Equilibria in data injection attacks

Iñaki Esnaola, Samir M. Perlaza, H. Vincent Poor

Equilibria in Data Injection Attacks

Iñaki Esnaola, Samir M. Perlaza, and H. Vincent Poor

***Abstract*—Data injection attacks are studied in a game theoretic setting. Assuming that the network operator acquires the state variables via generalized least squares (GLS) estimation, different attack performance metrics are proposed. The scenarios defined by the performance metrics are then analyzed. In particular, closed form expressions for best response functions and Nash equilibria (NEs) are given. First the case in which the attack vector can be constructed without energy constraints is studied. It is shown that for unconstrained attacks infinitely many optimal attack vectors exist and that the construction requires knowledge of the state variables in the grid. Alternatively, when energy constraints are included, the attack vector construction does not depend on the state variables. As a consequence, the optimal energy constrained attack strategy follows a correlated multivariate Gaussian distribution. It is shown that for unconstrained attacks infinitely many NEs exist and that in the constrained case at least one NE exists.**

I. Introduction

The smart grid paradigm is founded on the integration of existing power grids with sophisticated sensing and communication infrastructures. While the benefits provided by this setting are crucial for the future development of power grids, it also paves the way for cyber-security threats [1]. Recently, data injection attacks [2], [3], [4] have been proposed as a feasible risk to electricity grids. The fundamental assumption to perform data injection attacks is that a malicious attacker has access to metering units and thus, is able to tamper with network measurements to distort the state estimate obtained by the network operator. In [2] and [3] unobservable attacks are studied and construction procedures for attackers with access to a limited number of meters are presented.

The analysis in [2] relies on algebraic tools and assumes that the detector ignores the stochastic nature of the state variables. However, with growing data mining and analysis capabilities provided by modern computing, it is reasonable to assume that network operators can learn the statistical structure of the system and use attack detection strategies that incorporate the underlying stochastic process governing the network. In [5] it is assumed that training data is available to the network operator and the attack detection is posed as a statistical learning problem. Detection techniques in which a Bayesian model of the random state variables is assumed are studied in [3], [6] and [7]. In contrast, a scenario in which multiple attackers are present and/or limited communication is available among different instantiations of the same attacker, raises the notion of distributed attacks. Distributed attack and detection strategies when only local information is available are investigated in [8].

This paper considers the case in which second order statistics about the additive disturbance to the measurement vector are available to a centralized network operator. That being the case, it is assumed that the network operator acquires the state variables through a GLS estimator which incorporates the information of second order statistics into the weighting matrix [9]. Similarly, the attacker is assumed to have access to all the data in the network, which poses a worst case scenario for analyzing the vulnerability of the network operator to this type of threat.

There is an intrinsic conflict between the objective of the operator and the goals of the attacker. Moreover, the nature of the setting dictates that both the attacker and operator act without coordinating their actions. For that reason, in this paper, the data injection problem is cast in a game theoretic framework. Interestingly, studying the performance of the attacks boils down to analyzing the equilibria of the game formed by considering the attacker and the operator as competing entities, i.e., as players. In that setting, the trade-off between the damage to the network, e.g., the excess distortion term, and the ability to remain hidden from the network operator, e.g, to keep the probability of attack detection under a given threshold is studied. With practical limitations in mind, the impact of energy constraints on the attack construction is discussed as well. Ultimately, the main results of this paper are inscribed in the context of best response and NE analysis of data injection attacks.

The remainder of the paper is organized as follows. The next section describes the state estimation setting and the system model for the data injection attacks. In Section III the data injection problem is cast in a game theoretic framework and the main concepts used in the analysis of the NEs are introduced. The main results reporting the performance metrics, the structure of the best response functions, and the NEs for unconstrained and constrained attacks are presented in Sections IV and V respectively. Because of space limitations we omit the proofs of the results presented in this paper.

II. State Estimation Problem

Consider an electric power distribution network with $N \in \mathbb{N}$ buses and let $V_i \in \mathbb{C}$ be the voltage at bus i,

The authors are with Department of Electrical Engineering at Princeton University, Princeton, NJ 08544, USA.

Iñaki Esnaola is also with the Department of Automatic Control and Systems Engineering, University of Sheffield, S1 3JD, UK.

Samir M. Perlaza is also with the CITI Lab of the Institut National de Recherche en Informatique et en Automatique (INRIA), Université de Lyon and Institut National des Sciences Appliquées (INSA) de Lyon. 6 Av. des Arts 69621 Villeurbanne, France. (esnaola@sheffield.ac.uk, samir.perlaza@inria.fr, poor@princeton.edu).

This work was supported in part by the U.S. National Science Foundation under Grant CMMI-1435778.

with $i \in \{1, \ldots, N\}$. Denote by δ_i and $|V_i|$ the phase and the magnitude of the voltage V_i, respectively. Hence, the state of the grid is fully described by the vector [10]:

$$\mathbf{x} = (\delta_2, \ldots, \delta_N, |V_1|, \ldots, |V_N|) \in \mathbb{R}^{(N-1)} \times \mathbb{R}_+^N. \quad (1)$$

Without any loss of generality, all phases are measured with respect to the phase of voltage V_1 and thus, δ_1 is excluded from the state vector $\mathbf{x}$. The network operator collects M measurements of different variables of the network, i.e., real and/or reactive power injected at a given bus; and/or active and reactive power flows between buses. Note that in some cases, some voltage magnitudes can also be measured. However, in general, all measured variables are functions of the state variables $\delta_2, \ldots, \delta_N$ and $|V_1|, \ldots, |V_N|$. Let $F : \mathbb{R}^{(2N-1)} \to \mathbb{R}^M$ be a multi-dimensional function that describes the relation between the measured variables and the state variables. Variables other than the state variables and the measured variables, such as line-charging susceptances, series susceptances, etc., are considered to be known by the network operator. Denote by $\mathbf{y} \in \mathbb{R}^M$ the vector of measurements obtained by the network operator. Hence, the following equality holds:

$$\mathbf{y} = F(\mathbf{x}) + \mathbf{z} + \mathbf{a}, \quad (2)$$

where $\mathbf{z} \sim \mathcal{N}(0, \mathbf{I}_M)$ is the additive white Gaussian noise arising from the thermal noise added by the metering devices and $\mathbf{a} \in \mathbb{R}^M$ accounts for external malicious data injections. When $\mathbf{a} \neq (0, \ldots, 0)$, it is assumed that there exists an external entity able to tamper with the measurements obtained by the network operator. Often, the attack vector $\mathbf{a}$ is sparse as only a set of measurements is compromised [2]. The construction of the data injection vector $\mathbf{a}$ is thoroughly discussed in Sec. III. The function F in (2) is non-linear and has an involved description determined by the underlying physical laws describing the relation between the measured variables and the state variables $\delta_1, \delta_2, \ldots, \delta_N$ and $|V_1|, \ldots, |V_N|$. Following the reasoning in [10], a linearized model of the form

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z} + \mathbf{a} \quad (3)$$

is considered, where the matrix $\mathbf{H} \in \mathbb{R}^{M \times N}$, with entries $(\mathbf{H})_{i,j} = \dfrac{\partial F(x_i)}{\partial x_j}$, is the Jacobian of function F. It is important to highlight that the structure of the matrix $\mathbf{H}$ is strongly influenced by the indices used to identify each bus. That is, a different labeling of the buses induces a different matrix structure. With the proper labeling, diagonally dominant matrices can be constructed. However, the problem of labeling the buses of the network to induce particular properties in the matrix $\mathbf{H}$ is beyond the scope of this work. In the following, the analysis of the state estimation is kept independent of the resulting structure of $\mathbf{H}$.

Within this context, the state estimation problem consists of constructing an estimation function $f : \mathbb{R}^M \to \mathbb{R}^N$ such that, given an estimate of the form $\hat{\mathbf{x}} = f(\mathbf{y})$, it minimizes a given distortion metric with respect to the actual state vector $\mathbf{x}$. Particularly, this work focuses in the case in which the operator performs GLS estimation. One of the reasons for the widespread acceptance of this technique in the power systems community is that it is frequently used even when the errors introduced in the measurements are not Gaussian. Interestingly, from the point of view of the operator, the attack vector and the thermal noise can be jointly considered as the total error introduced in the measurements, which yields, in general, a non Gaussian error term. That being the case, it is reasonable to assume the operator performs weighted least-squares estimation, in which the weights include the distinct impact of the attack vector. More formally, the weighting matrix is denoted by $\mathbf{W} \in \mathcal{S}_{++}^M$, where $\mathcal{S}_{++}^M$ denotes the set of M-dimensional positive definite matrices. In that setting, the GLS estimation vector $\hat{\mathbf{x}}_{\mathsf{LS}}$ satisfies the following equality:

$$\begin{aligned} \hat{\mathbf{x}}_{\mathsf{LS}} &= \min_{\mathbf{x} \in \mathbb{R}^N} \|\mathbf{W}(\mathbf{y} - \mathbf{H}\mathbf{x})\|_2^2 \\ &= (\mathbf{W}\mathbf{H})^+ \mathbf{W}\mathbf{y}, \end{aligned} \quad (4)$$

where $(\cdot)^+$ denotes the Moore-Penrose inverse [11]. For ease of notation $\mathbf{P} \triangleq (\mathbf{W}\mathbf{H})^+ \mathbf{W}$ is defined. Note that in the special case in which $\mathbf{W}\mathbf{H}$ has full column rank, the Moore-Penrose pseudo-inverse is identical to $(\mathbf{W}\mathbf{H})^{-1}$. Hence, from (4), it follows that

$$\hat{\mathbf{x}}_{\mathsf{LS}} = \mathbf{P}(\mathbf{H}\mathbf{x} + \mathbf{z}) + \mathbf{P}\mathbf{a},$$

and the perturbation $\boldsymbol{\psi}_{\mathsf{LS}}$ induced by the attacker, with the vector $\mathbf{a}$, in the estimation is

$$\boldsymbol{\psi}_{\mathsf{LS}} = \mathbf{P}\mathbf{a}. \quad (5)$$

III. Game Formulation

This section studies the interaction between the network operator and a single malicious entity able to perform a global attack, i.e., to simultaneously tamper with any set of sensors of the network. Under this condition, the attack vector $\mathbf{a}$ can be any vector in the M-dimensional real space $\mathbb{R}^M$. In order to combat the attack, the network operator choses the weighting matrix $\mathbf{W}$ for (4).

Denote by $u : \mathbb{R}^M \times \mathcal{S}_{++}^M \to \mathbb{R}$ a function used by the network operator to choose between two weighting matrices $\mathbf{W}_1$ and $\mathbf{W}_2$, given that the attacker chooses the attack vector $\mathbf{a}$. More specifically, $u(\mathbf{a}, \mathbf{W}_1) > u(\mathbf{a}, \mathbf{W}_2)$ implies that, given attack vector $\mathbf{a}$, weighting matrix $\mathbf{W}_1$ is preferred to $\mathbf{W}_2$. When equality holds, i.e., $u(\mathbf{a}, \mathbf{W}_1) = u(\mathbf{a}, \mathbf{W}_2)$, the network operator indifferently chooses either $\mathbf{W}_1$ or $\mathbf{W}_2$ against the vector attack $\mathbf{a}$. Similarly, denote by $v : \mathbb{R}^M \times \mathcal{S}_{++}^M \to \mathbb{R}$ a function used by the attacker to choose between two attack vectors $\mathbf{a}_1$ and $\mathbf{a}_2$, given that the network operator uses the weighting $\mathbf{W}$. Hence, $v(\mathbf{a}_1, \mathbf{W}) \geqslant v(\mathbf{a}_2, \mathbf{W})$ implies that the vector $\mathbf{a}_1$ is preferred instead of $\mathbf{a}_2$, given the weighting $\mathbf{W}$ and no preference is established when equality holds.

This interaction can be modeled as a game in normal-form [12] denoted by

$$\mathcal{G} = \left\{\mathcal{K}, \{\mathcal{S}_{++}^M, \mathbb{R}^M\}, \{u, v\}\right\}, \quad (6)$$

where $\mathcal{K} = \{\text{Operator}, \text{Attacker}\}$ is the set of players; the set of $M-$dimensional positive-definite matrices $\mathcal{S}_{++}^M$ and the set of real $M-$dimensional vectors $\mathbb{R}^M$ are the set of actions of the operator and the attacker, respectively. Finally, u and v are the utility functions of the operator and the attacker, respectively. The aim of both the operator and the attacker is to maximize their individual utilities. Hence, given any attack $\mathbf{a}$, the best response of the operator is a correspondence $\text{BR}_u : \mathbb{R}^M \to \mathcal{S}_{++}^M$ that maps $\mathbf{a}$ into a set of utility-maximizing M-dimensional positive definite weighting matrices. Thus, BR_u satisfies the following condition:

$$\text{BR}_u(\mathbf{a}) = \{\mathbf{W}^* : u(\mathbf{a}, \mathbf{W}^*) \geqslant u(\mathbf{a}, \mathbf{W}), \forall \mathbf{W} \in \mathcal{S}_{++}^M\}. \tag{7}$$

The best response of the attacker $\text{BR}_v : \mathcal{S}_{++}^M \to \mathbb{R}^M$ is a correspondance that maps any weighting vector into a set of attack vectors. The set $\text{BR}_v(\mathbf{W})$ satisfies the following condition:

$$\text{BR}_v(\mathbf{W}) = \{\mathbf{a}^* : v(\mathbf{a}^*, \mathbf{W}) \geqslant v(\mathbf{a}, \mathbf{W}), \forall \mathbf{a} \in \mathbb{R}^M\}. \tag{8}$$

A. Operator-Attacker Equilibria

Among the many solutions of the game $\mathcal{G}$ [13], an interesting one, denoted by $(\mathbf{a}^*, \mathbf{W}^*)$, is the case in which the weighting $\mathbf{W}^*$ is optimal with respect to the attack vector $\mathbf{a}^*$, and, the attack vector $\mathbf{a}^*$ is optimal with respect to the weighting $\mathbf{W}^*$. This type of outcome is known as a Nash equilibrium (NE) [14]. An NE is stable in the sense that neither the network operator nor the attacker possesses a weighting vector or an attack vector that is preferred to $\mathbf{W}^*$ and $\mathbf{a}^*$, respectively. The following definition formalizes this game solution concept.

Definition 1 *Let the action profile $(\mathbf{a}^*, \mathbf{W}^*)$ be a Nash equilibrium of the game $\mathcal{G}$. Let also $\text{BR} : \mathbb{R}^M \times \mathcal{S}_{++}^M \to \mathbb{R}^M \times \mathcal{S}_{++}^M$ be a correspondence satisfying*

$$\text{BR}(\mathbf{a}, \mathbf{W}) = \text{BR}_u(\mathbf{a}) \times \text{BR}_v(\mathbf{W}). \tag{9}$$

Then, the Nash equilibrium action profile $(\mathbf{a}^*, \mathbf{W}^*)$ satisfies the fixed-point equation

$$(\mathbf{a}^*, \mathbf{W}^*) = \text{BR}(\mathbf{a}^*, \mathbf{W}^*). \tag{10}$$

IV. Unconstrained Attacks

A. Performance Metric of the Network Operator

The main objective of the network operator is to choose a matrix $\mathbf{W}$ such that it obtains a reliable estimate $\hat{\mathbf{x}}$ of the state variables $\mathbf{x}$ in the grid. Such reliability can be measured in terms of a squared ℓ_2-norm of a weighted version of the residual. More specifically, let the function u_1 be defined as

$$\begin{aligned} u_1(\mathbf{a}, \mathbf{W}) &= -\|\mathbf{W}(\mathbf{y} - \mathbf{H}\hat{\mathbf{x}})\|_2^2 \\ &= -\|\mathbf{W}(\mathbf{I} - \mathbf{HP})(\mathbf{Hx} + \mathbf{z} + \mathbf{a})\|_2^2. \end{aligned} \tag{11}$$

Note that the utility function (11) is parametrized by $\mathbf{H}$, $\mathbf{x}$ and $\mathbf{z}$. However, the knowledge of $\mathbf{x}$ and $\mathbf{z}$ is not required at the network operator as the value of $u_1(\mathbf{a}, \mathbf{W})$ can be calculated from the linearization matrix $\mathbf{H}$ and the vector of measurements $\mathbf{y}$, i.e., $u_1(\mathbf{a}, \mathbf{W}) = -\|\mathbf{Ly}\|_2^2$, where $\mathbf{L} \triangleq \mathbf{W}(\mathbf{I} - \mathbf{HP})$ is defined for notational simplicity.

B. Performance Metric of the Attacker

The design of the attack vector, $\mathbf{a}$, faces an important trade-off between maximizing the distortion introduced into the measurements and minimizing the probability of attack detection. Increasing the distortion increases the probability of detection and vice versa. A natural measure of the distortion induced by the vector attack $\mathbf{a}$ is the ℓ_2-norm of the excess distortion term, $\boldsymbol{\psi}_{\mathsf{LS}}$ in (5). On the other hand, a traditional attack detection technique consists of comparing the ℓ_2-norm of a weighted residual $\|\mathbf{Ly}\|_2^2$ with a given threshold τ. Thus, when $\|\mathbf{Ly}\|_2^2 > \tau$ holds, then the existence of an attack is declared by the network operator. Conversely, when $\|\mathbf{Ly}\|_2^2 \leqslant \tau$, the operator assumes that $\mathbf{a}$ is the null vector and thus, no attack is present in the network. Following this reasoning, the aim of the attacker can be modeled by a function v_1 defined as

$$v_1(\mathbf{a}, \mathbf{W}) = \|\mathbf{Pa}\|_2^2 - \lambda\|\mathbf{L}(\mathbf{Hx} + \mathbf{z} + \mathbf{a})\|_2^2. \tag{12}$$

C. Nash Equlibria

The NEs of the game $\mathcal{G}$ are studied in the case in which the attacker has access to full system information, i.e., the realizations of $\mathbf{H}$, $\mathbf{x}$, and $\mathbf{z}$. Contrastingly, the network operator estimates the state with knowledge only of the second order statistics of the disturbance, i.e., $\boldsymbol{\Sigma}_{\mathbf{a}+\mathbf{z}} = \mathbb{E}[(\mathbf{a}+\mathbf{z})(\mathbf{a}+\mathbf{z})^T]$. The rationale for this setting is to determine the security limits by investigating the least favorable situation for the network operator. Note that neither the attacker nor the network operator has knowledge about the statistics of the state variables. When only the measurements, $\mathbf{y}$, and the Jacobian, $\mathbf{H}$, are available to the network operator, the strategy reduces to $\text{BR}_u(\mathbf{a}) = \mathbf{I}$. To avoid this trivial solution, the knowledge of $\boldsymbol{\Sigma}_{\mathbf{a}+\mathbf{z}}$ is incorporated into the construction of the weighting matrix, $\mathbf{W}$, of the GLS estimator.

The following theorem describes the NE arising from the utility functions described above.

Theorem 1 *Let $(\mathbf{a}_1^*, \mathbf{W}_1^*)$ be an NE action profile of the game $\mathcal{G}$. Assume that $\lambda \geqslant 0$, $\mathbf{P}^* \triangleq (\mathbf{W}_1^*\mathbf{H})^+ \mathbf{W}_1^*$; $\mathbf{L}^* \triangleq \mathbf{W}_1^*(\mathbf{I} - \mathbf{HP}^*)$; and $\mathbf{T}^* \triangleq (\mathbf{P}^*)^T\mathbf{P}^* - \lambda(\mathbf{L}^*)^T\mathbf{L}^*$. Then, when $\mathbf{T}^*$ is negative definite, $(\mathbf{a}_1^*, \mathbf{W}_1^*)$ satisfies the fixed-point equation in* (9) *with*

$$\text{BR}_{u_1}(\mathbf{a}_1^*) = \boldsymbol{\Sigma}_{\mathbf{z}+\mathbf{a}_1^*}^{-\frac{1}{2}}, \tag{13}$$

$$\begin{aligned} \text{BR}_{v_1}(\mathbf{W}_1^*) = \lambda &\left((\mathbf{T}^*)^- + \mathbf{B} - (\mathbf{T}^*)^-\mathbf{T}^*\mathbf{B}\mathbf{T}^*(\mathbf{T}^*)^-\right) \\ &\times (\mathbf{L}^*)^T\mathbf{L}^*(\mathbf{Hx} + \mathbf{z}), \end{aligned} \tag{14}$$

where $(\cdot)^-$ *denotes any weak inverse and* $\mathbf{B} \in \mathbb{R}^{M\times M}$ *is any arbitrary matrix.*

The following corollaries follow immediately from Theorem 1:

Corollary 1 *When* $\mathbf{T}^*$ *is positive definite, the game* $\mathcal{G}$ *possesses infinitely many NEs.*

The fact that there are infinitely many NEs and attack vectors that maximize the performance metric v_1, makes defining a strategy for the operator hard. Indeed, due to the assumption that $\mathbf{\Sigma}_{\mathbf{z}+\mathbf{a}}$ is known, the operator is exposed to a random attack strategy. For instance, the attacker can generate a random matrix $\mathbf{B}$ with each attack construction, which effectively complicates the estimation of $\mathbf{\Sigma}_{\mathbf{z}+\mathbf{a}}$ for the network operator.

Corollary 2 *When* $\mathbf{T}^*$ *is not positive definite, the game game* $\mathcal{G}$ *does not possess an NE.*

The following section presents an energy constrained attack scenario in which the existence of at least one NE is ensured and the optimal attack vector does not depend on the state variables.

V. Energy Constrained Attacks

The attack construction in Section IV does not impose any energy constraint on the attacker. However, the Jacobian is often rank deficient and the construction of the optimal attack vector, $\mathbf{a}_1^*$, in Theorem 1 can be ill-conditioned and possibly very large. To that end, in this section a new performance metric is proposed that effectively limits the amount of energy that is available to the attacker.

A. Performance Metric in the Energy Constrained Case

In the case of the network operator, the same function as in Section IV is used and therefore

$$u_2\left(\mathbf{a}, \mathbf{W}\right) = -\|\mathbf{L}\mathbf{y}\|_2^2. \tag{15}$$

On the other hand, a new term capturing the amount of energy used in the attack construction is added to the performance metric given in (12), which results in

$$v_2(\mathbf{a}, \mathbf{W}) = \|\mathbf{P}\mathbf{a}\|_2^2 - \lambda_1\|\mathbf{L}\left(\mathbf{H}\mathbf{x}+\mathbf{z}+\mathbf{a}\right)\|_2^2 - \lambda_2\|\mathbf{a}\|_2^2. \tag{16}$$

In this case the trade-offs of the performance metric are expressed through the parameters λ_1 and λ_2 which scale the impact of detectability and available attack energy, respectively.

B. Nash Equilibria with Energy Constraints

The following theorem reflects the impact on the NE in Theorem 1 of adding energy constraints to the attacker.

Theorem 2 *Let* $(\mathbf{a}_2^*, \mathbf{W}_2^*)$ *be an NE action profile of the game* $\mathcal{G}$*. Assume that* $\lambda_1 \geqslant 0$ *and* $\lambda_2 \geqslant 0$*. Define* $\mathbf{P}^* \triangleq \left(\mathbf{W}_2^*\mathbf{H}\right)^+ \mathbf{W}_2^*$*;* $\mathbf{L}^* \triangleq \mathbf{W}_2^*\left(\mathbf{I} - \mathbf{H}\mathbf{P}^*\right)$*; and* $\mathbf{T}^* \triangleq \left(\mathbf{P}^*\right)^T\mathbf{P}^* - \lambda_1\left(\mathbf{L}^*\right)^T\mathbf{L}^*$*. Then, when* $\mathbf{T}^* - \lambda_2\|\mathbf{a}_2^*\|_2^2$ *is negative definite,* $(\mathbf{a}_2^*, \mathbf{W}_2^*)$ *satisfies the fixed-point equation in* (9) *with*

$$\mathrm{BR}_{u_2}(\mathbf{a}_2^*) \;\; = \mathbf{\Sigma}_{\mathbf{z}+\mathbf{a}_2^*}^{-\frac{1}{2}}\,, \tag{17}$$

$$\mathrm{BR}_{v_2}(\mathbf{W}_2^*) = \lambda_1\left(\mathbf{T}^* - \lambda_2\mathbf{I}\right)^+\left(\mathbf{L}^*\right)^T\mathbf{L}^*\mathbf{z}. \tag{18}$$

Interestingly, in this case the construction of the optimal attack is easier for the attacker as it requires less knowledge than the one proposed in (14). Indeed, the attack vector does not depend on the state variables of the grid and is obtained by linearly projecting the noise vector $\mathbf{z}$. Remarkably, this result shows that the optimal attack construction is a Gaussian distributed vector with covariance matrix defined by the Jacobian of the grid. Another interesting aspect is the fact that the construction in (18) is the attack vector with the minimum-norm solution in the set $\mathrm{BR}_{v_2}(\mathbf{W}_2^*)$. In fact, the following corollary gives an upper bound on the amount of energy that the attacker requires.

Corollary 3 *Let* $\mathbf{T}^* - \lambda_2\mathbf{I}$ *have rank* $r < M$ *and let* σ_{min} *be the smallest non-zero singular value of* $\mathbf{T}^* - \lambda_2\mathbf{I}$*. Then* $\mathbf{a}_2^*$ *is the minimum-norm solution in the set* $\mathrm{BR}_{v_2}(\mathbf{W}_2^*)$ *and*

$$\|\mathbf{a}_2^*\| \leq \frac{\|\lambda_1(\mathbf{L}^*)^T\mathbf{L}^*\mathbf{z}\|_2^2}{\sigma_{min}}. \tag{19}$$

VI. Conclusion

Different performance metrics give rise to different models of the interaction between the attacker and the network operator. The choice of performance metrics conditions the structure of the game and the meaning of the corresponding NEs. In this paper, two performance metrics have been proposed and the induced sets of NEs have been fully characterized. From the perspective of the attacker, these new performance metrics capture the intrinsic antagonism between distortion maximization, energy limitation, and minimization of the attack detection probability. In the case of the network operator, the performance metric is determined by the distortion in the estimation of the state variables. At the equilibrium points the effect of adding an energy constraint is that the optimal attack vector does not depend on the state variables.

References

[1] E. Hossain, Z. Han, and H. V. Poor, *Smart Grid Communications and Networking*, Cambridge University Press, 2012.

[2] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," in *Proc. of ACM Conference on Computer and Communications Security*, Chicago, IL, USA, Nov. 2009, pp. 21–32.

[3] O. Kosut, L. Jia, R. J. Thomas, and L. Tong, "Malicious data attacks on the smart grid," *IEEE Transactions on Smart Grid*, vol. 2, no. 4, pp. 645–658, Oct. 2011.

[4] T. T. Kim and H. V. Poor, "Strategic protection against data injection attacks on power grids," *IEEE Transactions on Smart Grid*, vol. 2, no. 2, pp. 326–333, Jun. 2011.

[5] M. Ozay, I. Esnaola, F. T Yarman-Vural, S. R. Kulkarni, and H. V. Poor, "Smarter security in the smart grid," in *Proc. of IEEE International Conference on Smart Grid Communications*, Nov. 2012.

[6] S. Cui, Z. Han, S. Kar, T. T. Kim, H. V. Poor, and A. Tajer, "Coordinated data-injection attack and detection in the smart grid: A detailed look at enriching detection solutions," *IEEE Signal Processing Magazine*, vol. 29, no. 5, pp. 106–115, Sep. 2012.

[7] A. Tajer, "Energy grid state estimation under random and structured bad data," in *Proc. IEEE Sensor Array and Multichannel Signal Processing Workshop*, Jun. 2014.

[8] M. Ozay, I. Esnaola, F. Yarman-Vural, S. Kulkarni, and H. V. Poor, "Sparse attack construction and state estimation in the smart grid: Centralized and distributed models," *IEEE Journal on Selected Areas in Communications: Smart Grid Communications Series*, vol. 31, no. 7, pp. 1306–1318, Jul. 2013.

[9] G. H. Golub and C. F. Van Loan, *Matrix Computations*, JHU Press, Baltimore, MD, USA, 2012.

[10] J. J. Grainger and W. D. Stevenson, *Power System Analysis*, vol. 621, McGraw-Hill New York, Hightstown, NJ, USA, 1994.

[11] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, UK, 2012.

[12] D. Fudenberg and J. Tirole, *Game Theory*, The MIT Press, Cambridge, MA, USA, 1991.

[13] S. M. Perlaza and S. Lasaulce, *Game-Theoretic Solution Concepts and Learning Algorithms*, in Mechanisms and Games for Dynamic Spectrum Allocation, *Eds.* T. Alpcan, H. Boche, M. Honig, and H.V. Poor, Cambridge University Press, New York, NY, USA, 2014.

[14] J. F. Nash, "Equilibrium points in n-person games," *Proc. National Academy of Sciences of the United States of America*, vol. 36, no. 1, pp. 48–49, Jan. 1950.