Detecting Fraud in Demand Response Programs

Carlos Barreto, *Student Member, IEEE,* and Alvaro A. Cárdenas, *Member, IEEE,*

***Abstract*—In this paper we formulate a new way to defraud the electricity system without the risks of being detected. Specifically, by attacking the signals sent by the demand response provider or electric utility the attacker can affect the behavior of a large sector of the population, and instruct them to behave in a manner beneficial for the attacker. Moreover, the attack scheme allows to define a large set of users who get benefits from the attack, making impossible the precise identification of the culprit. We analyze ways to detect attacks (i.e., detect that an attack is occurring, but not who is responsible for it), and propose some ideas for how to design the market in a way that attackers will not have an incentive to defraud the system.**

***Index Terms*—Electricity market, direct load control, dynamic pricing, security.**

I. Introduction

Most of the literature about fraud in the electric distribution system is centered around attacking smart meters to report less electricity consumption, or tapping directly into distribution lines bypassing meters (electricity theft). While these attacks are beneficial for a fraudster in the short-term, if the attack gets detected, the fraudster can be identified and be penalized severely.

In contrast, Demand Response (DR) programs can give attackers a new way to defraud the electricity system without the risks of being detected. By attacking the DR signals sent by the DR provider or electric utility (as shown in Fig. 1), the attacker can affect the behavior of a large sector of the population, and instruct them to behave in a manner beneficial for the attacker. For example the attacker can select a subset of consumers and instruct them to reduce electricity consumption (i.e., a set $\mathcal{V}$ of victims), reducing the cost of electricity for the population and therefore enabling another set of the population to consume larger amounts of electricity at reduced prices (i.e., a set $\mathcal{A}$ of consumers that benefit from the attack).

Notice that if a sabotaged smart meter is detected, then the attacker can be easily identified (the smart meter is attached to the property of the attacker); however, as described in the previous paragraph, defrauding the electricity system by attacking control signals from DR programs will add a layer of indirection that will make detection of attacks harder. In fact, if the set $\mathcal{A}$ of consumers who benefit from the attack (most of the elements of $\mathcal{A}$ can be honest and unaware that the attack is happening) is large enough, a forensic analyst trying to identify the attacker will see that a large set of users is benefiting, and all of them can claim plausible deniability, making impossible the precise identification of the culprit.

C. Barreto, and A. A. Cárdenas are with the Department of Computer Science, University of Texas at Dallas, Richardson, TX, USA. email: carlos.barretosuarez, alvaro.cardenas@utdallas.edu.

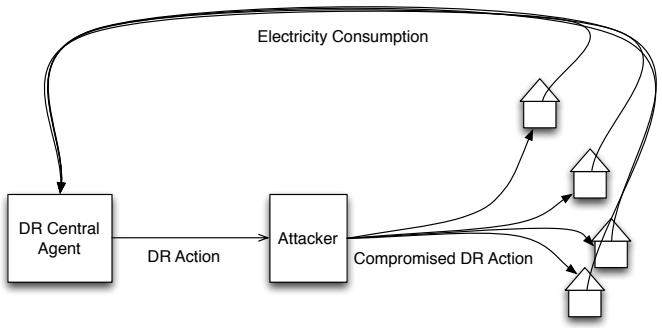


Fig. 1: Adversary Model: by compromising DR signals, the attacker can affect the behavior of a large sector of the population, and instruct them to behave in a manner beneficial for the attacker (e.g., force them to reduce electricity consumption so the attacker can get electricity at reduced rates).

In this paper we formulate this novel problem, analyze ways to detect attacks (i.e., detect that an attack is occurring, but not who is responsible for it), and propose some ideas for how to design the market in a way that attackers will not have an incentive to defraud the system (at the cost of some market efficiency when there are no attacks).

We model a general scenario in which there is a central planner who designs some mechanism to maximize the social welfare. This mechanism can be seen as a DR scheme that is designed to achieve efficiency. We focus on two previously proposed DR systems with two general information constraints, namely full information and asymmetric information of users preferences. These constraints are prevailing in many DR schemes, such as direct load control and dynamic pricing schemes.

We assume that the central planer (defender) knows i) the consumption of all users before an attack, and ii) the true consumption of all users (i.e., in Fig. 1 the smart meters reporting consumption back to the DR agent are not compromised).

We model an attacker whose objective is to maximize its own utility. We neglect limitations on the attacker's actions (such as cost or reach of the attack) and assume that the attacker has freedom to manipulate the demand profile of the population. The only restriction is that the attacker cannot falsify the consumption sent to the utility by her smart meter. That is, the attacker cannot modify its own bill. Therefore, the attacker might manipulate other components of the system. To model the practical scenario, we assume that the attacker is the only one that knows the parameters of the attack, and therefore, an anomaly detection algorithm will need to estimate the parameters of the attacks.

Section II introduces the notation and the model of the electricity system. Section III introduces the attack model and illustrates the effects of attacks on the electricity system. Section IV discusses how a central planner can detect an attack and identify the severity of the attack. Finally, Section V discusses how once we detect the population that is benefiting from the attack (even if they are not guilty of launching the attack) and the severity of the attack, we can design penalties that will make it unprofitable to continue attacking the system.

II. Background

We consider an electricity system with N users. We denote with q_i the average electricity consumption of the i^{th} user. The demand profile of the population is represented by the vector $\boldsymbol{q} = [q_1, q_2, \ldots, q_N]^\top \in \mathbb{R}^N_{\geq 0}$. The aggregated demand is denoted using the 1-norm: $\|\boldsymbol{q}\| = \sum_{j=1}^N q_j$. Without loss of generality, we assume that the electricity consumption of the i^{th} user satisfies $q_i \geq \underline{Q_i}$, where $\underline{Q_i} > 0$ represents the minimum consumption level. A *valuation function* $v_i(q_i)$ models the *valuation* that the i^{th} user gives to an electricity consumption of q_i units. Moreover, $\dot{v}_i()$ denotes the marginal valuation, defined as $\dot{v}_i(q) = \frac{\partial}{\partial q_i} v_i(q_i)|_{q_i=q}$. Let $p(\cdot) : \Re \to \Re$ be the price of electricity charged to consumers.

Following the market models in [1], [2], we assume that there is an independent system operator (ISO), which is in charge of clearing the market. Thus, we can express the profit function of each individual as their valuation of electricity minus their electricity bill, i.e.,

$$U_i(\boldsymbol{q}) = v_i(q_i) - q_i p\left(\|\boldsymbol{q}\|\right). \tag{1}$$

In this case we assume that the generation cost is quadratic (this is supported by [3]). Hence, the unitary price charged to costumers is a linear function, defined as $p(z) = \beta z + b$, where $\beta > 0$, $b \geq 0$ are parameters of the generation system.

One of the main reasons to use markets is to coordinate producers and consumers to achieve efficient outcomes. Particularly, the social optimal is the outcome that maximizes the profit of all users, which can be seen as the solution to the following optimization problem:

$$\begin{aligned} \underset{\boldsymbol{q}}{\text{maximize}} \quad & \sum\nolimits_{i=1}^N U_i(\boldsymbol{q}) \\ \text{subject to} \quad & q_i \geq \underline{Q_i}, i = \{1, \ldots, N\}. \end{aligned} \tag{2}$$

Here we make some assumptions on the problem characteristics in order to guarantee that the problem has a maximum and it is unique.

Assumption 1.

i. The valuation function $v_i^t(\cdot)$ is differentiable, concave, and non-decreasing.
ii. The price $p(\cdot)$ is differentiable, convex, and non-decreasing.

Assumption 2. The maximum of a concave function is inside the feasible set, i.e., the following inequality is satisfied for all i: $\frac{\partial}{\partial q_i} U_i([\underline{Q_1}, \ldots, \underline{Q_N}]^\top) > 0$.

Thus, the optimal outcome, denoted by $\boldsymbol{\mu}$, satisfies the following first order conditions (FOC):

$$\dot{v}_i(q_i) - p(||\boldsymbol{q}||) - \beta \|\boldsymbol{q}\| \big|_{\boldsymbol{q}=\boldsymbol{\mu}} = 0, \tag{3}$$

for every agent $i \in \{1, \ldots, N\}$.

III. Attack on the Electricity Market

In this case we model an attacker whose objective is to maximize its own utility [4]. We neglect limitations on the attacker's actions (such as cost or reach of the attack) and assume that the attacker has some freedom to manipulate the demand profile of the population. Here we consider only two restrictions: on one hand, the attacker cannot falsify the consumption sent to the utility by the smart meters. Therefore, the attacker cannot modify the bill of any user (but she might manipulate other components of the system). In second place, the attacker has as much information as the central planner. This means that, depending on the DR scheme, the attacker might have access to the consumption valuation of users.

Let us represent the attacker's objective with the following optimization problem (which can be implemented even with asymmetric information [4]):

$$\begin{aligned} \underset{\boldsymbol{q}}{\text{maximize}} \quad & \lambda \sum\nolimits_{h\in\mathcal{A}} U_h(\boldsymbol{q}) + \sum\nolimits_{h\in\mathcal{V}} U_h(\boldsymbol{q}) \\ \text{subject to} \quad & q_i \geq \underline{Q_i}, i = \{1, \ldots, N\}. \end{aligned} \tag{4}$$

The attack model uses two parameters, namely the severity of the attack $\lambda > 1$ and the proportion of attackers $0 < \gamma < 1$. On the one hand, λ allows us to adjust the impact of the attack on the population. For large λ, this optimization problem can lead to the maximum benefit for the users that belong to $\mathcal{A}$, because the utility of victims becomes irrelevant [4]. In second place, γ let us partition the population into two subsets $\mathcal{V} = \{1, \ldots, \lfloor(1-\gamma)N\rfloor\}$ and $\mathcal{A} = \{\lfloor(1-\gamma)N\rfloor + 1, \ldots, N\}$, whose members are either victims or take advantage of an attack, respectively. In this way, we consider multiple users who get benefits from the attack, because of coalitions or because attackers will share benefits with other users in an attempt to make the attack unattributable. Note that users who get benefits from the attack might or might not be aware of this.

The Lagrangian associated with the problem in Eq. (4) is

$$\begin{aligned} L(\boldsymbol{q}, \boldsymbol{\nu}) = \lambda \sum\nolimits_{h\in\mathcal{A}} U_h(\boldsymbol{q}) + \sum\nolimits_{h\in\mathcal{V}} U_h(\boldsymbol{q}) \\ + \sum\nolimits_{i=1}^N \nu_i \cdot (q_i - \underline{Q_i}). \end{aligned}$$

Thus, the demand profile under an attack, denoted by $\boldsymbol{x}$, must satisfy the following optimality conditions:

$$\lambda\left(\dot{v}_i(x_i) - p(||\boldsymbol{x}||) - \beta \|\boldsymbol{x}_{\mathcal{A}}\|\right) - \beta \|\boldsymbol{x}_{\mathcal{V}}\| + \nu_i = 0, \tag{5}$$

$$\dot{v}_j(x_j) - p(||\boldsymbol{x}||) - \beta \|\boldsymbol{x}_{\mathcal{V}}\| - \lambda\beta \|\boldsymbol{x}_{\mathcal{A}}\| + \nu_j = 0, \tag{6}$$

$$q_h - \underline{Q_h} \geq 0,$$

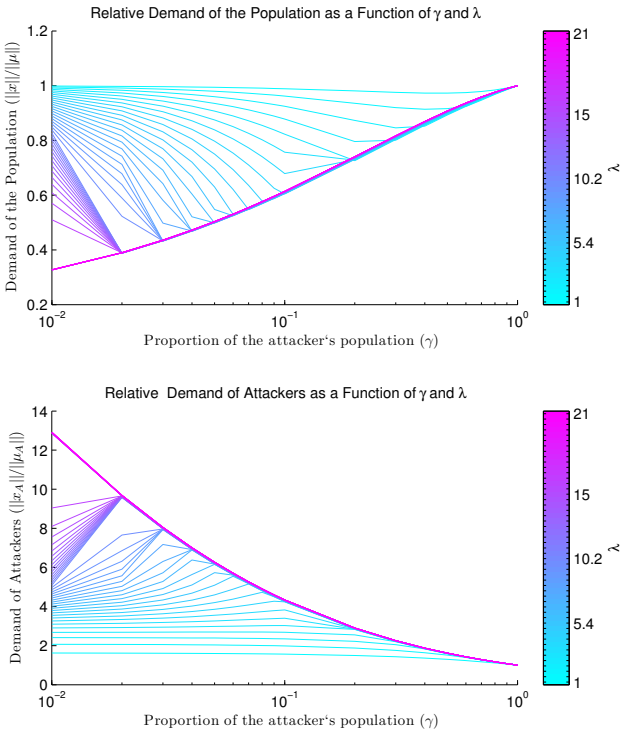


Fig. 2: Aggregated demand of attackers and victims as a function of γ. Attacker's consumption is higher if the benefits of the attack are shared with fewer users (e.g., γ is small). Attacker's benefits are accompanied by reduction in total demand.

$$\nu_h \geq 0,$$

$$(q_h - \underline{Q_h})\nu_h = 0,$$

for all $i \in \mathcal{A}$, $j \in \mathcal{V}$, and $h \in \{1, \ldots, N\}$. Let us denote by $\boldsymbol{x}_{\mathcal{A}}$ and $\boldsymbol{x}_{\mathcal{V}}$ the vectors with the consumption of attackers and victims, respectively. Thus, $||\boldsymbol{x}_{\mathcal{A}}|| = \sum_{i\in\mathcal{A}} x_i$ and $||\boldsymbol{x}_{\mathcal{V}}|| = \sum_{j\in\mathcal{V}} x_j$.

A. Illustration of Attacks

We illustrate the effect of attacks using some typical functions previously used in the literature that satisfy Assumptions 1, and 2 [2], [5] (a detailed implementation of the simulations can be found in [6]):

$$v_i^t(q_i^t) = \alpha_i^t \log(1 + q_i^t),\ \alpha_i^t > 0,$$
$$p(\|\boldsymbol{q}\|) = \beta\|\boldsymbol{q}\| + b,\ \beta > 0.$$

Fig. 2 shows the total electricity demand of both the entire population and attackers ($||\boldsymbol{x}||$ and $||\boldsymbol{x}_{\mathcal{A}}||$ respectively) as a function of the proportion of attackers γ, for different values of λ (severity of the attack). In this case the electricity demand is normalized with respect to the ideal demand ($||\boldsymbol{\mu}||$ and $||\boldsymbol{\mu}_{\mathcal{A}}||$ respectively). Note that the demand of attackers can have high values if the severity of the attack is high and the proportion of attackers is small. Particularly, in this example an attacker might consume more than 12 times using the attack of highest severity ($\lambda = 21$). On the other hand, as the severity of the attack increases, the total demand decreases even more than half of the ideal case. Simulations are made with $N = 100$, $\beta = b = 1$, $\alpha_i = 96.63$, $Q_i = 0.07$.

In summary, some consequences of the attack are: i) reduction of demand by victims; ii) increased demand by attackers; iii) reduction in the total demand of the population. This properties are formally proved in the following proposition:

Proposition 1. *Let $\boldsymbol{\mu}$ be the ideal equilibrium and $\boldsymbol{x}$ be the equilibrium with an attack associated with the optimization problems in Eqs. (2) and (4), respectively. If there is an attack with $\lambda > 1$, then the consumption of attackers (or victims) increases (or decreases) and the total demand decreases, with respect to the ideal case. That is,*

$$x_i > \mu_i,$$
$$x_j < \mu_j,$$
$$\|\boldsymbol{x}\| < \|\boldsymbol{\mu}\|,$$

for every attacker $i \in \mathcal{A}$ and victim $j \in \mathcal{V}$.

Sketch proof. We can evaluate the derivative of the attacker's objective function (see Eq. (4)) in the ideal outcome $\boldsymbol{\mu}$ to obtain

$$\lambda\left(\dot{v}_i(\mu_i) - p(||\boldsymbol{\mu}||) - \beta\|\boldsymbol{\mu}\|\right) + (\lambda - 1)\beta\|\boldsymbol{\mu}_{\mathcal{V}}\|,$$

$$\dot{v}_j(\mu_j) - p(||\boldsymbol{\mu}||) - \beta\|\boldsymbol{\mu}\| + (1 - \lambda)\beta\|\boldsymbol{\mu}_{\mathcal{A}}\|.$$

Note that the left hand side of the previous equations is precisely the FOC of the original optimization problem (see Eq. (3)). Thus, we can conclude that the derivative with respect to q_i is

$$(\lambda - 1)\beta\|\boldsymbol{\mu}_{\mathcal{V}}\| > 0,$$

and the derivative with respect to q_j is

$$(1 - \lambda)\beta\|\boldsymbol{\mu}_{\mathcal{A}}\| < 0.$$

Hence, we know that $x_i > \mu_i$ and $x_j > \mu_j$.

Now we are ready to prove that if $\lambda > 1$, then the total demand of the population is reduced. Recall that the valuation functions are concave, and consequently, the marginal valuations are convex decreasing functions. Therefore, we know that if $x_i > \mu_i$, then $\dot{v}_i(x_i) < \dot{v}_i(\mu_i)$. We can use Eq. (5) and (3) to obtain the following equivalent expression:

$$p(||\boldsymbol{x}||) + \beta||\boldsymbol{x}|| + \frac{1-\lambda}{\lambda}\beta\|\boldsymbol{x}_{\mathcal{V}}\| - \nu_i < p(||\boldsymbol{\mu}||) + \beta\|\boldsymbol{\mu}\|. \quad (7)$$

In this case $x_i > \underline{Q_i}$, and consequently, $\nu_i = 0$. Since $\lambda > 1$, from Eq. (7) it is clear that $\|\boldsymbol{x}\| < \|\boldsymbol{\mu}\|$. □

The intuition behind this result is that an attacker can increase her profit only if a significant reduction in the total demand is realized. Since the cost function is convex, an attacker can afford an increase in demand only if there is a reduction in the total demand of the population.

Note that the previous conclusions are satisfied regardless of the valuation function of each agent. Hence, the central planner can use this fact to determine the proportion of the attackers in the population γ (note that this information was only known by the attacker). Particularly, the partition of the population can be determined as follows:

$$\mathcal{A} = \{i | i \in \{1, \dots, N\},\, x_i > \mu_i\},$$
$$\mathcal{V} = \{j | j \in \{1, \dots, N\},\, x_j \leq \mu_j\}.$$

Thus, the proportion of attackers is $\gamma = {}^{|\mathcal{A}|}/_{N}$. This classification of users is used to design the detection mechanism in the following section.

IV. Detection of an Attack

An electricity utility aware of possible attacks on the system might raise alarms when the total demand falls below some threshold. Specifically, the threshold might be determined based on historic consumption data. For instance, alarms might be generated if the total demand falls more than $\epsilon = 5\%$ of the normal demand. In that case, an attacker aware of the detection mechanism might choose λ and γ that satisfy $\|\boldsymbol{x}\| \leq (1-\epsilon)\|\boldsymbol{\mu}\|$. From the example presented in Section III-A, we observe that with $\gamma = 0.7$ and $\lambda > 1.1$ the total demand is within the desired threshold. Moreover, the attacker increases her profit about 1.24 times.

This detection scheme has some drawbacks. Specifically, some attacks cannot be detected, and moreover, any demand beyond the threshold is considered an attack. However, we might design better detection mechanisms by using the characteristics of the attack developed in the last section. First, we analyze how to detect attacks in DR schemes with full information. Later we consider the detection problem with asymmetric information. In these cases we assume that $\boldsymbol{\mu}$ can be extracted from normal demand data and that β and b are known parameters.

A. DR with Full Information

From Eq. (6) we can extract the following relationship:

$$\lambda = \frac{\beta||\boldsymbol{x}_{\mathcal{V}}||}{\dot{v}_i(x_i) - 2\beta||\boldsymbol{x}_{\mathcal{A}}|| - \beta||\boldsymbol{x}_{\mathcal{V}}|| - b}, \tag{8}$$

If the utility company knows the valuation function of users, then it can use the previous equation to determine the value of λ. Note that $\lambda = 1$ indicates normal behavior, while $\lambda > 1$ suggests an attack. A drawback of this expression is that of $||\boldsymbol{x}_{\mathcal{V}}||$ must be different from zero. Otherwise, all the estimations of λ might be equal to zero.

Now, let us introduce an alternative method to detect attacks, which doesn't require $||\boldsymbol{x}_{\mathcal{V}}|| \neq 0$. Let us assume that there is a reduction in the demand of some users, which are classified as members of the set $\mathcal{V}$. Our objective is to find out if the reduction of demand was caused by an attacker. In this case we denote by $\boldsymbol{\zeta}$ and $\boldsymbol{\xi}$ the electricity demand associated to a normal and a fraudster behavior, respectively. We know that a normal and fraudster behaviors are determined by Eq. (3) and (5), which can be rewritten in this case as:

$$\dot{v}_i(\zeta_i) - 2\beta\left\|\boldsymbol{\zeta}_{\mathcal{A}}\right\| - b = 2\beta\left\|\boldsymbol{\zeta}_{\mathcal{V}}\right\|, \tag{9}$$

$$\dot{v}_i(\xi_i) - 2\beta\left\|\boldsymbol{\xi}_{\mathcal{A}}\right\| + \left(1 - \frac{1}{\lambda}\right)\beta\left\|\boldsymbol{\xi}_{\mathcal{V}}\right\| - b > 2\beta\left\|\boldsymbol{\xi}_{\mathcal{V}}\right\|.$$

Since we are interested in observing the reaction of a normal user and an attacker to a given demand of victims Q, we define $\|\boldsymbol{\zeta}_{\mathcal{V}}\| = \|\boldsymbol{\xi}_{\mathcal{V}}\| = \underline{Q}$. Moreover, for large $\lambda > 1$ we have

$$\dot{v}_i(\zeta_i) - 2\beta\left\|\boldsymbol{\zeta}_{\mathcal{A}}\right\| = \dot{v}_i(\xi_i) - 2\beta\left\|\boldsymbol{\xi}_{\mathcal{A}}\right\|.$$

Note that $\dot{v}_i(q_i) - 2\beta\left\|\boldsymbol{q}_{\mathcal{A}}\right\|$ is a decreasing function with respect to $\boldsymbol{q}$. Hence, we know that $\boldsymbol{\zeta} < \boldsymbol{\xi}$, that is, the demand of an attacker is always higher than the demand made by a normal user.

It is interesting that Eq. (8) doesn't use the information from normal behavior (e.g., $\boldsymbol{\mu}$). Hence, it is possible to distinguish attacks from failures that cause changes in demand. For example, if replace the normal demand evaluated in $\boldsymbol{\zeta}$ (see Eq. (9)) into Eq. (8) we obtain

$$\lambda = \frac{\beta\underline{Q}}{\beta\underline{Q}} = 1,$$

which indicates a normal behavior.

B. DR with Asymmetric Information

In this case, the utility company ignores the valuation functions of each user, e.g., the system might be decentralized. Without full information it is hard to know if the demand follows the pattern of an attack, because the results from the previous section relied on the knowledge of valuation functions.

The utility company can attempt to estimate the marginal valuation of users to determine if a particular demand profile matches the properties of an attack. Particularly, it is possible to find some boundaries on the values of $||\boldsymbol{x}_{\mathcal{A}}||$ and $||\boldsymbol{x}_{\mathcal{V}}||$ that indicate the presence of an attack. However, these boundaries rely on an upper bound on λ, which unfortunately has no boundary. First, we show that an attacker always get more benefit by implementing $\lambda \to \infty$. Latter we introduce the boundaries on the demand.

Proposition 2. *An attacker has more profit by setting $\lambda \to \infty$.*

Proof. The following is the optimization problem that represents the goal of the attackers:

$$\begin{aligned} \underset{q_i, \boldsymbol{q}_{-i}}{\text{maximize}} \quad & \sum\nolimits_{i \in \mathcal{A}} U_i(\left\|\boldsymbol{q}_{\mathcal{A}}\right\|, \left\|\boldsymbol{q}_{\mathcal{V}}\right\|) \\ \text{subject to} \quad & q_i \geq \underline{Q}_i, i = \{1, \dots, N\}. \end{aligned}$$

One of the optimality conditions of this problem is

$$\dot{v}_i(x_i) - \beta||\boldsymbol{x}|| - \beta\left\|\boldsymbol{x}_{\mathcal{A}}\right\| - b = 0.$$

Note that this condition is satisfied through the optimization problem in Eq. (4) if $\lambda \to \infty$. Hence, it is always better for the attacker to choose a large λ. □

This result is true regardless of the value of $\underline{Q_i}$ and γ. However, γ establishes a limit in the maximum profit that can be achieved by the attacker. Intuitively, demand increments are profitable as long as the total demand decreases. However, γ set a limit in the users that can reduce demand, and consequently, in the total demand that can be reduced. For this reason the detection scheme at the beginning of the section can mitigate the impact of attacks, even though it has drawbacks.

Now, let us introduce some boundaries on the demand of victims and attackers.

Proposition 3. *$\Omega(||\boldsymbol{x}_{\mathcal{V}}||, \lambda)$ and $\Lambda(||\boldsymbol{x}_{\mathcal{A}}||, \lambda)$ represent the lower and upper bound of $||\boldsymbol{x}_{\mathcal{A}}||$ and $||\boldsymbol{x}_{\mathcal{V}}||$, respectively. That is, $||\boldsymbol{x}_{\mathcal{A}}|| \geq \Omega(||\boldsymbol{x}_{\mathcal{V}}||, \lambda)$ and $||\boldsymbol{x}_{\mathcal{V}}|| \leq \Lambda(||\boldsymbol{x}_{\mathcal{A}}||, \lambda)$, where*

$$\Omega(||\boldsymbol{x}_{\mathcal{V}}||, \lambda) = \frac{2}{\beta(1+\lambda)} \left(||\boldsymbol{\mu}|| - \frac{||\boldsymbol{\mu}_{\mathcal{V}}||}{N_V} - ||\boldsymbol{x}_{\mathcal{V}}|| \frac{N_V - 1}{N_V} \right),$$

and

$$\Lambda(||\boldsymbol{x}_{\mathcal{A}}||, \lambda) = \frac{2\lambda}{(1+\lambda)} \left(||\boldsymbol{\mu}|| - \frac{||\boldsymbol{\mu}_{\mathcal{A}}||}{N_S} - ||\boldsymbol{x}_{\mathcal{A}}|| \frac{N_S - 1}{N_S} \right).$$

Proof. First let us introduce a more general expression that can be obtained by summing Eqs. (5) and (6) over all the elements of each population. Thus, we get

$$||\boldsymbol{x}_{\mathcal{A}}|| = \frac{1}{\beta(1+\lambda)} \left(\frac{1}{N_V} \sum_{j \in \mathcal{V}} \dot{v}_j(x_j) - 2\beta ||\boldsymbol{x}_{\mathcal{V}}|| - b \right), \tag{10}$$

and

$$||\boldsymbol{x}_{\mathcal{V}}|| = \frac{\lambda}{\beta(1+\lambda)} \left(\frac{1}{N_S} \sum_{i \in \mathcal{S}} \dot{v}_i(x_i) - 2\beta ||\boldsymbol{x}_{\mathcal{A}}|| - b \right). \tag{11}$$

From the FOC of the original system (Eq. (2)) we know that

$$\dot{v}_i(\mu_i) = 2\beta ||\boldsymbol{\mu}|| + b, \tag{12}$$

for all $i \in \mathcal{N}$. We know that the valuation of each user $v_i(\cdot)$ is a concave function. Hence, we know that the marginal valuation $\dot{v}_i(\cdot)$ is convex decreasing and non-negative. Thus, from Proposition 1 follows

$$\dot{v}_i(x_i) \leq \dot{v}_i(\mu_i),$$
$$\dot{v}_j(x_j) \geq \dot{v}_j(\mu_j).$$

The previous equations can be used along Eq. (12) to extract the following inequalities:

$$\dot{v}_i(x_i) \leq \dot{v}_i(\mu_i) \leq 2\beta(||\boldsymbol{\mu}|| - \mu_i + x_i) + b, \tag{13}$$

$$\dot{v}_j(x_j) \geq \dot{v}_j(\mu_j) \geq 2\beta(||\boldsymbol{\mu}|| - \mu_j + x_j) + b, \tag{14}$$

Eqs. (13) and (14) can be replaced in Eqs. (10) and (11), respectively, to obtain:

$$||\boldsymbol{x}_{\mathcal{A}}|| \geq \frac{2}{\beta(1+\lambda)} \left(||\boldsymbol{\mu}|| - \frac{||\boldsymbol{\mu}_{\mathcal{V}}||}{N_V} - ||\boldsymbol{x}_{\mathcal{V}}|| \frac{N_V - 1}{N_V} \right) = \Omega(||\boldsymbol{x}_{\mathcal{V}}||, \lambda) \tag{15}$$

and

$$||\boldsymbol{x}_{\mathcal{V}}|| \leq \frac{2\lambda}{(1+\lambda)} \left(||\boldsymbol{\mu}|| - \frac{||\boldsymbol{\mu}_{\mathcal{A}}||}{N_S} - ||\boldsymbol{x}_{\mathcal{A}}|| \frac{N_S - 1}{N_S} \right) = \Lambda(||\boldsymbol{x}_{\mathcal{A}}||, \lambda), \tag{16}$$

where $||\boldsymbol{\mu}_{\mathcal{A}}||$ and $||\boldsymbol{\mu}_{\mathcal{V}}||$ represent the normal consumption of attackers and victims, respectively. On the other hand, $\Omega(||\boldsymbol{x}_{\mathcal{V}}||)$ and $\Lambda(||\boldsymbol{x}_{\mathcal{A}}||)$ represent the lower and upper bound of $||\boldsymbol{x}_{\mathcal{A}}||$ and $||\boldsymbol{x}_{\mathcal{V}}||$, respectively. □

The previous inequalities can be used to determine if there is an attack. The following theorem states that the demand in an attack must satisfy the boundaries introduced before.

Theorem 1. *The demand of an attack must satisfy*

$$||\boldsymbol{x}_{\mathcal{A}}|| \geq \Omega(||\boldsymbol{x}_{\mathcal{V}}||, \lambda)$$

and

$$||\boldsymbol{x}_{\mathcal{V}}|| \leq \Lambda(||\boldsymbol{x}_{\mathcal{A}}||, \lambda)$$

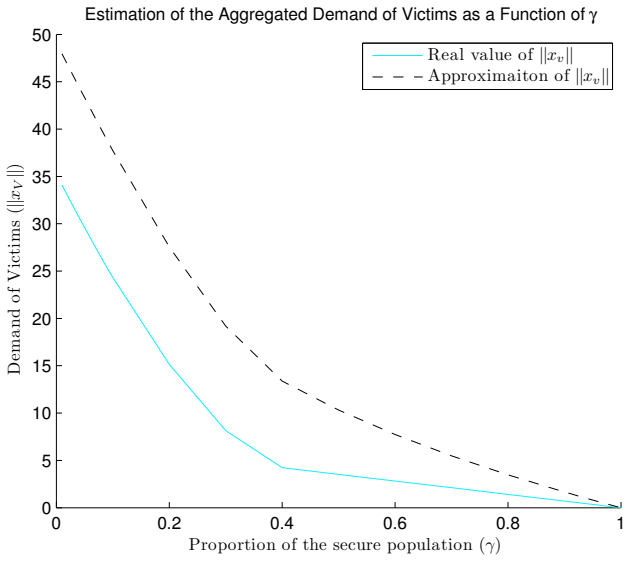
Proof. Note that for some λ and γ the demand profile with an attack is unique. Furthermore, the demand of attackers and victims is intimately related, making it possible to detect an attack if the demand matches the relations in Eqs. (15) and (16). An important characteristic is that the estimation of $||\boldsymbol{x}_{\mathcal{A}}||$ is found based $||\boldsymbol{x}_{\mathcal{V}}||$, i.e., the demand of attackers is found based on the demand of victims. Hence, the demand of victims can be used to estimate the demand of attackers that might be associated to it. Thus, the demand matches an attack as long as $||\boldsymbol{x}_{\mathcal{A}}|| > \Omega(||\boldsymbol{x}_{\mathcal{V}}||, \lambda)$. Similarly, the converse is also true, i.e., $||\boldsymbol{x}_{\mathcal{V}}|| < \Lambda(||\boldsymbol{x}_{\mathcal{A}}||, \lambda)$. □

A drawback is that we need an estimation of λ to obtain the boundaries. Therefore, the estimation of λ must overestimate the real parameter to guarantee that the estimation is correct.

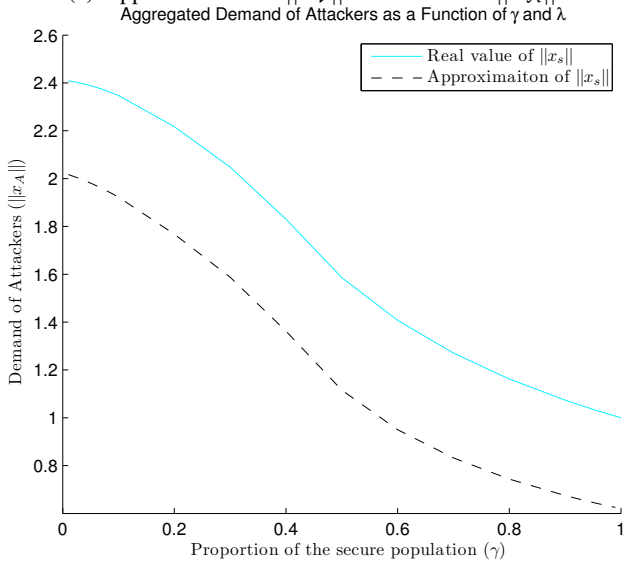
Example 1. Figs. 3 and 4 shows an example of the estimations obtained for an attack with $\lambda = 1.9$ and an estimation $\tilde{\lambda} = 2$. Note that the boundaries $\Omega(||\boldsymbol{x}_{\mathcal{V}}||, \tilde{\lambda})$ and $\Lambda(||\boldsymbol{x}_{\mathcal{A}}||, \tilde{\lambda})$ are linear with respect to $||\boldsymbol{x}_{\mathcal{V}}||$ and $||\boldsymbol{x}_{\mathcal{A}}||$, respectively. This property can be observed in Fig. 4, which also shows a parametric plot of $||\boldsymbol{x}_{\mathcal{A}}||$ and $||\boldsymbol{x}_{\mathcal{V}}||$. On the other hand, a case with no attack ($\lambda = 1$) is shown in a dotted line that connects the extremes of the boundary curves.

V. Design of Penalties

In this section we analyze the design of penalties imposed to agents once an attack is detected. The penalties are designed to make unprofitable the attacks–and might be defined in the contract between users and the company. In general, even though attacks are unattributable, the utility company can impose penalties on all users that benefit from the attack. While this might be unfair with agents who involuntarily get benefit from the attack, this action might prevent rational agents from launching attacks in the first place. Below we analyze two alternatives to design penalties for the cases with full information and asymmetric information.



(a) Approximation of $||\boldsymbol{x}_{\mathcal{V}}||$ in function of $||\boldsymbol{x}_{\mathcal{A}}||$.



(b) Approximation of $||\boldsymbol{x}_{\mathcal{A}}||$ in function of $||\boldsymbol{x}_{\mathcal{V}}||$.

Fig. 3: Demand estimations with $\lambda = 1.9$ and $\tilde{\lambda} = 2$. With an accurate estimation of λ it is possible to bound the demand with attacks.

A. Penalties with Full Information

Intuitively, the penalties should be equal to the losses caused by the attack, i.e., the attackers should be responsible for the looses caused to the population (this is similar to the Clark pivot mechanism [7]). Looses are defined as

$$\sum\nolimits_{j\in\mathcal{V}} U_j(\boldsymbol{\mu}) - U_j(\boldsymbol{x}),$$

and can be computed if the DR has full information about users' preferences.

This scheme is particularly desirable for the utility company, because allows it to save expenses for repairing the damage caused to victims. Thus, the utility company (and victims) might not have losses due to attacks. A drawback of this approach is that there might not be enough incentives for investing in security, since 1) attacks can be detected; and 2) losses can be covered by users who benefit from the attack.

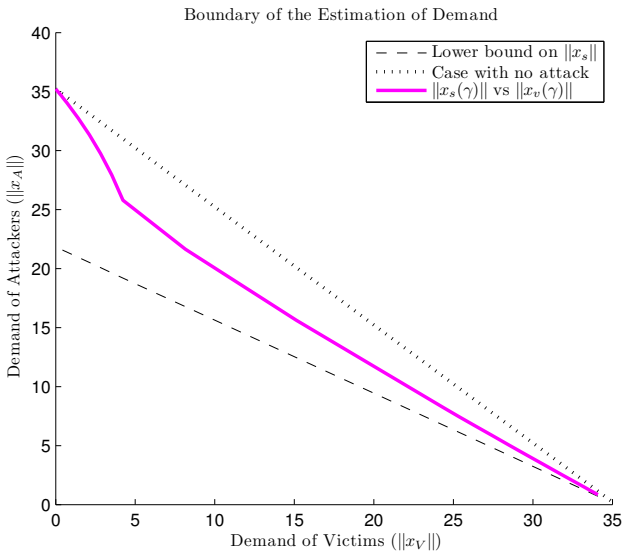


Fig. 4: Parametric plot of $||\boldsymbol{x}_{\mathcal{A}}||$ and $||\boldsymbol{x}_{\mathcal{V}}||$ in function of γ and the estimated boundaries with $\lambda = 1.9$ and $\tilde{\lambda} = 2$. With an accurate estimation of λ it is possible to detect attacks.

B. Penalties with Asymmetric Information

There are some difficulties to implement previous approach in DR schemes with asymmetric information. Specifically, in cases of asymmetric information it is necessary to estimate the profit losses. However, the estimation might be lower than its real value. To see the reason, let us calculate the profit reduction on the j^{th} victim as the utility change between the optimal outcome $\boldsymbol{\mu}$ and the outcome with an attack $\boldsymbol{x}$. The profit reduction can be approximated using the concavity property of the profit function in Eq. (1) as follows:

$$\begin{aligned} U_j(\boldsymbol{\mu}) - U_j(\boldsymbol{x}) &\leq \nabla U_j(\boldsymbol{x})(\boldsymbol{\mu} - \boldsymbol{x}) \\ &= \sum\nolimits_{h=1}^{N} \frac{\partial}{\partial q_h} U_j(\boldsymbol{q})\Big|_{\boldsymbol{q}=\boldsymbol{x}} (\mu_h - x_h), \end{aligned}$$

where j represents a victim. The marginal utility with respect to q_i is equal to

$$\frac{\partial}{\partial q_i} U_i(\boldsymbol{q}) = \dot{v}_i(q_i) - p(||\boldsymbol{q}||) - \beta q_i, \tag{17}$$

$$\frac{\partial}{\partial q_j} U_i(\boldsymbol{q}) = -\beta q_i. \tag{18}$$

Recall that the marginal valuation $\dot{v}_j(x_j)$ is unknown. Note that it is hard to find an accurate upper estimation, because the optimality condition in Eq. (6) depends on an unknown term ν_j. Hence, we are forced to underestimate the marginal valuation, using for instance the approximation given by Eq.

(14). This leads to the following inequality:

$$U_j(\boldsymbol{\mu}) - U_j(\boldsymbol{x}) \leq \nabla U_j(\boldsymbol{x})(\boldsymbol{\mu} - \boldsymbol{x}) \geq \sum\nolimits_{h\in\mathcal{N}} \frac{\partial}{\partial q_h}\tilde{U}_j(\boldsymbol{q})\Big|_{\boldsymbol{q}=\boldsymbol{x}}(\mu_h - x_h), \quad (19)$$

where $\tilde{U}_j(\boldsymbol{q})$ denotes the approximation using the marginal valuation. Note that the right part of Eq. (19) is the estimation of profit loss, but there is no guarantee that this estimation is greater than the real value. Hence, the loss of profit perceived by the victims cannot be used to design penalties, because it is possible to underestimate the losses.

Another alternative is to design penalties based on the profit earned by each attacker. In this case, the increase of utility can be estimated using

$$U_i(\boldsymbol{x}) - U_i(\boldsymbol{\mu}) \leq \nabla U_i(\boldsymbol{x})(\boldsymbol{x} - \boldsymbol{\mu}) = \sum\nolimits_{h\in\mathcal{N}} \frac{\partial}{\partial q_h}U_i(\boldsymbol{q})\Big|_{\boldsymbol{q}=\boldsymbol{x}}(x_h - \mu_h). \quad (20)$$

Using Eq. (17) and (18) we can rewrite the estimation as

$$\nabla U_i(\boldsymbol{x})(\boldsymbol{x} - \boldsymbol{\mu}) = (\dot{v}_i(\mu_i) - \beta||\boldsymbol{\mu}|| - b)(x_i - \mu_i) - \sum\nolimits_{h\neq i}\mu_h(x_h - \mu_h).$$

Note that we know the marginal valuation of the i^{th} user at the optimal outcome (see Eq. (12)). Thus, the estimation can be expressed as:

$$\nabla U_i(\boldsymbol{x})(\boldsymbol{x} - \boldsymbol{\mu}) = \beta||\boldsymbol{\mu}||x_i - \beta||\boldsymbol{x}||\mu_i + \mu_i(x_i - \mu_i). \quad (21)$$

Thus, from Eq. (20) we know that Eq. (21) gives an upper bound on the profit increase of an user $i \in \mathcal{A}$. Hence, if an attacker is charged according to Eq. (21), her profit is lower than the profit with no attack, i.e.,

$$U_i(\boldsymbol{x}) - \nabla U_i(\boldsymbol{x})(\boldsymbol{x} - \boldsymbol{\mu}) \leq U_i(\boldsymbol{\mu}).$$

This implies that an attacker might obtain lower profits by launching an attack (only if the attack is detected). Fig. 5 shows an example of the effect of penalties on the profit of attackers.

Note that if $\lambda = 1$ and $\|\boldsymbol{x}\| = \|\boldsymbol{\mu}\|$, then the penalties are zero. However, the penalties might fail when the there is a deviation from the expected behavior. Recall from Section IV-A that an event might change the normal electricity demand from $\|\boldsymbol{\mu}\|$ to $\|\boldsymbol{\zeta}\|$, where $\|\boldsymbol{\mu}_{\mathcal{V}}\| > \|\boldsymbol{\zeta}_{\mathcal{V}}\| = \underline{Q}$. Since the reaction to any user is to increase her demand (even if she is honest), we have $\|\boldsymbol{\mu}_{\mathcal{A}}\| < \|\boldsymbol{\zeta}_{\mathcal{A}}\|$. In this case, the users that belong to $\mathcal{A}$ might be charged with penalties, even if $\|\boldsymbol{\zeta}_{\mathcal{A}}\|$ satisfies the normal behavior stated in Eq. (3).

Note that with asymmetric information there is no guarantee that the penalties are enough to cover the total losses on the system. Hence, the utility company might have incentives to invest in security.

VI. Conclusions

We propose and analyze general mechanisms to detect and penalize attacks on DR schemes with full information and asymmetric information. First, we showed how in a DR system with full information it is straightforward to detect attacks and to design penalties that cover losses caused to others. Furthermore, an attacker with access to private information of users might have more flexibility to design attacks.

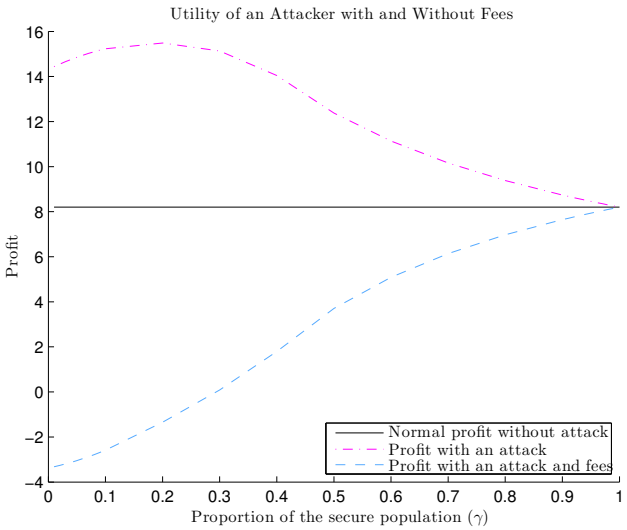


Fig. 5: The design of penalties on the attacker's profit make it unprofitable to launch attacks, even with asymmetric information.

Second, we find that the case of DR with asymmetric information is more challenging, but also has some desired properties. On the one hand, even though attacks are harder to detect, asymmetric information limits the actions of an attacker. However, there is no guarantee that the penalties distributed to the population benefiting from the attack are enough to cover the total losses on the system.

A limitation of this work is that the analysis relies on capturing the normal behavior of consumers (i.e., how they react to DR signals). It is interesting to investigate more general models that include uncertainties in that information. Another interesting direction is to identify the properties of the system that can affect the impact of the attacks, such as the size of the population, among others.

References

[1] C. Barreto, E. Mojica-Nava, and N. Quijano, "Design of mechanisms for demand response programs," in *2013 IEEE 52nd Annual Conference on Decision and Control (CDC)*, Dec. 2013.

[2] M. Roozbehani, M. Rinehart, M. A. Dahleh, S. K. Mitter, D. Obradovic, and H. Mangesius, "Analysis of competitive electricity markets under a new model of real-time retail pricing," in *2011 8th International Conference on the European Energy Market (EEM)*, may 2011, pp. 250–255.

[3] A. J. Wood and B. F. Wollenberg, *Power generation, operation, and control*. John Wiley & Sons, 2012.

[4] C. Barreto, A. A. Cárdenas, N. Quijano, and E. Mojica-Nava, "Cps: Market analysis of attacks against demand response in the smart grid," in *Proceedings of the 30th Annual Computer Security Applications Conference*, ser. ACSAC '14. New York, NY, USA: ACM, 2014, pp. 136–145. [Online]. Available: http://doi.acm.org/10.1145/2664243.2664284

[5] A. Mohsenian-Rad, V. Wong, J. Jatskevich, R. Schober, and A. Leon-Garcia, "Autonomous demand-side management based on game-theoretic energy consumption scheduling for the future smart grid," *IEEE Transactions on Smart Grid*, vol. 1, no. 3, pp. 320–331, dec 2010.

[6] C. Barreto, E. Mojica-Nava, and N. Quijano, "Incentives-based mechanism for efficient demand response programs," *arXiv preprint arXiv:1408.5366*, 2014.

[7] N. Nisan, T. Roughgarden, É. Tardos, and V. V. Vazirani, *Algorithmic Game Theory*. 32 Avenue of the Americas, New York, NY 10013-2473, USA: Cambridge University Press, 2007.