equilibrium results.

Keywords: Energy economics · Equilibrium analysis · Repeated games · Electricity market · Penalty mechanisms

1 Introduction

Consider the following repeated game where a center owns resources and one or more strategic players pay the center to consume the resources. In every round, a player self-reports their usage, which will then be used to determine their payment to the center. However, it is not always possible for the center to verify the submitted information from the players. Instead, only part of the actual consumption is revealed to the center based on some publicly known distribution.

A player can report any value that is at least the revealed amount. Without any external interference, a player will naturally report exactly the revealed amount (potentially lower than the true consumption) to minimize their payment. The center then needs to determine a payment mechanism such that each player is incentivized to report their true value.

The electricity market is facing precisely the described problem. As the number of electricity *prosumers* increases each year, new rate structures are designed to properly calculate the electricity bill for this special type of consumer while ensuring that every customer is still paying their fair share of the network costs. Prosumers are those who not only consume energy but also produce electricity via distributed energy resources such as rooftop solar panels. Among different rate structures, *net metering* is a popular billing mechanism that is currently adopted in more than 40 states in the US [22]. Net metering charges prosumers a payment proportional to their net consumption, i.e., gross consumption minus the production [26], demonstrated in Fig. 1. The payment includes the electricity usage as well as grid costs that are incurred by using the electricity network.
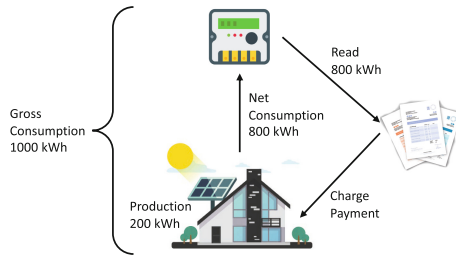

Fig. 1. Net metering for electricity prosumers.

The controversy in net metering lies in that prosumers fail to pay their share of the grid costs when they do not have local storage equipment [9]. In the United States, only 4% of the solar panel owners also own a battery to store the produced solar energy [18]. For those who do not own battery storage, the generated power has to be transmitted back to the grid. Accordingly, the daily consumption of power by these prosumers also needs to come from the grid instead of directly from the solar panels. In this way, most prosumers have under-paid their share of the network costs and become "free-riders" of the electricity grid. The grid is often subject to costly line upgrades and net metering unevenly shifts such costs to traditional consumers, who usually come from lower-income households [15]. Indeed, previous research works have suggested that prosumers should pay a part of the grid costs proportionally to their gross consumption, not net consumption [9,16]. However, the gross consumption is hidden from the utility companies since only net consumption can be observed from the meter. Meanwhile, there is no incentive for prosumers to voluntarily report their true consumption as it will only increase their electricity bills.

Fortunately, the production from solar panels usually follows some pattern while the gross consumption of electricity for a typical household stays relatively constant, which is especially true for industrial sites – the major consumers for utilities [21,27]. Thus, the observed consumption can be assumed to follow some natural distribution and the center is able to detect dishonesty when a player's report differs from their reporting history. With this idea, we propose a simple penalty mechanism, the *flux mechanism*, that elicits truthful reports from players in a repeated game setting when only partial verification is possible. Particularly, a player is charged their reported value as well as a penalty due to inconsistency in consecutive reports in each round. The main goal is to ensure that every player reports their true values and no penalty payment is collected. We show that the combining effect of (i) the penalty rate and (ii) the length of the game is sufficient for inducing truthful behavior from the player for the entire game. As the horizon of the game increases, the minimum penalty rate for truth-telling as an optimal strategy decreases. In other words, it is the fear of uncertainty in the future that incentivizes the player to be truthful today.

1.1   Our Contribution

We address the problem of eliciting truthful reports when the center is able to observe a part of the player's private value based on some publicly known distribution. The strategic player reports some value that is at least the publicly revealed value and is charged a payment accordingly. We propose a truth-eliciting mechanism, *flux mechanism*, that utilizes the player's fear of uncertainties to achieve truthfulness. In each round, the player is charged a "regular payment" proportional to the consumption they report. Starting from the second round, the player is charged an additional "penalty payment", which is r times the (absolute) difference between the reports in the current and the previous round, where the penalty rate r is set by the center before the game starts.

Intuitively, a player can save their regular payment by under-reporting their consumption, but they will then face the uncertainty of paying penalties in future rounds due to inconsistent reports. Under most settings, if r is set to be infinitely high, the players will be completely truthful to avoid any penalty payment. However, a severe punishment rule is undesirable and discourages players from participating. Therefore, we want to understand the following question.

What is the minimum penalty rate such that the player is willing to report their true value?

We observe that no finite penalty can achieve complete truthfulness for arbitrary distributions as a player's true consumption may never be revealed exactly. We can, however, obtain approximate truthfulness for a general distribution by analyzing complete truthfulness for a corresponding Bernoulli distribution. For $Ber(p)$, the partial signal equals the true consumption with probability p and 0 with probability $1 - p$ for $p \in (0, 1)$. We give results for Bernoulli distributions in Main Results 1 and 2. For arbitrary distributions, we redefine p as the probability of having a partial signal that is at least α times the true consumption, for $\alpha \in [0, 1]$, to obtain α-truthfulness (Main Result 3).

Main Result 1 (Theorem 1). *For a T-round game with Bernoulli distribution Ber(p), the player is completely truthful if and only if the penalty rate is at least*

$$\frac{1 - (1 - p)^T}{p - p(1 - p)^{T-1}}.$$

Main Result 1 gives the minimum penalty rate that guarantees complete truthfulness for $Ber(p)$ distributions. We also want to understand how players would behave if the penalty rate is not as high, which describes the situation when the center is willing to sacrifice some degree of truthfulness by lowering the penalty rate. Given any penalty rate, we show that a player's optimal strategies can be described as one or a combination of three basic strategies, *lying-till-end, lying-till-busted* and *honest-till-end*. Specifically, with a low penalty rate, the player is always untruthful to save regular payment, i.e., lying-till-end is optimal. As the penalty rate increases, the player's optimal strategy *gradually* moves to lying-till-busted, which is to be untruthful until the partial signal is revealed as the true consumption for the first time and then stays truthful for the rest of the game. When the penalty rate is sufficiently high, the player would avoid lying completely and reports the truth, i.e., she is honest-till-end.

Table 1. Optimal strategy given penalty rate r under $Ber(p)$ distributions

Bernoulli Prob.	Penalty Rate	Optimal Strategy
$p \geq 0.5$	$r \leq \frac{1}{2p}$	lying-till-end
	$\frac{1}{2p} < r \leq 1$	lying-till-busted + lying last round
	$1 < r < \frac{1-(1-p)^T}{p-p(1-p)^{T-1}}$	lying-till-busted
	$r \geq \frac{1-(1-p)^T}{p-p(1-p)^{T-1}}$	honest-till-end
$p < 0.5$	$r \leq 1$	lying-till-end
	$h(t-1) < r \leq h(t)$	lying-till-end first t rounds + lying-till-busted for rest
	$h(T-1) < r < \frac{1-(1-p)^T}{p-p(1-p)^{T-1}}$	lying-till-busted
	$r \geq \frac{1-(1-p)^T}{p-p(1-p)^{T-1}}$	honest-till-end

Main Result 2 (Theorems 1 and 2). *For a T-round game with Bernoulli distribution Ber(p), given any penalty rate r, the player's optimal strategy is summarized in Table 1, where*

$$h(t) = \frac{1 - (1 - p)^t}{2p - p(1 - p)^{t-1}}, for\, 1 \leq t \leq T.$$

For arbitrary distributions, including uniform distributions, it is impossible to obtain complete truthfulness without setting the penalty to infinity. Main Result 3 gives a reduction from Bernoulli distributions to general distributions for approximate truthfulness.

Main Result 3 (Theorem 3). *Given $\alpha \in [0,1]$ and an arbitrary distribution with CDF F, if a penalty rate r achieves complete truthfulness for $Ber(p)$ where $p = 1 - F(\alpha D)$ and D is the player's true gross consumption, then the same r achieves α-approximate truthfulness for distribution F.*

Finally, we extend our results to multiple players. We note that if the players are charged independently, applying the flux mechanism to each individual elicits truthful reports. A more complicated and realistic setting is the cost-sharing problem where the players split an overhead cost based on their submitted reports. We propose the *multi-player flux mechanism* where the penalty payment is the same as before but the regular payment is now a share of some overhead cost. Again, if the penalty rate is sufficiently high, the players stay truthful, regardless of others' behavior, to avoid any penalty payment, i.e., the truthful report profile forms a *dominant strategy equilibrium*. As the penalty rate decreases, the truthfulness of a player may depend on other players' actions. That is, with a lower penalty rate, a truthful report profile forms a *Nash equilibrium*. For both equilibrium definitions, we are interested in the following question.

> *What is the minimum penalty rate for the truthful report profile to form a dominant strategy or Nash equilibrium?*

We give exact penalty thresholds for both truthful equilibria under Bernoulli distributions and use a reduction to obtain approximate results under arbitrary distributions in Main Result 4.

Main Result 4 (Theorems 4, 5, 6 and 7). *For any T-round game with distribution $Ber(p)$, a truthful strategy profile is a dominant strategy equilibrium if and only if*

$$r \geq \frac{C}{nD} \frac{1 - (1-p)^{n-1}}{p} \frac{1 - (1-p)^T}{p - p(1-p)^{T-1}},$$

and a Nash equilibrium if and only if

$$r \geq \frac{C}{nD} \frac{1 - (1-p)^T}{p - p(1-p)^{T-1}}.$$

Given $\alpha \in [0,1]$ and any distribution with cumulative distribution function F, let $p = 1 - F(\alpha D)$, where D is the true gross consumption. Then α-approximate truthful profile is a Nash equilibrium if

$$r \geq \frac{1}{\alpha} \frac{C}{nD} \frac{1 - (1-p)^T}{p - p(1-p)^{T-1}},$$

and the α-approximate truthful profile is a dominant strategy equilibrium if

$$r \geq \frac{1}{\alpha} \frac{C}{nD} \frac{1 - (1-p)^n}{p} \frac{1 - (1-p)^T}{p - p(1-p)^{T-1}}.$$

1.2   Related Works

The economic effect of the net metering policy has been explored for different countries and regions [6,7,17,24,28]. It has been observed that net metering can cause inequality issues for traditional energy consumers [8,15,17,23]. Accordingly, alternative pricing mechanisms and tariff structures have been proposed to fairly compensate the energy production [4,9,10,16,25]. In particular, Gautier *et al.* [9] and Khodabakhsh *et al.* [16] proposed that individuals should be charged based on their true consumption, not net consumption. Our work is a continuation of [16], where a primitive version of the penalty mechanism is first proposed for promoting a fairer electricity rate structure. We formally define the mechanism and provide the corresponding theoretical analysis.

More broadly, fairness for the power grid has become an increasingly popular subject. First, Heylen *et al.* provided various indices to measure fairness and inequality in power system reliability [13]. Fairness is also explored for load shedding plans [14], electric vehicle charging schemes [1], demand response [2], etc. Moret and Pinson showed fairness can be improved with a "community-based electricity market", where prosumers are allowed to share their production on the community level [20]. Our model, on the other hand, addresses the fairness issues by modifying the current electricity structure, which is easier for utility companies to adopt.

Theoretically, our work is related to information elicitation with limited verification ability. Caragiannis *et al.* [5] and Ball *et al.* [3] worked on probabilistic verification where a lying player may be caught by a probability based on her type. Our work can be viewed as an extension of probabilistic verification under a repeated game setting where the verification is implicit and the main goal is to incentivize truthful reports. Another related problem is strategic classification, in which individuals can manipulate their input to obtain a better classification outcome [11,12,19,30]. Although we also consider the strategic behavior of players in a sequential game, our model is quite different. Strategic classification allows the center to learn the patterns of the players via a private classifier. In our setting, the scoring rule is transparent to the players and an additional measure (e.g., a penalty) has to be invoked to incentivize truthfulness.

2   Problem Statement

We formally define our problem under the single player setting and defer the extension to multiple players to Sect. 5. The player has a gross consumption $D \geq 0$, which is her private information. The game has T rounds where $T > 1$ as otherwise the flux mechanism becomes invalid. In each round t, the center observes a partial signal, $y_t \leq D$, which is randomly and independently drawn from a distribution F supported on $[0, D]$. We use $r \geq 0$ to denote the penalty rate. In a flux mechanism, a player cares more about the number of rounds left in the future rather than the number of rounds has passed. Thus we use $t = T, T - 1, \cdots, 1$ to denote the current round, where t *means there are t rounds left*, including the current round. For example, the first round is round

T, the last round is round 1, and the previous round of round t is round $t + 1$. For round $t \leq T$, the flux mechanism runs as follows.

- The center observes the player' net consumption $y_t \sim F$.
- The player submits their reported gross consumption which is at least the net consumption, $b_t \geq y_t$. The player may not be truthful, i.e., $b_t \leq D$.
- When $t < T$, the player's payment consists of regular payment b_t and penalty payment $r \cdot |b_{t+1} - b_t|$. When $t = T$, the player only pays the regular payment.

For $t < T$, we call b_{t+1} the *history* of round t. In each round t, the player wants to pay the lowest expected total payment by reporting b_t without knowing the partial signals for future rounds. We call a mechanism *truthful* if the player reports D for all rounds. When two reports bring the same expected payment, we break tie in favor of truthfulness. We adopt the assumption from Khodabakhsh *et al.* [16] that D does not vary with t. We explain in the full version of this paper an easy extension where D_t is drawn from a known range $[\underline{D}, \overline{D}]$.

3   Bernoulli Distributions

We start with the analysis of Bernoulli distribution as we show later a reduction from an arbitrary distribution to a Bernoulli distribution. We prove it is only optimal for a player to report zero or their true consumption in each round. The optimal strategies can then be characterized by three basic strategies (Definition 1). The penalty thresholds are computed by comparing the different combinations of the basic strategies. Due to space limit, we defer most proofs to the full version of this paper and focus on explaining the intuition in this section.

3.1   Basic Strategies

In a Bernoulli distribution setting, in each round t, the partial signal y_t is D with probability p and 0 with probability $1 - p$. When the partial signal equals to the private value, i.e., $y_t = D$, we say that the player is *"busted"* in round t. We first define three basic strategies.

Definition 1 (Basic Strategies). *For Bernoulli distributed net consumption $y_t \sim Ber(p)$, we define the following as the three basic strategies:*

- *lying-till-end: Report $b_t = 0$ when $y_t = 0$ and $b_t = D$ otherwise;*
- *lying-till-busted: Report $b_t = 0$ until $y_t = D$ for the first time, then report D for all future rounds;*
- *honest-till-end: Report $b_t = D$ for all rounds.*

We note that a player's optimal strategy for a given penalty rate r can be solved by backward induction. Let $OptCost(t, r, b_{t+1})$ denote the optimal expected cost for a player starting in round t with penalty rate r and report b_{t+1} for the previous round. Then

$$OptCost(t, r, b_{t+1}) = \min_{b_t} ExpCost(t, r, b_{t+1}, b_t),$$

where $ExpCost(t, r, b_{t+1}, b_t)$ is the expected cost for the player starting in round t and reporting b_t (if she is allowed to), with penalty rate r and history b_{t+1},

$$
\begin{aligned}
ExpCost&(t, r, b_{t+1}, b_t) \\
&= \mathbb{E}_{y_t}[\max\{y_t, b_t\} + r|\max\{y_t, b_t\} - b_{t+1}| + OptCost(t-1, r, \max\{y_t, b_t\})] \\
&= p\big(D + r(D - b_{t+1}) + OptCost(t-1, r, D)\big) \\
&\qquad + (1-p)\big(b_t + r|b_t - b_{t+1}| + OptCost(t-1, r, b_t)\big).
\end{aligned}
$$

The first term on the right side of the equation above refers to the cost when the partial signal is revealed as D and the player has to report D. The second term refers to the cost when the partial signal is 0 and the player chooses to report b_t. Let $OptCost(0, r, b_1) = 0$ for all b_1. When $t = T$, i.e., the first round, there is no history b_{T+1} and the player wants to minimize the following total cost,

$$
\begin{aligned}
OptCost(T, r) &= \min_{b_T} ExpCost(T, r, b_T) \\
&= p(D + OptCost(t-1, r, D)) + (1-p)(b_T + OptCost(t-1, r, b_T)).
\end{aligned}
$$

Solving the recursion will give the characterization of optimal strategies in Table 1, as we demonstrate in the full version of this paper. In what follows, we discuss a surprisingly simpler and more constructive proof by exploiting the properties of the flux mechanism, which may be of independent interest.

3.2   Main Theorems

We observe that there are two key elements that influence the decision making of the player.

(1) The player's history, b_{t+1} for $t < T$. The value of b_{t+1} directly affects the penalty payment in round t. Intuitively, a player is more reluctant to lie if b_{t+1} is high and better off lying if b_{t+1} is small.
(2) The number of rounds left to play, i.e., t, indirectly influences the probability and the number of times a player will be busted in the remaining rounds.

Via Lemmas 1–4, we show these are the *only two* elements that determine a rational player's action. The following lemma shows that it is not optimal for a player to report a value strictly between 0 and D. Moreover, if a player is untruthful in the previous round, it is better off to remain untruthful. With this lemma, we largely reduce the strategy space we need to consider.

Lemma 1. *For any round $t \leq T$, given $y_t = 0$, the optimal report in round t is $b_t \in \{0, D\}$. Moreover, if $t < T$ and $b_{t+1} = y_t = 0$, then the optimal report is $b_t = 0$.*

Next, we prove that in each round, the optimal strategy is determined by a penalty threshold such that a player will be truthful if and only if the penalty rate r is above the threshold. We call them *critical thresholds*.

Lemma 2 (Critical Thresholds). *For $t = T$, there is a threshold penalty rate $r_T^{(\emptyset)} \geq 0$ such that reporting D is optimal if and only if the penalty rate is at least $r_T^{(\emptyset)}$; For $t < T$, there is a threshold penalty rate $r_t^{(b_{t+1})} \geq 0$ such that reporting D is optimal for a player in round t with history b_{t+1} if and only if the penalty rate is at least $r_t^{(b_{t+1})}$.*

Lemmas 1 and 2 together imply that the optimal strategy can only be one or a combination of the basic strategies. In particular, by Lemma 1, $r_t^{(0)} = \infty$ for any t. Moreover, since b_{t+1} can only be 0 or D, by Lemma 2, we only need to determine the values of $r_T^{(\emptyset)}$ and $r_t^{(D)}$ for $t < T$ to complete the picture of optimal strategies. We now give some properties of these thresholds.

Lemma 3. $r_t^{(\emptyset)} \geq r_t^{(D)}$ *for $t \in \{1, \ldots, T\}$.*

Given the same t rounds left, Lemma 3 says a player is more inclined to lie without a history than with a truthful history. This is straightforward as lying with a truthful history results in an additional penalty payment.

Lemma 4. *Given $r_t^{(\emptyset)} \geq \frac{1}{p}$, $r_t^{(\emptyset)}$ decreases as t increases.*

Lemmas 3 and 4 together tell us the player is least incentivized to be truthful on the first round and $r_T^{(\emptyset)}$ is the penalty threshold that ensures truthfulness for the game. We give this important threshold in Theorem 1.

Theorem 1. *The minimum penalty for truthful reporting in a game of T rounds with $Ber(p)$ distribution is*

$$r_T^{(\emptyset)} = \frac{1 - (1-p)^T}{p - p(1-p)^{T-1}}. \tag{1}$$

We see $r_T^{(\emptyset)} \to 1/p$ as $T \to \infty$ and $r_T^{(\emptyset)}$ decreases as T increases. This implies the increasing length of the game incentivizes the player to speak the truth today, even when they do not have to. To understand Theorem 1, we observe that it is sufficient to compare lying-till-busted and honest-till-end since $r_T^{(\emptyset)}$ ensures the player to stay truthful after being busted. Before the player is busted for the first time, it is not optimal to oscillate between lying and truth-telling, as it is strictly dominated by lying completely. Therefore, the only viable strategies are lying-till-busted and honest-till-end, and the desired threshold sets the expected cost of these two strategies equal.

With a more involved argument, we get the exact values for the truthful threshold given a truthful history, i.e., the $r_t^{(D)}$'s. The values of $r_T^{(\emptyset)}$ and $r_t^{(D)}$ characterize the optimal strategies for a player and are an alternative representation of Table 1.

Theorem 2. *For $p \leq \frac{1}{2}$, $r_t^{(D)} = \frac{1-(1-p)^t}{2p-p(1-p)^{t-1}}$. For $p > \frac{1}{2}$, $r_t^{(D)} = 1$ for $t = 1$ and $r_t^{(D)} = \frac{1}{2p}$ for $t \geq 2$.*

The optimal strategy is visualized in Figs. 2a and 2b for $p = 0.3$ and $p = 0.7$, respectively. The x-axis is the number of rounds left (t), and the y-axis is the penalty thresholds for truthfulness. We give examples of penalties via the red dashed lines. For the first round, the player refers to the blue dot representing $r_T^{(\emptyset)}$ and is truthful if and only if the penalty is above the blue dot. Afterwards, given t rounds left and history D, the player looks at the green curve representing $r_t^{(D)}$ and is only truthful if the penalty is above the curve. If the history is 0, she remains untruthful and reports 0. Figures 2a and 2b visualize the optimal strategies given in Table 1. Both green curves are closely related to $\frac{1}{2p}$. An intuition is that in any round $t < T$, a player pays D if she is truthful and roughly $2prD$ if she lies, where the penalty payment rD comes from the previous and the next round, each with probability p. The penalty that sets these two costs equal is $\frac{1}{2p}$. The actual $r_t^{(D)}$ thresholds vary upon values of t and p.
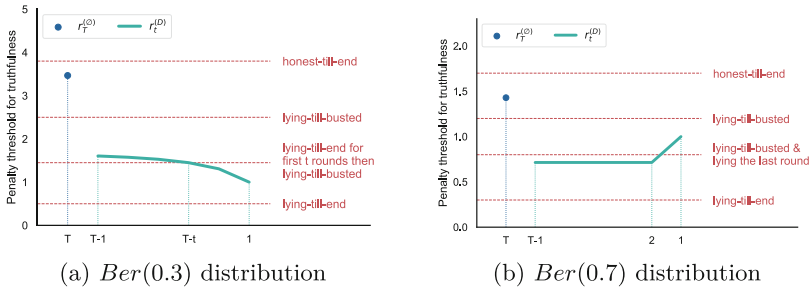

(a) $Ber(0.3)$ distribution   (b) $Ber(0.7)$ distribution

Fig. 2. Critical thresholds under two distributions with sample optimal strategies.

4 A Reduction for Arbitrary Distributions

As discussed in the introduction, only the infinite penalty rate will guarantee complete truthfulness under arbitrary distributions, yet there is still hope to obtain approximate results. The trick is to redefine being busted as having a partial signal that is less than α times the true consumption, for $\alpha \in [0, 1]$. Then any arbitrary distribution is reduced to $Ber(p)$ where p is the probability that the partial signal is at least αD. For approximate truthfulness, we define being α-truthful as reporting at least αD. We reuse the arguments of comparing basic strategies from Sect. 3 to determine an upper bound for the penalty rate that guarantees α-truthfulness. We introduce the notion of approximate truthfulness in Definition 2 and give the reduction in Theorem 3. We demonstrate the reduction with uniform distributions in Example 1.

Definition 2 (α-truthfulness). *A reporting $\boldsymbol{b}$ is α-truthful when $b_t \geq \alpha D$ for all $t = 1, \ldots, T$.*

Theorem 3. *Given $\alpha \in [0, 1]$ and an arbitrary distribution with CDF F, if a penalty rate r achieves complete truthfulness for $Ber(p)$ where $p = 1 - F(\alpha D)$, then the same r achieves α-approximate truthfulness for distribution F.*

Example 1. Assume partial signals follow a uniform distribution $U(0, D)$. Let r be the truthful threshold of $Ber(p)$ where $p = 1 - \alpha$, i.e. $r = \frac{1 - \alpha^T}{(1-\alpha)(1-\alpha^{T-1})}$. Then using r ensures α-truthfulness for $U(0, D)$ by Theorem 3. For uniform distributions, it is impossible to obtain complete truthfulness unless $r = \infty$, which can be verified by setting $\alpha = 1$.

5   Extension: A Cost-Sharing Model

We extend the problem to the multi-player setting and focus on cost-sharing among homogeneous players. Let N be the set of players with $n = |N| \geq 1$. Each player $i \in N$ has a private value $x^i \geq 0$, and we assume all players are symmetric, i.e., $x^i = D$ for all $i \in N$ (see the full version of this paper for a relaxation). All players in N split an overhead cost C, which is at least the total gross consumption, i.e., $C \geq nD$. The game has T rounds in total. Given the penalty rate r, we analyze the following multi-player flux mechanism.

- The center observes a partial signal representing player i's net consumption $y_t^i \sim F$ for each player $i \in N$;
- Each player i submits their reported gross consumption that is at least their net consumption, $b_t^i \geq y_t^i$;
- If $t < T$, player i's pays regular payment $C \cdot \frac{b_t^i}{\sum_j b_t^j}$ and penalty payment $r \cdot |b_{t+1}^i - b_t^i|$. If $t = T$, the players only pay regular payments.

We call b_{t+1}^i the *history* for player i in round t and $\boldsymbol{b}_{t+1}$ the *group history*. If everyone lies in a round, the overhead cost is split evenly among all players. A mechanism is *truthful* if every player reports D for every round. We are interested in computing the minimum penalty rates such that truthful reports form a Nash equilibrium (NE) or a dominant strategy equilibrium (DSE). Informally, a strategy profile is a NE if no player wants to unilaterally deviate, and it is a DSE if no player wants to deviate no matter what the other players do. We show that approximate results for any arbitrary distribution can be deducted from an exact analysis for a Bernoulli distribution. Due to space limit, we defer all proofs to the full version of this paper.

Similar to the single-player setting, we avoid solving the recursion by exploiting the properties of the mechanism. Again, we start our analysis with F being a Bernoulli distribution and provide a reduction for approximate truthfulness when F is an arbitrary distribution. In the single-player model with Bernoulli-distributed F, we have shown that it is only optimal for a player to report 0 or her actual consumption D. We claim it is the same case for multiple players. Moreover, if a player lied yesterday and also has an observed consumption of 0 today, they will report 0 regardless of other players' actions.

Lemma 5. *For Bernoulli-distributed F, reporting anything strictly between 0 and D is sub-optimal in a multi-player flux mechanism. Moreover, if $b_{t+1}^i = y_t^i = 0$, it is optimal to report $b_t^i = 0$.*

Starting from this point, we assume that every player reports either 0 or D. When $n = 2$, we show that the multi-player model reduces to the single-player model with a multiplicative factor of $\frac{C}{2D}$. The reason for the reduction is that the savings of switching to lying from being truthful for a player are always $\frac{C}{2}$, regardless of what the other player does.

Lemma 6. *When $n = 2$, the multi-player model reduces to a single-player model. The truthful penalty threshold is $\frac{C}{2D}$ times (1).*

For general n, we show it is sufficient to analyze the maximum difference between lying and truth-telling for player i in round t given group history $\boldsymbol{b}_{t+1}$. In a DSE, a player achieves the biggest gain from lying if all players were lying in the previous round. We then use $\boldsymbol{b}_{t+1} = \boldsymbol{0}$ to compare lying and truth-telling for a player.

Theorem 4. *For the $Ber(p)$ distribution, a truthful strategy profile forms a dominant strategy equilibrium if and only if*

$$r \geq \frac{C}{nD} \frac{1 - (1 - p)^{n-1}}{p} \frac{1 - (1 - p)^T}{p - p(1 - p)^{T-1}}. \tag{2}$$

If we slowly lower the penalty from (2), we will hit a threshold such that truth-telling is an NE. The difference between the truthful NE and the DSE is that now we can assume that every player $j \neq i$ is truthful in the first round and show that player i would not deviate unilaterally. However, we shall not assume that player $j \neq i$ remains truthful for the rest of the game. This is because if player i lies in the first round, player j can observe the report of i in the second round and deviate from truthful behavior. We first show that if $r \geq \frac{C}{nD} \frac{1}{p}$, players with truthful history stay truthful. Then we can safely assume player $j \neq i$ remains truthful throughout the game. In this way, truthful NE is reduced to the case where there is one strategic player and $n - 1$ truthful players. It is not hard to see the threshold is precisely $\frac{C}{nD} \frac{1-(1-p)^T}{p-p(1-p)^{T-1}}$.

Theorem 5. *For the $Ber(p)$ distribution, a truthful strategy profile forms a Nash equilibrium if and only if*

$$r \geq \frac{C}{nD} \frac{1 - (1 - p)^T}{p - p(1 - p)^{T-1}}. \tag{3}$$

We visualize $Ber(p)$ penalty thresholds in Fig. 3 for different T's and p's. The x-axis is the total number of rounds for a game and the y-axis is the penalty rate that guarantees the specified equilibrium. The blue and orange lines are penalty thresholds for $p = \frac{1}{3}$ and $\frac{2}{3}$, respectively. The solid and dashed lines are thresholds for truthful DSE and NE, respectively. All four thresholds in Fig. 3 decrease as T increases, suggesting that the increasing length of the game promotes truthful equilibria. From expressions (2) and (3), we see that the DSE and NE thresholds tend to be the same as p approaches 1.
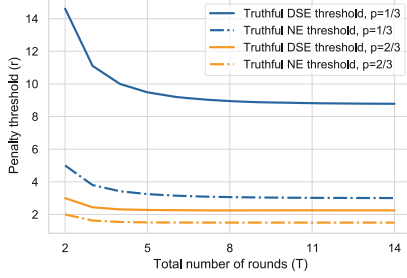
Fig. 3. Exact penalty thresholds for truthful DSE and NE, given the number of rounds T for $Ber(p)$ distributions. We assume $n = 20$, $D = 1$ and $C = n \cdot D = 20$.

Similar to the single-player model, we extend the results for Bernoulli distributions to approximate results for general distributions. Given $\alpha \in [0, 1]$, we redefine being busted as having an observed consumption of at least αD. For the dominant strategy equilibrium, we find the threshold that being α-truthful is a dominant strategy. For Nash equilibrium, we first define the approximate truthful NE, a natural extension of the complete truthful NE.

Theorem 6. *Given $\alpha \in [0, 1]$ and some general distribution F, let $p = 1 - F(\alpha D)$. The α-truthful strategy profile forms a dominant strategy equilibrium if*

$$r \geq \frac{1}{\alpha} \frac{C}{nD} \frac{1 - (1 - p)^n}{p} \frac{1 - (1 - p)^T}{p - p(1 - p)^{T-1}}. \tag{4}$$

Definition 3 (α-truthful Nash equilibrium). *Given $\alpha \in [0, 1]$, a reporting profile $\boldsymbol{b} \in [0, D]^{n \times T}$ is an α-truthful Nash equilibrium if $\boldsymbol{b}_t^i \geq \alpha D$ for all i, t and no player wants to deviate from being α-truthful in any round.*

Theorem 7. *Given $\alpha \in [0, 1]$ and some general distribution F, let $p = 1 - F(\alpha D)$. The α-truthful strategy profile forms a Nash equilibrium if*

$$r \geq \frac{1}{\alpha} \frac{C}{nD} \frac{1 - (1 - p)^T}{p - p(1 - p)^{T-1}}. \tag{5}$$

We see that both the penalty thresholds, (4) and (5) are close to $\frac{1}{\alpha}$ times their Bernoulli thresholds, (2) and (3), for $p = 1 - F(\alpha D)$. Recall that in the single-player model, α-truthfulness can be obtained by directly using the Bernoulli threshold with $p = 1 - F(\alpha D)$. In the multi-player model, however, we have to multiply the Bernoulli threshold with a factor of $\frac{1}{\alpha}$, which suggests it is more difficult to get every player to speak the truth under the cost-sharing setting. We note that both penalty rates (4) and (5) are upper bounds for the actual thresholds. This is because we treat any report greater than αD as αD. We conjecture that the exact thresholds are not far from thresholds (4) and (5).

6 Conclusion and Open Problems

We propose a penalty mechanism for eliciting truthful self-reports when only partial signals are revealed in a repeated game. A player faces trade-off between under-reporting today and paying a penalty in the future due to the uncertainty of partial signals. We find that the length of the game naturally reduces the minimum penalty rate that incentivizes truth-telling. Given any penalty rate, we give a characterization of the optimal strategies under both single- and multiple-player settings for any distribution. We identify a penalty rate that achieves complete truthfulness for Bernoulli distributions, which can be used in a reduction to obtain approximate truthfulness for arbitrary distributions.

A possible future direction is to extend our results to asymmetric multi-player settings where players do not have the same gross consumption or the same distribution for partial signals. For heterogeneous players, we may then consider, in addition to truthfulness, the fairness of the mechanism. It would be interesting to develop a definition of fairness for the cost sharing model and compute the fairness ratios accordingly. It is also worthwhile to derive other truthful and fair mechanisms that do not involve penalty.

References

1. Aswantara, I.K.A., Ko, K.S., Sung, D.K.: A centralized EV charging scheme based on user satisfaction fairness and cost. In: 2013 IEEE Innovative Smart Grid Technologies-Asia (ISGT Asia), pp. 1–4. IEEE (2013)
2. Baharlouei, Z., Hashemi, M., Narimani, H., Mohsenian-Rad, H.: Achieving optimality and fairness in autonomous demand response: benchmarks and billing mechanisms. IEEE Trans. Smart Grid **4**(2), 968–975 (2013)
3. Ball, I., Kattwinkel, D.: Probabilistic verification in mechanism design. In: Proceedings of the 2019 ACM Conference on Economics and Computation, pp. 389–390 (2019)
4. Burger, S., Schneider, I., Botterud, A., Pérez-Arriaga, I.: Fair, equitable, and efficient tariffs in the presence of distributed energy resources. Consumer, Prosumer, Prosumager: How Service Innovations will Disrupt the Utility Business Model, p. 155 (2019)
5. Caragiannis, I., Elkind, E., Szegedy, M., Yu, L.: Mechanism design: from partial to probabilistic verification. In: Proceedings of the 13th ACM Conference on Electronic Commerce, pp. 266–283 (2012)
6. Darghouth, N.R., Barbose, G., Wiser, R.: The impact of rate design and net metering on the bill savings from distributed PV for residential customers in California. Energy Policy **39**(9), 5243–5253 (2011)
7. Dufo-López, R., Bernal-Agustín, J.L.: A comparative assessment of net metering and net billing policies. Study cases for Spain. Energy **84**, 684–694 (2015)

8. Eid, C., Guillén, J.R., Marín, P.F., Hakvoort, R.: The economic effect of electricity net-metering with solar PV: consequences for network cost recovery, cross subsidies and policy objectives. Energy Policy **75**, 244–254 (2014)

9. Gautier, A., Jacqmin, J., Poudou, J.-C.: The prosumers and the grid. J. Regul. Econ. **53**(1), 100–126 (2018). https://doi.org/10.1007/s11149-018-9350-5

10. Glass, E., Glass, V.: Power to the prosumer: a transformative utility rate reform proposal that is fair and efficient. Electr. J. **34**(9), 107023 (2021)

11. Haghtalab, N., Immorlica, N., Lucier, B., Wang, J.Z.: Maximizing welfare with incentive-aware evaluation mechanisms. arXiv preprint arXiv:2011.01956 (2020)

12. Hardt, M., Megiddo, N., Papadimitriou, C., Wootters, M.: Strategic classification. In: Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science, pp. 111–122 (2016)

13. Heylen, E., Ovaere, M., Proost, S., Deconinck, G., Van Hertem, D.: Fairness and inequality in power system reliability: summarizing indices. Electr. Power Syst. Res. **168**, 313–323 (2019)

14. Heylen, E., Ovaere, M., Van Hertem, D., Deconinck, G.: Fairness of power system load-shedding plans. In: 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 1404–1409. IEEE (2018)

15. Hoarau, Q., Perez, Y.: Network tariff design with prosumers and electromobility: who wins, who loses? Energy Econ. **83**, 26–39 (2019)

16. Khodabakhsh, A., Horn, J., Nikolova, E., Pountourakis, E.: Prosumer pricing, incentives and fairness. In: Proceedings of the Tenth ACM International Conference on Future Energy Systems, pp. 116–120 (2019)

17. Koumparou, I., Christoforidis, G.C., Efthymiou, V., Papagiannis, G.K., Georghiou, G.E.: Configuring residential PV net-metering policies–a focus on the Mediterranean region. Renew. Energy **113**, 795–812 (2017)

18. Leavitt, L.: Solar batteries: how renewable battery backups work (2021). https://www.cnet.com/home/energy-and-utilities/solar-batteries-how-renewable-battery-backups-work/. Accessed 13 Aug 2022

19. Liang, A., Madsen, E.: Data and incentives. In: Proceedings of the 21st ACM Conference on Economics and Computation, pp. 41–42 (2020)

20. Moret, F., Pinson, P.: Energy collectives: a community and fairness based approach to future electricity markets. IEEE Trans. Power Syst. **34**(5), 3994–4004 (2018)

21. Nadel, S., Young, R.: Why is electricity use no longer growing? In: American Council for an Energy-Efficient Economy Washington (2014)

22. National Conference of State Legislators: State net metering policies (2017). https://www.ncsl.org/research/energy/net-metering-policy-overview-and-state-legislative-updates.aspx. Accessed 13 Aug 2022

23. Negash, A.I., Kirschen, D.S.: Combined optimal retail rate restructuring and value of solar tariff. In: 2015 IEEE Power & Energy Society General Meeting, pp. 1–5. IEEE (2015)

24. Schelly, C., Louie, E.P., Pearce, J.M.: Examining interconnection and net metering policy for distributed generation in the United States. Renew. Energy Focus **22**, 10–19 (2017)

25. Singh, S.P., Scheller-Wolf, A.: That's not fair: tariff structures for electric utilities with rooftop solar. Manuf. Serv. Oper. Manag. **24**(1), 40–58 (2022)

26. Solar Energy Industry Associations: Net metering (2017). https://www.seia.org/initiatives/net-metering. Accessed 01 Sept 2021

27. United States Energy Information Administration: Hourly electricity consumption varies throughout the day and across seasons (2020). https://www.eia.gov/todayinenergy/detail.php?id=42915. Accessed 03 Nov 2021

28. Vieira, D., Shayani, R.A., De Oliveira, M.A.G.: Net metering in Brazil: regulation, opportunities and challenges. IEEE Lat. Am. Trans. **14**(8), 3687–3694 (2016)
29. Wu, Y., Khodabakhsh, A., Li, B., Nikolova, E., Pountourakis, E.: Eliciting information with partial signals in repeated games. CoRR abs/2109.04343 (2021)
30. Zhang, H., Conitzer, V.: Incentive-aware PAC learning. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, pp. 5797–5804 (2021)