Data Attacks on Power System State Estimation: Limited Adversarial Knowledge vs. Limited Attack Resources

Kaikai Pan*, André Teixeira†, Milos Cvetkovic* and Peter Palensky*
*Intelligent Electrical Power Grids
Faculty of EEMCS, Delft University of Technology, Delft, The Netherlands
†Engineering Systems and Services
Faculty of TPM, Delft University of Technology, Delft, The Netherlands

Abstract—**It has shown that with perfect knowledge of the system model and the capability to manipulate a certain number of measurements, the false data injection (FDI) attacks, as a class of data integrity attacks, can coordinate measurements corruption to keep stealth against the bad data detection schemes. However, a more realistic attack is essentially an attack with limited adversarial knowledge of the system model and limited attack resources due to various reasons. In this paper, we generalize the data attacks that they can be pure FDI attacks or combined with availability attacks (e.g., DoS attacks) and analyze the attacks with limited adversarial knowledge or limited attack resources. The attack impact is evaluated by the proposed metrics and the detection probability of attacks is calculated using the distribution property of data with or without attacks. The analysis is supported with results from a power system use case. The results show how important the knowledge is to the attacker and which measurements are more vulnerable to attacks with limited resources.**

I. Introduction

The integration of information and communication technology (ICT) and power systems makes the intelligent power grids typical cyber-physical systems. These systems are operated by means of complex distributed software systems which transmit information through wide and local area networks [1]. Thus the intelligent power grids would be exposed to a large number of security threats [2], [3]. The heterogeneity, diversity, and complexity of intelligent power grids may introduce new vulnerabilities that may lead to severe consequences [4].

The State Estimation (SE) within modern energy management systems (EMS) is an instance of such dependency. It is supported by the Supervisory Control and Data Acquisition (SCADA) system for data delivery and provides important information to the EMS for power grids monitoring and control. SE uses measurements collected by the Remote Terminal Units (RTUs) in substations and transmitted through the SCADA communication network to the control center. There is a built-in bad data detection (BDD) process in SE to detect erroneous measurements. The estimated state information is then processed by other applications in EMS such as Optimal Power Flow and Contingency Analysis to compute optimal control action while ensuring reliability and safety, as is indicated in Figure 1. The critical nature of SE highlights the importance of making it accurate and secure for power grid operations. In order to increase the security of SE and EMS, one needs to conduct vulnerability and attack impact assessment. Some of the literature has already tackled these problems. Vulnerability of SE to data integrity attack (e.g., FDI attack) is quantified by computing the attack resources needed by the adversary to keep stealth against the BDD [5]–[7]. The attack impacts of FDI attacks on SE, such as introduced estimate errors [8], potential economic loss in market operation [9], [10], physical damaging like line overflows, are also well presented and analyzed. Besides, the FDI attacks with limited knowledge or limited resources are discussed by restricting the knowledge of the adversary to a part of the system model [11] or a subset of the network [12], or restricting the capability of the attacker to manipulate the number of sensors [8]. Our recent work [13] extends the attack scenarios that the SE can be corrupted by FDI attacks and data availability attacks (e.g., Denial of Service (DoS) attacks) simultaneously.

In this paper we aim to contribute in analyzing data attacks with limited adversarial knowledge and limited attack resources. Here the data attacks are "generalized" that can be pure integrity attacks (i.e. FDI attacks) or combined integrity and availability attacks. In order to achieve this, we introduce attack vectors for FDI attacks and availability attacks respectively and formulate attack strategies under both scenarios of limited knowledge and limited resources. To compare attacks under these two scenarios, we also propose attack impact metrics for evaluating attack impact on load estimate and provide methods to calculate detection probability of the attacks under these two scenarios. We show how important the knowledge of the system model is to the adversary and which measurements have the priority to be protected when attacks have limited resources. The analysis is supported with results from a case study.

The outline of the paper is as follows. Section II details the state estimation techniques, the optimal attack strategy with full knowledge and attack resources. The attack impact metrics are also proposed to evaluate impact on load estimate. In Section III, the first scenario that attacks with limited adversarial knowledge is discussed. The method to calculate

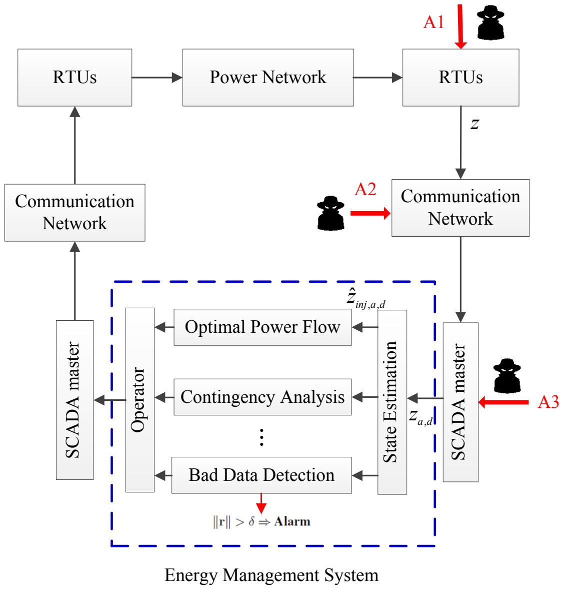

Figure 1. A schematic block diagram of the power system, SCADA system and EMS. An attacker can launch integrity attacks or availability attacks or both simultaneously on the measurements at various levels A1-A3 in the system. The figure is based on [6, Fig.1].

the detection probability of the attacks is discussed and a special case of limited knowledge scenario is specified. The second scenario that attacks with limited attack resources is presented in detail in Section IV, including the optimal data attacks with limited resources and the computation solution for solving it. Section V shows the results from the case study. The conclusion remarks are in Section VI.

II. System Model and Data Attacks

In this section, we review the state estimation problem and discuss the optimal attacks with perfect knowledge and full resources. The attack impact metrics are also formulated.

A. DC State Estimation and Bad Data Detection

A power system model has a number of buses connected by transmission lines. The measurement data collected by RTUs includes line power flow measurements and bus power injection (generation, load) measurements. In the formulation, we use the DC power system model which is commonly employed in security analysis of SE by neglecting the reactive power, line losses and assuming the voltage magnitudes to be constant.

We assume that there are $n+1$ buses and n_t transmission lines in the power network. Line power flow and bus power injection measurements are collected by RTUs in each substation. These m power flow measurements are denoted by $z = [z_1, \ldots, z_m]^T$. The DC SE solves the following problem,

$$z = \begin{bmatrix} P_1 W B^T \\ -P_2 W B^T \\ P_3 B_0 W B^T \end{bmatrix} x + e := Hx + e, \tag{1}$$

where $H \in \mathbb{R}^{m \times n}$ represents the system model, depending on the parameters of transmission lines (i.e., matrix W), the network topology (i.e., matrix B_0) and the placement of RTUs (i.e., matrices P_1, P_2, P_3). Here the topology is described by a directed incidence matrix $B_0 \in \mathbb{R}^{(n+1)\times n_t}$ in which the directions of the lines can be arbitrarily specified [7]. Matrix $B \in \mathbb{R}^{n \times n_t}$ is the truncated incidence matrix with the row in B_0 corresponding to the reference bus removed. Matrix $W \in \mathbb{R}^{n_t \times n_t}$ is a diagonal matrix whose diagonal entries are the reciprocals of transmission line reactance. Matrices P_1, P_2 and P_3 are stacked by the rows of identity matrices, indicating whether a particular line power flow at the both sides of lines or bus power injection is measured. The total number of rows of P_1, P_2 and P_3 is m. The state vector $x = [x_1, \ldots, x_n]^T$ refers to phase angles on each bus except the reference one and $e \sim \mathcal{N}(0, R)$ is the measurement noise vector of independent zero-mean Gaussian variables with covariance matrix $R = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_m^2)$.

The state estimate $\hat{x}$ can by obtained using weighted least squares (WLS) estimate:

$$\hat{x} = \arg\min_x (z - Hx)^T \Sigma^{-1} (z - Hx), \tag{2}$$

which can be solved as

$$\hat{x} = (H^T R^{-1} H)^{-1} H^T R^{-1} z := Kz. \tag{3}$$

To validate state estimates, the bad data detection schemes are used to detect erroneous measurements. The algorithms of the BDD within SE are based on the measurement residual

$$r = z - H\hat{x} = (I - HK)z := S z, \tag{4}$$

where $r \in \mathbb{R}^m$ is the residual vector, $I \in \mathbb{R}^{m \times m}$ is an identity matrix, and S is the so-called residual sensitivity matrix [14]. The BDD is based on checking whether the p-norm of (weighted) measurement residual is below some threshold τ. In this paper we choose the $J(\hat{x})$-test based BDD, which uses the quadratic function $J(\hat{x}) = \|R^{-1/2} r\|_2^2$ to check if it follows the chi-squared distribution χ^2_{m-n}. The BDD scheme becomes

$$\begin{cases} \text{Good data, if} & \|R^{-1/2} r\|_2^2 \le \tau(\alpha), \\ \text{Bad data, if} & \|R^{-1/2} r\|_2^2 > \tau(\alpha), \end{cases} \tag{5}$$

where $\tau(\alpha)$ is the threshold corresponding to the false alarm rate α. Defining the probability distribution function (PDF) of χ^2_{m-n}, $\tau(\alpha)$ can be obtained by solving

$$\int_0^{\tau(\alpha)} f(x)dx = 1 - \alpha. \tag{6}$$

B. Optimal Data Attacks with Perfect Knowledge and Full Resources

An adversary aims to perturb the state estimate and keep stealth against the BDD, by tampering of RTUs, the SCADA system or even the SCADA master in the control center. To generalize the potential data attack, we assume that the attacker will use all tools available and can launch both data integrity and availability attacks. Corrupted by such data attacks, the measurement vector z is changed into

$$\bar{z} := (I - \mathrm{diag}(d))z + a, \tag{7}$$

where $a \in \mathbb{R}^m$ is the *FDI attack vector*, $d \in \{0,1\}^m$ is the *availability attack vector* and $I \in \mathbb{R}^{m \times m}$ is an identity matrix. Throughout this paper we define $diag(d)$ as the $m \times m$ diagonal matrix with the elements of vector d on the main diagonal. To describe the number of attacked measurements needed by the adversary, we have the following definition of a (k_a, k_d)-tuple attack,

Definition 1 ((k_a,k_d)-tuple attack)**.** *A data attack with an FDI attack vector $a \in \mathbb{R}^m$ and an availability attack vector $d \in \{0,1\}^m$ described above is called a (k_a,k_d)-tuple attack if $\|a\|_0 = k_a$, $\|d\|_0 = k_d$.*

In our recent work [13], we showed that if the attacker has the perfect knowledge of the system model (i.e., H) and can manipulate certain number of measurements with full attack resources to perform a (k_a,k_d)-tuple attack, it can keep stealth against the BDD by introducing the two attack vectors a and d satisfying $a = (I - \text{diag}(d))Hc$ where $c \in \mathbb{R}^n$ is non-zero. Under such an attack, we define the matrix of the system model and the noise vector as a function of the attack vector d,

$$H_d := (I - \text{diag}(d))H, \quad e_d := (I - \text{diag}(d))e. \tag{8}$$

where H_d denotes the model of the remaining measurements and it is obtained from H by replacing some rows with zero row vectors due to availability attacks on these measurements, e_d is the noise vector for remaining measurements and the entries of it are zero if the corresponding measurements are unavailable. According to the formulation of matrix K in (3) and matrix S in (4), we can also have

$$K_d := (H_d^T R^{-1} H_d)^{-1} H_d^T R^{-1}, \tag{9}$$

$$S_d := I - H_d K_d. \tag{10}$$

An optimal attack pursue minimum attack resources. To simplify the discussion, we assume that an optimal attack can have the minimum attack resources when it needs to corrupt the minimum number of measurements. For the adversary with the capability to manipulate a certain number of measurements and perfect knowledge of the system model, we formulate the optimal attack strategy as the following optimization problem,

$$\begin{aligned} (c^*, d^*) := \arg\min_{c,d} \quad & \|a\|_0 + \|d\|_0 \\ \text{s.t.} \quad & a = H_d c, && \text{(11a)} \\ & H_d = (I - \text{diag}(d))H, && \text{(11b)} \\ & a(j) = \mu, && \text{(11c)} \\ & d(i) \in \{0,1\} \quad \text{for all } i. \end{aligned}$$

Here in (11b) we assume $a(j) = \mu$ where μ is the non-zero attack magnitude on the target measurement j. For measurement j, the optimal attack with attack vectors d^* and $a^* = H_{d^*}c^*$ has the minimum number of measurements to corrupt and correspondingly has the minimum attack resources. To solve the optimization problem above, we propose a computation solution which uses the big M method:

$$\begin{aligned} (c^*, w^*, d^*) := \arg\min_{c,w,d} \quad & \sum_{i=1}^{m} w(i) + \sum_{k=1}^{m} d(k) \\ \text{s.t.} \quad & Hc \leq M(w + d), && \text{(12a)} \\ & -Hc \leq M(w + d), && \text{(12b)} \\ & H(j,:)c = \mu, && \text{(12c)} \\ & w(i) \in \{0,1\} \quad \text{for all } i, && \text{(12d)} \\ & d(k) \in \{0,1\} \quad \text{for all } k, && \text{(12e)} \end{aligned}$$

where $w, d \in \{0,1\}^m$ in (12d) and (12e). $w(i) = 1$ and $d(k) = 1$ means that FDI attack and data availability attack take place on measurement i and k respectively. The following theorem, which is adopted from our work [13], states that the optimal solution to (11) can be exactly obtained by solving (12).

Theorem 1. *For any measurement index $j \in \{1, \ldots, m\}$ and non-zero μ, let (c^*, w^*, d^*) be an optimal solution to* (12)*. Then an optimal solution to* (11) *can be computed as (c^*, d^*).*

C. Attack Impact Metric

As the work in [15] shows, the Optimal Power Flow application uses the load estimate as the inputs provided by SE. If data attacks take place and pass the BDD, the load estimate gets perturbed and it will influence the control actions in the next time interval. Therefore, we consider the impact metric as a function of the bias introduced by the attack on the load estimate.

Assuming that there are m_{inj} injection (with load) measurements, we consider the impact on the errors of power injection (with load) estimate, which can be described as

$$\epsilon = \hat{z}_{inj,a,d} - z_{inj}, \tag{13}$$

where $z_{inj} \in \mathbb{R}^{m_{inj}}$ is the vector of power injection (with load) measurements without attacks and $\hat{z}_{inj,a,d} \in \mathbb{R}^{m_{inj}}$ is the vector of estimated injection (with load) measurements under a (k_a,k_d)-tuple attack. We can further obtain

$$\epsilon = H_{inj}\hat{x}_{a,d} - (H_{inj}x + e_{inj}), \tag{14}$$

where $\hat{x}_{a,d} = K_d\bar{z} = x + K_d e_d + K_d a$, $H_{inj} \in \mathbb{R}^{m_{inj} \times n}$ denotes the submatrix of H by keeping the rows corresponding to injection (with load) measurements, i.e. $H_{inj} = M_{inj}H$ where $M_{inj} \in \mathbb{R}^{m_{inj} \times m}$ is the incidence matrix for selecting the rows in H corresponding to injection (with load) measurements, $e_{inj} \in \mathbb{R}^{m_{inj}}$ is the noise vector for injection (with load) measurements. Thus $\epsilon = H_{inj}K_d a - M_{inj}S_d e$. The expected value of injection (with load) estimate errors is

$$\mathbb{E}(\epsilon) = H_{inj}K_d a. \tag{15}$$

We have the following definition for the attack impact metric,

Definition 2. *The impact metric $\mathbf{I}(a,d)$ for quantifying attack impact of a (k_a,k_d)-tuple attack with attack vectors a and d on load estimate is defined as the 2-norm of $H_{inj}K_d a$, i.e. $\mathbf{I}(a,d) = \|H_{inj}K_d a\|_2$.*

III. Scenario 1: Attacks with Limited Adversarial Knowledge

In this section, we consider the first scenario that the adversary has limited knowledge of the system model and discuss how this would affect the detectability of data attacks.

A. Perturbed Model Known by the Attacker

In Section II-B, the attacks are assumed to have perfect knowledge of the system model. This requires knowledge on topology matrix B_0, the line parameter matrix W and the RTU placement matrix P_1, P_2, P_3. Such knowledge can be obtained by recording and analyzing data sent from RTUs using statistical methods. However, due to restricted access to the power grid, errors in data collection and analyzing may essentially result in an attack with limited knowledge of the system model. We denote the perturbed system model known by the attacker as $\tilde{H}$, such that

$$\tilde{H} = H + \Delta H, \tag{16}$$

where $\Delta H \in \mathbb{R}^{m \times n}$ denotes the part of model uncertainty because of the issues discussed above.

B. Detection Probability of Attacks

When the measurements are corrupted by a (k_a, k_d)-tuple attack, the measurement residual $r(a,d)$ can be written as

$$r(a,d) = S_d\bar{z} = S_d e_d + S_d a. \tag{17}$$

As discussed in Section II-B, when the vectors of the (k_a, k_d)-tuple attack satisfy $a = H_d c$, the residual $r(a,d) = S_d e_d + S_d H_d c = S_d e_d$ due to $S_d H_d = 0$. Then the residual is not affected by a and no increase of alarms would triggered in the BDD since the BDD would treat the measurements attacked by availability attacks as a case of data missing. However, when the attacker has limited knowledge of the system model, the attack vector a becomes $a = (I - \mathrm{diag}(d))\tilde{H}c := \tilde{H}_d c$ and $S_d a$ may be non-zero in this case. The measurement residual is incremented and the attack can be detected with some probability. In the following we show how the detection probability can be calculated.

Note that the quadratic cost function with a (k_a, k_d)-tuple attack becomes $J_{a,d}(\hat{x}) = \|R^{-1/2} r(a,d)\|_2^2$. We can further obtain $J_{a,d}(\hat{x}) = \|R^{-1/2} S_d e_d + R^{-1/2} S_d a\|_2^2$. Here the mean of $(R^{-1/2} S_d e_d + R^{-1/2} S_d a)$ becomes non-zero $R^{-1/2} S_d a$ incremented by the attack. Under the (k_a, k_d)-tuple attack, $J_{a,d}(\hat{x})$ has a *generalized non-central chi-squared distribution* with $m - n - k_d$ degrees of freedom. We use $J_{a,d}(\hat{x})$ as an approximation of having the *non-central chi-squared distribution* $\chi^2_{m-n-k_d}(\|R^{-1/2} S_d a\|_2^2)$ to calculate the detection probability, where $\lambda_{a,d} = \|R^{-1/2} S_d a\|_2^2$ is the non-centrality parameter. In [16], we have validated such approximation using empirical results from Monte Carlo simulation. It implies that data attacks with limited adversarial knowledge would increase the possibility to trigger alarms in BDD due to the introduced non-zero non-centrality parameter. We can get

$$\int_0^{\tau_d(\alpha)} f_{\lambda_{a,d}}(x)dx = 1 - \delta_{a,d}, \tag{18}$$

where $f_{\lambda_{a,d}}(x)$ is the PDF of $\chi^2_{m-n-k-d}(\|R^{-1/2} S_d a\|_2^2)$, $\tau_d(\alpha)$ is the threshold set in the BDD using (6) but with the PDF of $\chi^2_{m-n-k_d}$, and $\delta_{a,d}$ is the detection probability of the (k_a, k_d)-tuple attack. In order to keep stealth, the attacker has to minimize $\delta_{a,d}$ as close to the false alarm rate α as possible.

C. A Special Case of Limited Adversarial Knowledge

The model uncertainty defined in (16) is "general" and can be any values. An interesting analysis is to understand what the model uncertainty ΔH could be to the adversary. In particular, we are interested in the case where the adversary knows the exact topology of the power network and the placement of RTUs, but has limited information of the line parameters. This can be expected due to various practical reasons as explained in [11], e.g., limited access to the knowledge of exact position of the tap changer and the exact length of the transmission line and type of the conductor being used. Thus the perturbed system model known by the adversary becomes

$$\tilde{H} = P \begin{bmatrix} (W + \Delta W)B^T \\ -(W + \Delta W)B^T \\ B_0(W + \Delta W)B^T \end{bmatrix}, \tag{19}$$

where $\Delta W \in \mathbb{R}^{n_t \times n_t}$ denotes the error on estimate of transmission line reactance.

IV. Scenario 2: Attacks with Limited Attack Resources

In this section we consider the second scenario where the adversary has limited attack resources but still targets to keep stealth against the BDD and have maximum attack impact.

A. Optimal Data Attacks with Limited Attack Resources

The attack policies in Section II-B for the (k_a, k_d)-tuple attacks follow the linear model and the adversary is assumed to be able to attack a certain number of measurements, i.e., $k_a + k_d \geq \min \|a^*\|_0 + \|d^*\|_0$ where attack vectors d^* and $a^* = H_{d^*} c^*$ are obtained by solving (12) for any measurement index j. Now we consider the following scenario that the attacker has limited attack resources that $k_a + k_d < \min \|a^*\|_0 + \|d^*\|_0$ and thus can not follow the linear attack policies above. For the sake of comparison, in this scenario we assume that the attacker has full knowledge of the system model and enough computational capability. In the following we construct the optimal attack strategy for the (k_a, k_d)-tuple attacks with limited attack resources.

The objective of the attack is to gain an ability to introduce error on load estimate. The adversary tries to achieve the goal by maximizing the impact metrics, i.e., maximizing $\|H_{inj} K_d a\|_2$ for a (k_a, k_d)-tuple attack.

An optimal attack also targets to keep stealth against the BDD, or at least minimize the detection probability. In the case that the adversary has limited attack resources, the attacks can be detected since the term $S_d a$ in (17) may be non-zero. We assume there exists an upper limit of detection probability $\overline{\delta}$ which is acceptable for the adversary. Thus according to Subsection III-B, for the (k_a, k_d)-tuple attack, we can obtain

$$\int_0^{\tau_d(\alpha)} f_{\lambda_{a,d}}(x)dx \geq 1 - \overline{\delta}, \tag{20}$$

where the non-centrality parameter is $\lambda_{a,d} = \|R^{-1/2} S_d a\|_2^2$ and $f_{\lambda_{a,d}}(x)$ is the PDF of $\chi^2_{m-n-k_d}(\|R^{-1/2} S_d a\|_2^2)$. For a given $\overline{\delta}$, the non-centrality parameter satisfies $\lambda_{a,d} = \|R^{-1/2} S_d a\|_2^2 \leq \overline{\varepsilon}_{a,d}$ where $\overline{\varepsilon}_{a,d}$ can be determined using (20).

Based on the aforementioned intuition, we consider the following optimization problem for the optimal attack strategy of the (k_a, k_d)-tuple attack under the relaxation on the assumption of attack resources,

$$\begin{aligned} \max_{a,d} \quad & \|H_{inj} K_d a\|_2^2 \\ \text{s.t.} \quad & \|R^{-1/2} S_d a\|_2^2 \leq \overline{\varepsilon}_a, \\ & \|a\|_0 + \|d\|_0 \leq \overline{R}. \end{aligned} \tag{21}$$

where $\overline{R}$ denotes the maximum number of measurements that the attacker can manipulate.

B. Computation Solution

The optimization problem of (21) for a (k_a, k_d)-tuple attack is non-convex and difficult to solve without the specifications of the attack vectors [8]. We consider to add more constraints on the attack vectors by setting the measurements which could be attacked to be determined for a given k_a and k_d which satisfy $k_a + k_d \leq \overline{R}$. Thus for this specifications of the non-zero entries in attack vectors a and d, the above problem can be equivalent to solve the following one,

$$\begin{aligned} \min_{a} \quad & \overline{\varepsilon}_{a,d} \varphi \\ \text{s.t.} \quad & Q - \varphi W \leq 0, \end{aligned} \tag{22}$$

where $Q = Qs^T Qs$, $W = Ws^T Ws$, and for a givn d, Qs, Ws are the submatrices of $H_{inj} K_d$ and $R^{-1/2} S_d$ corresponding to the non-zero entries of attack vector a. (22) also implies that φ is the maximum generalized eigenvalue of the matrix pair (Q, W), i.e., $\varphi = \lambda_{max}(Q, W)$ [17].

It should be noted that though the attack vectors are constrained in order to solve the optimization problem, the optimal attacks for any given k_a and k_d which satisfy $k_a + k_d \leq \overline{R}$ can still be obtained using exhaustive search over all possible attacked measurement sets. Some computationally efficient algorithms can also be developed to solve (21). We leave this for future work.

V. Case Study

In this section we apply the attack scenarios of limited adversarial knowledge and limited attack resources to the IEEE 14-bus system use case. Full measurements placement is employed that power flow measurements are placed on all the buses and transmission lines to provide large redundancy. The per-unit system is used and the power base is $100MW$. The power flow measurements are generated by the DC model with Gaussian noise ($\sigma_i = 0.02$ for all measurements). For the limited knowledge scenario, we assume that the adversary knows the exact topology of the system but has an estimated line parameters with errors up to ±10%, ±20% and ±30%.

With different levels of error on estimation of line parameters, the detection probability of the attacks can be obtained according to (18). We pick such case that measurement $j = 9$. According to the optimal solution of (12) w.r.t. $\tilde{H}$ in (19), a number of 11 measurements need to be manipulated by the attacker. Figure 2 shows the detection probability of the (5,6)-tuple attacks targeting on these 11 measurements. The x-axis indicates the attack magnitude μ on measurement j using the attack vector $a = \tilde{H}_d c$. We can see that the detectability of attack is intimately related to the error on estimation of line parameters. With larger model uncertainty in building attack vectors, the attack has higher possibility to be detected by the BDD. The detection probability becomes much higher when the error on estimation of line parameters increases. This implies that in order to keep stealth, the adversary do need good knowledge of the system model.

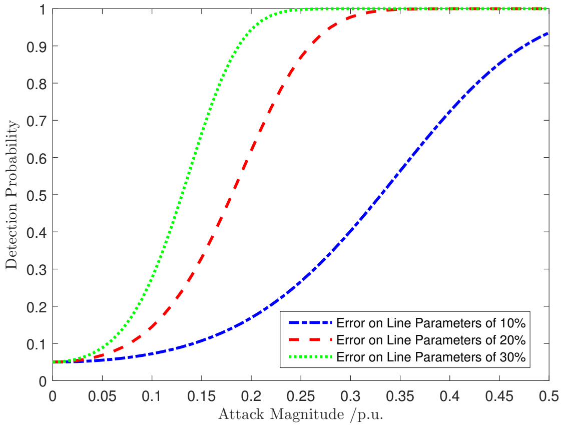

Figure 2. The detection probability is plotted versus the magnitude μ on measurement $j = 9$. The attacks are under there levels of error on estimation of parameters: ±10%, ±20%, ±30%. The false alarm rate is set to be $\alpha = 0.05$.

Next the simulations are conducted in the scenario where the adversary has limited attack resources but full knowledge of the system and the optimal attack strategy described in (21) is used. The measurements that could be manipulated are specified on a determined measurement set which is the same as the one in the previous case, i.e., the set with 11 measurements containing measurement $j = 9$. If all the measurements in this set are corrupted with enough attack resources, the attack with full knowledge can perform the optimal attack strategy in (11) and keep stealth. However, in this scenario the attacker could only corrupt part of the measurements in the set thus can be observed by the BDD. Using (21), we compare the attack scenarios where the attacker has limited resources to corrupt part of the measurements but full knowledge and has limited knowledge but full resources to corrupt all the measurements in this measurement set, as shown in Figure 3. We can see that though the attacker with limited resources can manipulate fewer measurements for (10,0)-tuple attack and (4,6)-tuple attack with full knowledge, they can have larger impact metrics comparing with the (11,0)-tuple attack and (5,6)-tuple attack with limited knowledge of the system. This implies the importance of the knowledge of system model. From the system operator's view, the system

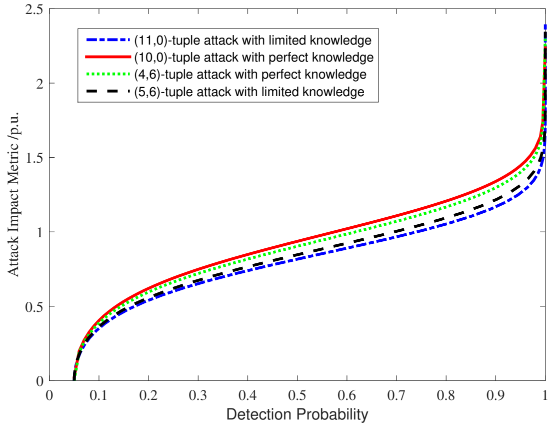

Figure 3. The attack impact metric is plotted versus the detection probability under the attacks with limited knowledge or resources respectively. For the limited knowledge scenario, error on line parameters of ±20% is employed. The false alarm rate is set to be $\alpha = 0.05$.

Table I
SPARSE OPTIMAL ATTACKS WITH LIMITED RESOURCES

Attacks	Index	$\mathbf{I}(a,d)$ /p.u. ($\bar{\delta} = 0.1$)	$\mathbf{I}(a,d)$ /p.u. ($\bar{\delta} = 0.2$)
(3,0)-tuple attack	(7,27,45)	0.0585	0.0898
(2,0)-tuple attack	(42,45)	0.0440	0.0676
(1,0)-tuple attack	(44)	0.0354	0.0544
(2,1)-tuple attack	(7,27,45)	0.0574	0.0881
(1,2)-tuple attack	(7,27,45)	0.0551	0.0847
(1,1)-tuple attack	(7,44)	0.0412	0.0633

model kept in the database of SCADA should be protected well. Besides this also can be used to implement mitigation schemes by misleading attacks to use perturbed or even faked system model thus making them detectable.

Then the scenario that the attacker can corrupt even fewer measurements is considered. Such (k_a, k_d)-tuple attack becomes a “sparse” attack. All of the possible attacked measurement sets can be examined and the worst case with largest attack impact metrics is selected. Table I gives the results of the “sparse” optimal attacks. With a given $\bar{\delta}$ and k_a, k_d, the optimal attacks with largest attack impact metrics are obtained. Here the measurements index denotes the measurement set manipulated by the attacker. Index 7 and 27 are line power flow measurements on branch 7 (from bus 4 to bus 5) and 42, 44, 45 are bus power injection measurements on bus 2, 4, 5. We can see that with more measurements corrupted by FDI attack, the maximum impact metrics on errors of load estimate can be larger. The results also indicate that the measurement set (e.g., line power flow measurements on branch 7 and injection measurements on bus 2,4,5) are vulnerable to the “sparse” attacks from the view of the system operation and has the priority of be equipped with mitigation schemes.

VI. CONCLUSION

In this paper, we consider more realistic attacks with limited adversarial knowledge and limited attack resources. The attack is also generalized to include both data integrity and data availability attacks. We show that the detection probability of attacks increase when the error on parameter estimation increase for the attack. The optimal attacks with limited attack resources but full knowledge can be more damaging than the ones with limited knowledge but enough attack resources according to the detection probability and attack impact metrics, which implies the importance of the knowledge of the system model to the attack. The results also suggest which measurements are more vulnerable to “sparse” data attacks that need to be protected with priority. Future work includes the computationally efficient algorithms, using various attack cost on different measurements, mitigation schemes, etc.

REFERENCES

[1] A. Teixeira, S. Amin, H. Sandberg, K. H. Johansson, and S. S. Sastry, “Cyber security analysis of state estimators in electric power systems,” in *Proc. 49th IEEE Conf. Decision and Control (CDC)*, Dec. 2010, pp. 5991–5998.

[2] T. M. Chen and S. Abu-Nimeh, “Lessons from stuxnet,” *Computer*, vol. 44, no. 4, pp. 91–93, 2011.

[3] J. Hong, Y. Chen, C.-C. Liu, and M. Govindarasu, “Cyber-physical security testbed for substations in a power grid,” in *Cyber Physical Systems Approach to Smart Electric Power Grid*. Springer, 2015, pp. 261–301.

[4] Y. Mo, T. H.-J. Kim, K. Brancik, D. Dickinson, H. Lee, A. Perrig, and B. Sinopoli, “Cyber–physical security of a smart grid infrastructure,” *Proceedings of the IEEE*, vol. 100, no. 1, pp. 195–209, 2012.

[5] G. Hug and J. A. Giampapa, “Vulnerability assessment of AC state estimation with respect to false data injection cyber-attacks,” *IEEE Transactions on Smart Grid*, vol. 3, no. 3, pp. 1362–1370, Sep. 2012.

[6] H. Sandberg, A. Teixeira, and K. H. Johansson, “On security indices for state estimators in power networks,” in *First Workshop on Secure Control Systems (SCS), Stockholm*, 2010.

[7] A. Teixeira, K. C. Sou, H. Sandberg, and K. H. Johansson, “Secure control systems: A quantitative risk management approach,” *IEEE Control Systems*, vol. 35, no. 1, pp. 24–45, 2015.

[8] 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, 2011.

[9] L. Xie, Y. Mo, and B. Sinopoli, “Integrity data attacks in power market operations,” *IEEE Transactions on Smart Grid*, vol. 2, no. 4, pp. 659–666, 2011.

[10] L. Jia, J. Kim, R. J. Thomas, and L. Tong, “Impact of data quality on real-time locational marginal price,” *IEEE Transactions on Power Systems*, vol. 29, no. 2, pp. 627–636, Mar. 2014.

[11] M. A. Rahman and H. Mohsenian-Rad, “False data injection attacks with incomplete information against smart power grids,” in *Global Communications Conference (GLOBECOM), 2012 IEEE*. IEEE, 2012, pp. 3153–3158.

[12] J. Zhang, Z. Chu, L. Sankar, and O. Kosut, “False data injection attacks on power system state estimation with limited information,” in *Power and Energy Society General Meeting (PESGM), 2016*. IEEE, 2016, pp. 1–5.

[13] K. Pan, A. M. H. Teixeira, M. Cvetkovic, and P. Palensky, “Combined data integrity and availability attacks on state estimation in cyber-physical power grids,” in *Proc. IEEE Int. Conf. Smart Grid Communications (SmartGridComm)*, Nov. 2016, pp. 271–277.

[14] A. Abur and A. G. Exposito, *Power system state estimation: theory and implementation*. CRC press, 2004.

[15] J. Liang, L. Sankar, and O. Kosut, “Vulnerability analysis and consequences of false data injection attack on power system state estimation,” *IEEE Transactions on Power Systems*, vol. 31, no. 5, pp. 3864–3872, Sep. 2016.

[16] K. Pan, A. Teixeira, M. Cvetkovic, and P. Palensky, “Cyber risk analysis of combined data attacks against power system state estimation,” unpublished.

[17] S. Boyd and L. El Ghaoui, “Method of centers for minimizing generalized eigenvalues,” *Linear algebra and its applications*, vol. 188, pp. 63–111, 1993.