On the Security of Distributed Power System State Estimation under Targeted Attacks

Ognjen Vuković
Laboratory for Communication Networks
School of Electrical Engineering
KTH Royal Institute of Technology, Sweden
vukovic@ee.kth.se

György Dán
Laboratory for Communication Networks
School of Electrical Engineering
KTH Royal Institute of Technology, Sweden
gyuri@ee.kth.se

ABSTRACT

State estimation plays an essential role in the monitoring and control of power transmission systems. In modern, highly inter-connected power systems the state estimation should be performed in a distributed fashion and requires information exchange between the control centers of directly connected systems. Motivated by recent reports on trojans targeting industrial control systems, in this paper we investigate how a single compromised control center can affect the outcome of distributed state estimation. We describe five attack strategies, and evaluate their impact on the IEEE 118 benchmark power system. We show that that even if the state estimation converges despite the attack, the estimate can have up to 30% of error, and bad data detection cannot locate the attack. We also show that if powerful enough, the attack can impede the convergence of the state estimation, and thus it can blind the system operators. Our results show that it is important to provide confidentiality for the measurement data in order to prevent the most powerful attacks. Finally, we discuss a possible way to detect and to mitigate these attacks.

Categories and Subject Descriptors

[**Security and privacy**]: Distributed systems security; [**Power and energy**]: Smart grid

Keywords

SCADA/EMS, distributed state estimation, security, data integrity attacks, inter-control center communication

1. INTRODUCTION

Power system operators rely on Supervisory Control and Data Acquisition (SCADA) systems integrated with Energy Management Systems (EMS) to efficiently and safely operate the power grid. The SCADA system collects measurement data from the substations that belong to the operator into a control center. The measurement data are processed at the control center by the EMS. A core component of the EMS is the state estimator (SE), which allows the operator to get an accurate estimate of the state of the power system despite noisy or faulty measurement data by using a steady-state model of the power flows in the physical system [14, 1]. The state estimate is used by various EMS applications, such as contingency analysis and security constrained economic dispatch, and thus an accurate state estimate is crucial both for the safety and for the efficiency of the power system's operation.

In order to improve operational efficiency, modern power systems have become increasingly inter-connected and are managed by several independent operators. Each operator has its own SCADA/EMS system and control center, which it uses to manage a region of the entire system. Examples of inter-connected systems are the Western Interconnect (WECC) in the U.S., the ENTSO-E in Europe, and some major European national transmission systems managed by various operators. In the future smart grid, interconnected systems are expected to become even more prevalent, and it is expected that their control and supervision becomes fully distributed, without any central coordinator. The safety of an interconnected power system depends on the safety of its constituent regions, as demonstrated by recent cascading failures, e.g., the 2003 North-East blackout in the U.S. It is therefore important that the regional operators exchange timely and accurate information about each other's networks state. Due to the sensitivity of the data, the information exchange is in practice very limited. Nevertheless, the exchanged information is used in the regional control centers as an input to the SE. The resulting fully distributed SE [16, 2, 12] are effectively extensions of the basic SE algorithm and aim to achieve a consistent state estimate for the entire power system.

Motivated by recent reports on trojans targeting industrial control systems, such as Stuxnet and Duqu [17], in this work we address the security of distributed state estimation in the presence of a misbehaving control center. We consider an attacker that compromises a single control center so that it can manipulate the data that the control center exchanges with its neighbors. We define various attack strategies that differ in the attacker's knowledge about the system. We show via simulations on an IEEE benchmark power system that attacks can disturb the distributed state estimation in two ways. First, the distributed state estimation could yield a highly erroneous state estimate (up to 30% relative estimation error), and second, the distributed state estimation could fail to provide any state estimate. Moreover, our results show that it is important to protect the confidentiality of measurement data, since the attacker needs those data to perform the strongest kinds of attacks. Finally, we show a possible way to detect convergence problems as a consequence of an attack by relying on a contraction mapping interpretation of distributed state estimation. This detection is a complement to traditional bad data detection (BDD) algorithms, which require the SE to converge.

SAC'13 March 18-22, 2013, Coimbra, Portugal.

Several recent works focused on the security of standalone SEs for the case of so called stealth attacks [13, 18, 3, 11, 6, 19, 10, 8]. Stealth attacks are false data injection attacks against SCADA measurement data that bypass the model-based bad data detector used in the SE. The possibility of such attacks was pointed out in [13], and different mitigation schemes were proposed in [3] based on protecting individual data, by changing the bad data detection algorithm [11], and by protecting components of the SCADA network infrastructure [6, 19]. The problem of maintaining operator privacy for distributed state estimation was addressed recently in an information theoretic framework in [15]. To the best of our knowledge we are the first to consider the vulnerability of distributed state estimators to data integrity attacks.

The rest of the paper is organized as follows. In Section 2 we describe the system model and give an outline of distributed state estimation algorithms. In Section 3 we describe the attacker model and define various strategies. Section 4 provides an impact analysis of the attack strategies. In Section 5 we consider a possible detection and mitigation strategy, and Section 6 concludes the paper.

2. SYSTEM MODEL

We consider an inter-connected power system that spans several administrative areas, called regions. We denote the set of buses by $\mathcal{B}$, $|\mathcal{B}| = B$, and the set of regions by $\mathcal{R}$. Each bus belongs to a region, and we denote the set of buses that belong to region $r \in \mathcal{R}$ by $\mathcal{B}_r$.

We say that a transmission line $t_{b,b'}$ that connects $b \in \mathcal{B}_r$ and $b' \in \mathcal{B}_{r'}$ is a *tie line* between two regions if $r \neq r'$. We say that $b \in \mathcal{B}_r$ is a border bus to region r' if there is a tie line $t_{b,b'}$ for some $b' \in \mathcal{B}_{r'}$. We denote the set of all tie lines connecting region r to region r' by $\mathcal{T}_{r,r'} = \{t_{b,b'} \mid b \in \mathcal{B}_r,\ b' \in \mathcal{B}_{r'}\}$. The set of all border buses of region r to region r' is denoted by $\mathcal{B}_{r,r'} = \{b \mid \exists t_{b,b'} \in \mathcal{T}_{r,r'}\}$ ($B_{r,r'} = |\mathcal{B}_{r,r'}|$). Similarly, the set of border buses from all regions to region r' is denoted by $\mathcal{B}_{b,r'} = \cup_r \mathcal{B}_{r,r'}$ ($B_{b,r'} = |\mathcal{B}_{b,r'}|$). Finally, we say that two regions are *neighbors* if they share at least one tie line. We denote the set of neighbors of region r by $\mathcal{N}(r)$ ($N(r) = |\mathcal{N}(r)|$).

2.1 State Estimation

We consider models of the active and reactive power injections at every bus, and models of the active and reactive power flows between buses (over transmission lines) [14, 1]. The power flow and injection measurement values are denoted by the vector $z \in \mathbb{R}^M$, where M is the number of measurements. The value of a measurement m equals to $z_m = P_m + e_m$, where P_m is the actual power flow or injection (active or reactive) and e_m is independent random measurement noise. The noise is usually assumed to have a Gaussian distribution of zero mean, $e = (e_1, e_2, ..., e_M)^T \in N(0; R)$ where $W = Eee^T$ is the diagonal measurement covariance matrix.

The state-estimation problem consists of estimating B voltage phasor vectors, $\mathcal{V}_b = V_b e^{j\theta_b}\ \forall b \in \mathcal{B}$, given the power flow and injection measurement vector z. One (arbitrary) voltage phasor can be selected as the reference phasor, for example $\mathcal{V}_B = 1e^{j0}$, and then only $n = B - 1$ phasors have to be estimated. We denote by x, the *state vector*, which consists of the voltage phasor angles and magnitudes, i.e., $x = [\theta_1, V_1, \theta_2, V_2, ..., \theta_n, V_n]^T$, where θ_i and V_i are phase angle and voltage magnitude on bus b_i, respectively. We refer to a component of the vector x as a *state variable*.

The most widely used approach to solve the estimation problem is to minimize the squares of the weighted deviations of the estimated variables from the actual measurements [1], which can be formulated as

$$\min_x J(x) = \min_x [z - f(x)]^T [W^{-1}][z - f(x)], \tag{1}$$

where $f(x)$ is the vector of functions describing the measurements as a function of the state vector x. Since f is non-linear, the estimation is typically done using an iterative solution scheme known as the Gauss-Newton algorithm [1]. The recurrence relation of this iterative solution scheme is

$$x^{(k+1)} = x^{(k)} + \Delta x^{(k)}, \tag{2}$$

and the increment $\Delta x^{(k)}$ can be calculated as

$$\Delta x^{(k)} = [H^{(k)T} W^{-1} H^{(k)}]^{-1} H^{(k)T} W^{-1} \Delta z^{(k)}, \tag{3}$$

where $H^{(k)}$ is the Jacobian of vector $f(x^{(k)})$, $\Delta z^{(k)}$ is the measurement residual vector defined as $\Delta z^{(k)} = z - f(x^{(k)})$, and $x^{(k)}$ is the value of vector x at the k^{th} iteration. The algorithm is said to converge when for some k^* the maximum update of the state variables is smaller than the *convergence threshold* $\varepsilon > 0$, i.e., $||\Delta x^{(k^*)}||_\infty < \varepsilon$, where $||\cdot||_\infty$ denotes the maximum norm of a vector. We refer to the number of iterations k^* required for convergence as the *convergence time*.

Once the state estimator converges, a Bad Data Detection (BDD) algorithm is used to detect and identify faulty measurement data. The BDD algorithm analyses the measurement residual vector ($\Delta z^{(k^*)}$). The most widely used BDD algorithm is the *Largest Normalized Residual Test (LNRT)*. The LNRT suspects the measurement with highest normalized residual, i.e., the largest value of the measurement residual vector divided by its Euclidean norm ($\Delta z^{(k^*)} / ||\Delta z^{(k^*)}||_2$), as bad data, if the ratio is above a certain threshold. For a more complete treatment of BDD we refer to [14, 1].

2.2 Distributed State Estimation (DSE)

In an inter-connected power system each regional control center performs the state estimation based the topology and the parameters of the region, and based on the measurements taken in the region. Therefore, the state estimation problem in region r becomes a problem of estimating the voltage phasor vectors for the buses $b \in \mathcal{B}_r$, i.e., the state vector x_r. However, the power flow measurements on the tie lines $\mathcal{T}_{r,r'}$ ($r' \in \mathcal{N}(r)$), which we refer to as the *boundary* measurements, are a function of the state variables of the neighboring regions r' as well. Hence, the control center of region r needs to exchange a few state variables with the control centers of its neighboring regions. These state variables correspond to the buses at the two ends of the tie lines; the control center of region r sends the state variables for the buses in $\mathcal{B}_{r,r'}$ to the control center of region r'. In most of the recently proposed DSE algorithms, e.g., [16, 2, 12], state variables are exchanged at the beginning of every iteration. For the purpose of our study, we consider the algorithm described in [16].

We denote the vector of state variables communicated by region r to region r' (r' to r) at iteration k by $x^{(k)}_{r,r'}$ ($x^{(k)}_{r',r}$), and define it as

$$x^{(k)}_{r,r'} = [\theta^{(k)}_{i_1}\ V^{(k)}_{i_1}\ \theta^{(k)}_{i_2}\ V^{(k)}_{i_2}\ ...\ ]^T,\ \forall b_{i_j} \in \mathcal{B}_{r,r'}. \tag{4}$$

We denote the vector of state variables that region r receives from its neighbors at iteration k by

$$x^{(k)}_{b,r} = [x^{(k)T}_{r'_{i_1},r}\ x^{(k)T}_{r'_{i_2},r}\ ...\ ]^T,\ \forall r'_{i_j} \in \mathcal{N}(r).$$

The state estimator of region r uses $x^{(k)}_{b,r}$ to iteratively estimate x_r similar to (2) and (3), but the Jacobian and the measurement residual vector are calculated as

$$H^{(k)} = \left[\frac{\partial f(y^{(k)}_r)}{\partial x^{(k)}_r}\right] \quad \Delta z^{(k)} = z - f(y^{(k)}_r), \tag{5}$$

where $y_r^{(k)} = [x_r^{(k)T}\ x_{b,r}^{(k)T}]^T$ is the state vector extended with the boundary state variables received at the beginning of the current iteration, i.e., iteration k. The DSE is said to converge when all regional state estimators converge. If we denote by k_r^* the convergence time of region r, then the *total convergence time* is $c = \max_r(k_r^*)$.

3. ATTACK SCENARIO

DSE requires that neighboring control centers periodically exchange data with each other. The most widely used protocol for this purpose is the standardized Inter-Control Center Communications Protocol (ICCP or IEC 60870-6/TASE.2). ICCP defines data structures and encodings, and allows control centers to establish so called associations on a pairwise basis. An association allows bidirectional data exchange between two control centers. Using ICCP it is possible to implement access control, but ICCP provides no means for key-based authentication of the data sent.

The standard way of providing authentication for ICCP associations is to rely on the authentication provided by standard transport layer protocols, such as TLS and SSL [7], as mandated by IEC 62351. As an effect, ICCP messages might be passed in clear text to the TCP/IP protocol stack or to standard libraries providing authentication. An attacker that compromises the operating system and the TCP/IP protocol stack in a control center, e.g., by installing a trojan, can thus easily manipulate all incoming and outgoing ICCP messages at the compromised control center. The vulnerability of control systems to such an attack is aggravated by the fact that ICCP associations are often established between hosts in demilitarized zones.

3.1 Attack Model

We consider an attacker whose goal is to introduce disturbances in DSE. In order to achieve its goal, the attacker corrupts the control center of a single region $r^a \in \mathcal{R}$ so that it has access to the state variables exchanged between region r^a and its neighbors $\mathcal{N}(r^a)$ at the beginning of every DSE iteration. At iteration k, the state variables are elements of the vectors $x_{r,r^a}^{(k)}$, $\forall r \in \mathcal{N}(r^a)$, and the vectors $x_{r^a,r}^{(k)}$, $\forall r \in \mathcal{N}(r^a)$. In principle, the attacker can tamper with the entire vectors, but the relative differences in voltage magnitudes between neighboring buses are rather small and their manipulation may be easy to detect. Therefore, we focus on an attacker that tampers with the exchanged state variables that correspond to the phase angles. We describe the attack against the state variables sent from regions $r \in \mathcal{N}(r^a)$ to region r^a (from r^a to r) at the beginning of iteration k by the *attack vector* $a_{r,r^a}^{(k)}$ ($a_{r^a,r}^{(k)}$). We define the attack vector $a_{r,r^a}^{(k)}$ as the vector of phase angles

$$a_{r,r^a}^{(k)} = [\hat{\theta}_{i_1}^{(k)}\ \hat{\theta}_{i_2}^{(k)}\ \ldots]^T\ \ \forall b_{i_j} \in \mathcal{B}_{r,r^a}, \tag{6}$$

where element $\hat{\theta}_{i_j}^{(k)}$ corresponds to the value that the attacker adds to the phase angle $\theta_{i_j}^{(k)}$ that it wants to modify. The attack vector $a_{r,r^a}^{(k)}$ can be defined in a similar way. In the rest of this Section, we describe the attack against the state variables sent to region r^a from its neighbors $r \in \mathcal{N}(r^a)$. The attack against the state variables sent from region r^a to its neighbors can be described in a similar way, but we omit it for brevity.

Since the attack is additive and it concerns the phase angles of the exchanged vector of state variables $x_{r,r^a}^{(k)}$, it results in a corrupted

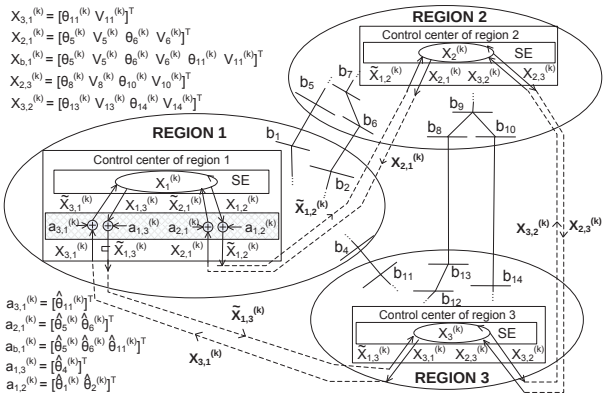

Figure 1: Interconnected power system with three regions. The attacker corrupts the control center of Region 1, and tampers with the state variables $x_{1,2}^{(k)}$ and $x_{1,3}^{(k)}$ sent from Region 1, and the state variables $x_{2,1}^{(k)}$ and $x_{3,1}^{(k)}$ received by Region 1. The symbol (+) indicates that the components of the attack vector are added to the corresponding components (phase angles) of the vector of exchanged state variables. The attacker cannot tamper with the state variables exchanged between Regions 2 and 3.

vector of state variables

$$\tilde{x}_{r,r^a}^{(k)} = x_{r,r^a}^{(k)} + Q_{r,r^a} \cdot a_{r,r^a}^{(k)}, \tag{7}$$

where $Q_{r,r^a} = [q_{i,j}]_{2\cdot B_{r,r^a} \times B_{r,r^a}}$ is a matrix used to insert the components that correspond to voltage magnitudes with values equal to 0. The elements of matrix Q_{r,r^a} are defined as: $q_{i,j} = 1$ if $\lceil j = i/2 \rceil$ and $i \bmod 2 = 1$, and $q_{i,j} = 0$ otherwise. The resulting vector $\tilde{x}_{r,r^a}^{(k)}$ is used as an input to the iteration k of DSE in region r^a, instead of the originally exchanged vector $x_{r,r^a}^{(k)}$.

For convenience, we introduce the attack vector $a_{b,r^a}^{(k)}$ for the state variables sent to region r^a from all its neighboring regions

$$a_{b,r^a}^{(k)} = [a_{r_{i_1},r^a}^{(k)T}\ a_{r_{i_2},r^a}^{(k)T}\ \ldots]^T\ \ \forall r_{i_j} \in \mathcal{N}(r^a), \tag{8}$$

and the corresponding corrupted vector of state variables

$$\tilde{x}_{b,r^a}^{(k)} = x_{b,r^a}^{(k)} + Q_{r^a} \cdot a_{b,r^a}^{(k)}, \tag{9}$$

where $Q_{r^a} = [q_{i,j}]_{2\cdot B_{b,r^a} \times B_{b,r^a}}$ is a matrix with the same structure as Q_{r,r^a}. Fig. 1 illustrates an attack on a power system with three regions. Observe that $\tilde{x}_{b,r^a}^{(k)}$ is the input to iteration k of DSE, and thus, the attack $a_{b,r^a}^{(k)}$ leads to a *corrupted* state vector update $\Delta\tilde{x}_{r^a}^{(k)}$.

We define *the size of the attack* as the Euclidean norm of the attack vector, i.e., $||a_{b,r^a}^{(k)}||_2$. We consider that the goal of the attacker is to find an attack vector with a small size but with a big impact on the convergence time c of the distributed state estimator, or formally

$$\max_{a_{b,r^a}^{(k)}, k=1,\ldots} \quad c \qquad s.t.\ ||a_{b,r^a}^{(k)}||_2 \leq \beta\ \ \forall k, \tag{10}$$

where $\beta > 0$ is the desired bound on the attack size. By definition, $c = \infty$ if the DSE does not converge.

3.2 Attack Strategies

Since the distributed state estimation problem is non-linear, solving (10) is non-trivial. In the following we describe five strategies to construct the attack vector $a_{b,r^a}^{(k)}$.

3.2.1 Maximal Update Vector Attack (MUV)

The MUV attack is an approximation of (10) done by maximizing the Euclidean norm of the corrupted state vector update in every iteration,

$$\max_{a_{b,r^a}^{(k)}} ||\Delta\tilde{x}_r^{(k)}||_2 \;\; s.t. \;\; ||a_{b,r^a}^{(k)}||_2 = \beta. \tag{11}$$

Recall that $\Delta\tilde{x}_r^{(k)}$ depends on $a_{b,r^a}^{(k)}$ through (3) and (5). The objective function and the constraints in (11) are quadratic functions, and therefore the vector $a_{b,r^a}^{(k)}$ can be obtained by solving a quadratically constrained quadratic program [4]. Observe that the attacker cannot solve (11) without knowing the entire state vector $x_r^{(k)}$ and the measurement vector z, but the vectors $x_r^{(k)}$ and z are not exchanged between the regions. We therefore use the MUV attack as a baseline for comparison.

3.2.2 First Singular Vector Attack (FSV)

The FSV attack also aims to solve (11) but in the cases when the vectors $x_r^{(k)}$ and z may be unknown to the attacker. We denote by $x_r^{a(k)}$ the attacker's knowledge of the vector $x_r^{(k)}$ at iteration k. Correspondingly, we denote by $x_{b,r}^{a(k)}$ and by $y_r^{a(k)}$ the attacker's knowledge of the vectors $x_{b,r}^{(k)}$ and $y_r^{(k)}$, respectively. In order to approximate (11), we linearize the function $f(y_r^{(k)})$ at $y_r^{a(k)}$ so that for the measurement residual vector $\Delta\tilde{z}^{(k)}$ we obtain

$$\begin{aligned}\Delta\tilde{z}^{(k)} &\approx z - (f(\begin{bmatrix} x_r^{a(k)} \\ x_{b,r}^{a(k)} \end{bmatrix}) + [H^{a(k)} H_b^{a(k)}] \begin{bmatrix} \mathbf{0} \\ Q_{r^a} \cdot a_{b,r^a}^{(k)} \end{bmatrix}) \\ &\approx \Delta z^{(k)} - [H^{a(k)} H_b^{a(k)}] \begin{bmatrix} \mathbf{0} \\ Q_{r^a} \cdot a_{b,r^a}^{(k)} \end{bmatrix} \approx \Delta z^{(k)} - H_b^{a(k)} \cdot Q_{r^a} \cdot a_{b,r^a}^{(k)},\end{aligned} \tag{12}$$

where $H^{a(k)}$ and $H_b^{a(k)}$ are the Jacobian matrices of $f(y_r^{(k)})$ evaluated at $x_r^{a(k)}$ and $x_{b,r}^{a(k)}$, respectively. After substituting (12) into (3), the corrupted vector $\Delta\tilde{x}_r^{(k)}$ can be approximated as

$$\Delta\tilde{x}_r^{(k)} = \Delta x_r^{(k)} - [H^{a(k)\,T} W^{-1} H^{a(k)}]^{-1} H^{a(k)\,T} W^{-1} H_b^{a(k)} \cdot Q_{r^a} \cdot a_{b,r^a}^{(k)}. \tag{13}$$

Observe that the subtrahend in (13) is a vector with the same number of elements as the vector $\Delta x_r^{(k)}$, and we refer to it as the *subtrahend vector*. The Euclidean norm of the subtrahend vector is maximized if the attack vector $a_{b,r^a}^{(k)}$ is aligned with the first right singular vector of the matrix $[H^{a(k)\,T} W^{-1} H^{a(k)}]^{-1} H^{a(k)\,T} W^{-1} H_b^{a(k)} \cdot Q_{r^a}$, that is, with the singular vector with highest singular value. The complexity of singular vector decomposition is $O(mn^2)$ [9], low enough for the computation to be done on-line.

Observe in (13) that size of the corrupted vector $\Delta\tilde{x}_r^{(k)}$ depends on the direction of the subtrahend vector, and consequently, on the direction of the first singular vector. Whether the attacker will choose the correct direction of the first singular vector depends on its knowledge of the state vector $x_r^{(k)}$, and on the measurement vector z. We consider two variants of the FSV attack.

FSV with State Information (FSV+ST): The FSV+ST attack assumes that the attacker knows the state vector $x_r^{(k)}$, but it does not know the measurement vector z and the correct direction. Consequently, $x_r^{a(k)} = x_r^{(k)}$ and $y_r^{a(k)} = y_r^{(k)}$ in (12) and (13). Since the attacker does not know the vector z, and thereby the update vector $\Delta x_r^{(k)}$ without attack, finding the correct direction is not trivial. In order to estimate the direction, we assume that the estimates of the active and reactive power flows on a tie line are closer to their actual values when using the most recent exchanged state variables. The attacker may tamper with the exchanged state variables such that the introduced estimation errors take the estimates closer to the estimates from the previous round. The direction which satisfies this for more tie line power flows is chosen by the attacker.

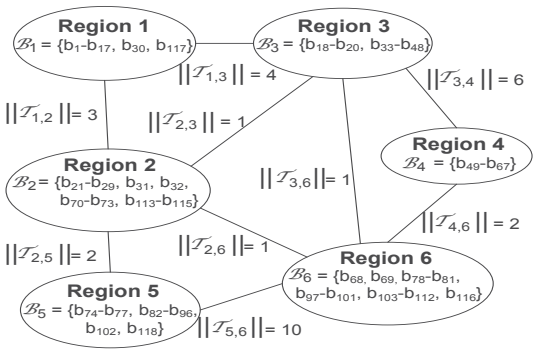

Figure 2: IEEE 118 bus system divided into six regions. Neighboring regions are connected by a line, $||\mathcal{T}_{r,r'}||$ is the number of tie-lines. The buses $\mathcal{B}_r$ are shown for each region.

FSV with Measurement Information (FSV+MEAS): The FSV+MEAS attack assumes that the attacker does not know the state vector $x_r^{(k)}$, but it knows the measurement vector z. Consequently, $x_r^{a(k)} = x_r^{(1)}$ and $y_r^{a(k)} = y_r^{(1)}$ in (12) and (13). The update vector $\Delta x_r^{(k)}$, and thereby the correct direction, is not known by the attacker. In order to estimate the direction, we use a similar approach as for the FSV+ST attack, but the attacker uses the actual measurements, rather than two estimates, when choosing the direction.

3.2.3 Uniform Rotation Attack (UR)

The third strategy we consider is rather naive. The attack vector rotates all voltage phasors by a constant ϕ, thus

$$a_{b,r^a}^{(k)} = \phi \cdot \mathbf{1}, \tag{14}$$

where $\mathbf{1}$ is the column vector of all ones of dimension B_{b,r^a}. The size of the attack is $||a_{b,r^a}^{(k)}||_2 = \phi \cdot \sqrt{B_{b,r^a}}$.

3.2.4 Sign Inversion Attack (SI)

The fourth strategy we consider is adaptive, similar to the FSV attack. The attack only requires knowledge of the exchanged state variables, and at every round it inverts the sign of exchanged phase angles,

$$a_{b,r^a}^{(k)} = [-2\theta_{i_1}^{(k)} \;\; -2\theta_{i_2}^{(k)} \;\; ...\;] \;\; \forall b_{i_j} \in \mathcal{B}_{b,r^a}. \tag{15}$$

The size of the attack depends on the system state.

3.2.5 Sign of Difference Inversion Attack (SDI)

The last strategy is based on the insight that the steady state active power flow on a tie line is an odd function of the phase angle difference between the border buses [1],

$$a_{b,r^a}^{(k)} = [-2(\theta_{i_1}^{(k)} - \theta_{i_1'}^{(k)}) \;\; ...\;] \forall b_{i_j} \in \mathcal{B}_{b,r^a} \text{ and } t_{b_i,b_i'} \in \mathcal{T}_{b,r^a}. \tag{16}$$

The attack effectively inverts the sign of the phase angle differences for every tie line, which corresponds to reverting the power flow on every tie line of region r^a. Again, the size of the attack depends on the system state.

4. ATTACK IMPACT

In the following we evaluate the impact of the attack strategies on the IEEE 118 bus power system. The power system is divided into six regions as shown in Fig. 2. We consider that the attacker corrupts the control center of region r_1, and performs the attacks

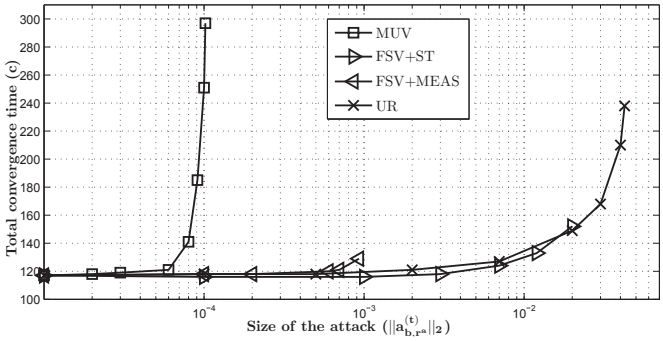

Figure 3: Total convergence time for cases when the DSE converges as a function of the attack size.

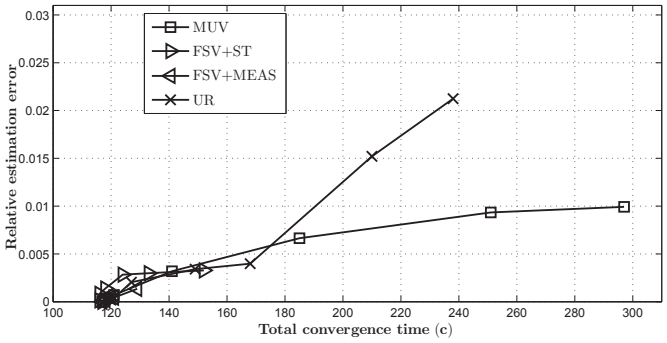

Figure 4: Relative estimation error (50th percentile) for the upper 50% utilized power flows and injections vs. total convergence time

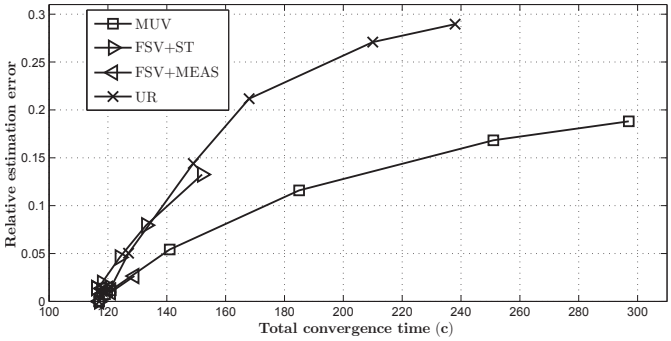

Figure 5: Relative estimation error (maximum) for the upper 10% utilized power flows and injections vs. total convergence time

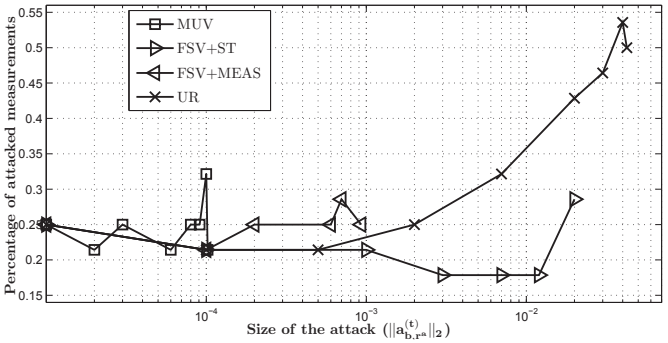

Figure 6: Percentage of the border measurements around the attacked region that are among top candidates for bad data vs. size of the attack

against the state variables sent from and to region r_1. Measurements are taken at every power injection and power flow (both active and reactive), and the convergence threshold is $\varepsilon = 10^{-3}$. Fig. 3 shows the total convergence time c as a function of the attack size for the DSE under the MUV, the FSV (both variants), and the UR attacks. The total convergence time increases monotonically with the attack size for all considered attacks. The MUV attack is the most powerful among the considered attacks: the increase of the convergence time is significantly higher for the same attack size, and the DSE stops converging for a much lower attack size. The results show that the FSV+MEAS attack is significantly more powerful than the FSV+ST attack. Therefore, it is important to prevent the attacker from obtaining the measurement data, e.g., by not exchanging the data between neighboring regions and by encrypting the data when transmitting them from the substations to the control center.

Although for small attacks the DSE converges, the estimated state and thus the estimated power flows could be erroneous. Fig. 4 and Fig. 5 show the 50th percentile and the maximum of the relative estimation error for the highest 50% and for the highest 10% of the power flows, respectively as a function of total convergence time (and thus the attack size). The relative estimation error increases monotonically with the total convergence time, and thereby the attack size, and can exceed 25% for some large power flows, which is a significant estimation error that can affect the outcome of EMS applications like contingency analysis.

In principle, the BDD algorithm should identify the measurements whose estimates significantly differ from the measured values (e.g., due to the attack) as bad data, and should thus detect the attack. In the following we use the centralized Largest Normalized Residual Test algorithm [16] for BDD to evaluate the efficiency of bad data detection under the considered attacks. We use a centralized BDD, because a centralized BDD is typically more efficient in identifying bad data than the fully distributed algorithms, e.g., [5]. We thus consider the strongest BDD possible. We focus on attacks that allow the DSE to converge, as the BDD cannot be performed if the DSE does not converge.

Since the attack concerns the power flow estimates at the tie lines connecting the attacked region with its neighbors, one would expect that the border measurements around the attacked region get identified by the BDD algorithm as bad data. If this was the case then by discarding those measurements, the BDD would isolate the rest of the system from the attacked region. However, this is not the case. Fig. 6 shows the percentage of the border measurements around the attacked region that are identified by the BDD algorithm as the top candidates for bad data as a function of the attack size for the MUV, the FSV (both variants), and the UR attacks. The percentage does not increase monotonically, and it is fairly constant even for strong attacks that cause significant estimation errors. Moreover, the percentage is relatively low for all attacks. This implies that the BDD algorithm may be misled: it can discard measurements in/between non-attacked regions, and does not locate the source of the attack.

Fig. 7 shows the maximum state vector update $||\Delta x||_\infty$ for the FSV+ST, the UR, the SI, and the SDI attacks in each iteration k. In order to make the results comparable, we scaled the FSV+ST and the UR attacks such that their attack size equals to the attack size of the SI attack in every iteration. Under the SI attack the DSE almost converges, but all attack strategies prevent the DSE

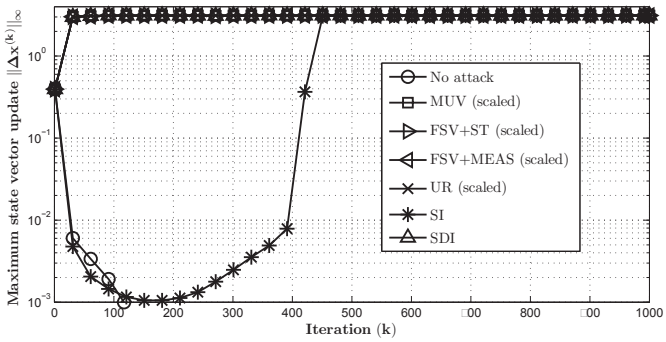

Figure 7: Evolution of the max. value of the state vector update in the entire system ($||\Delta x^{(k)}||_\infty$) with and without data integrity attacks in region r_1.

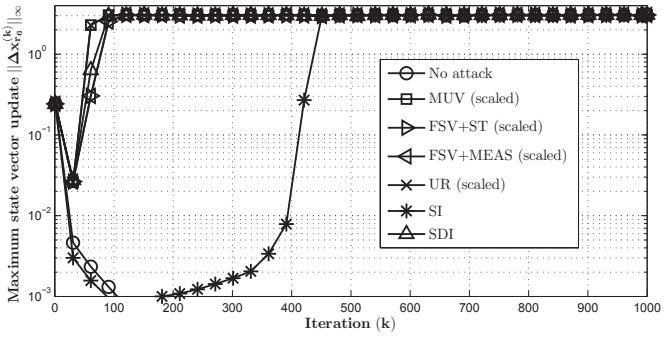

Figure 8: Evolution of the max. value of the state vector update in region r_6 ($||\Delta x_{r_6}^{(k)}||_\infty$) with and without data integrity attacks in region r_1.

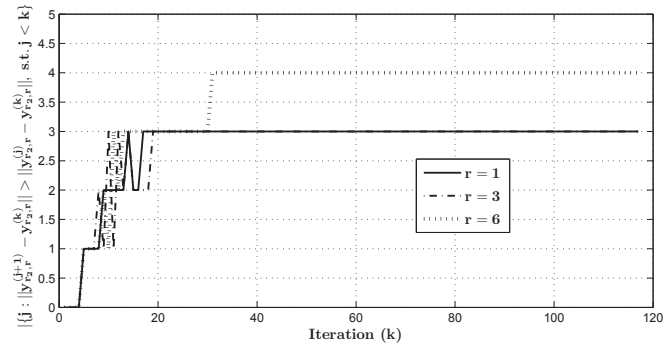

Figure 9: Evolution of the number of outlier state estimates based on $y_{r',r}^{(k)}$ in region $r' = 2$ vs. the number of rounds. No attack.

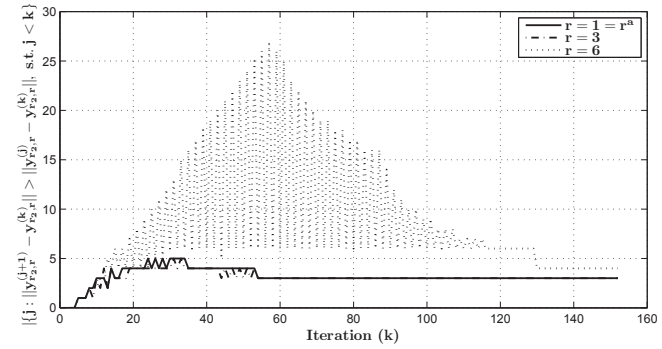

Figure 10: Evolution of the number of outlier state estimates based on $y_{r',r}^{(k)}$ in region $r' = 2$ vs. the number of rounds. FSV+MEAS attack at region $r^a = 1$ that admits convergence.

to converge eventually. One may assume that the DSE does not converge mainly due to the state vector updates in the corrupted region (r_1) and its neighbors, but Fig. 8 shows that this is not the case. Fig. 8 shows the maximum state vector update $||\Delta x_{r_6}||_\infty$ in the non-neighboring region r_6. While $||\Delta x_{r_6}||_\infty$ decreases initially for all attacks, it eventually starts increasing and diverges from the convergence threshold due to the resulting disturbances in r_1, r_2, and r_3 that propagate to the rest of the system through the state variables that are communicated. It is interesting that in the case of the SI attack the state estimator in region r_6 first converges, but not the DSE since at least one of the other regions has not converged yet, and as an effect $||\Delta x_{r_6}||_\infty$ starts increasing.

5. DETECTION AND MITIGATION

In the following we discuss a possible way for detecting an attack against the DSE. For the detection, let us first consider the evolution of the state vector without the data integrity attack. Observe that the evolution of the state vector in the DSE can be written as a recurrence relation $x^{(k+1)} = g(x^{(k)})$ for some non-linear mapping $g : \mathbb{R}^n \to \mathbb{R}^n$. Furthermore, when the DSE converges after k^* iterations, it holds that $x^{(k^*)} = g(x^{(k^*-1)}) \approx x^{(k^*-1)}$. In order for the DSE to converge, the mapping g has to satisfy certain conditions. One example is the sufficient condition formulated in [16, Proposition 5.2., Theorem 7.5.], which provides some insight into the behavior of the recurrence relation defined by g. The following proposition summarizes the condition.

Proposition 1. *If the iterative non-linear mapping function $g : \mathbb{R}^n \to \mathbb{R}^n$ is non-expansive in the Euclidean norm, then the set X^* of its fixed points is non-empty. If it satisfies*

$$||g(x) - x^*||_\infty \leq ||x - x^*||_\infty, \forall x \in \mathbb{R}^n, \forall x^* \in X^*, \qquad (17)$$

then the solution sequence $x^{(k)}$ converges to a fixed point x^*.

The above result does not imply that the subsequent state vector updates $\Delta x^{(k)}$ would form a non-increasing sequence in the max norm, i.e., $||g(x^{(k+1)}) - g(x^{(k)})||_\infty \not\leq ||g(x^{(k)}) - g(x^{(k-1)})||_\infty$. Furthermore, the set of fixed points X^* is not known. Nevertheless, for large values of k we can use the approximation that the estimate $x^{(k)}$ is close to a fixed point of g, and thus for large k and $k' < k$ we have

$$||x^{(k'+1)} - x^{(k)}||_\infty \leq ||x^{(k')} - x^{(k)}||_\infty \qquad (18)$$

assuming that the state estimator converges. In other words, when the state estimator is close to convergence to a fixed point, the distance of the points on the trajectory of convergence from the current estimate is a non-increasing function of the iteration k'. In the case of DSE the regional control centers only have access to their own state vector $x_r^{(k)}$ and to the last received state variables $x_{b,r}^{(k)}$ from their neighbors, i.e., to the vector $y_r^{(k)}$, and thus (18) has to be verified on these data. In the following we investigate how well (18) indicates convergence problems based on this data.

Fig. 9 shows for every iteration k the number of previous iterations j for which (18) does not hold for the vector $y_{r',r}^{(k)} = [x_{r'}^{(k)T} \; x_{r,r'}^{(k)}]^T$

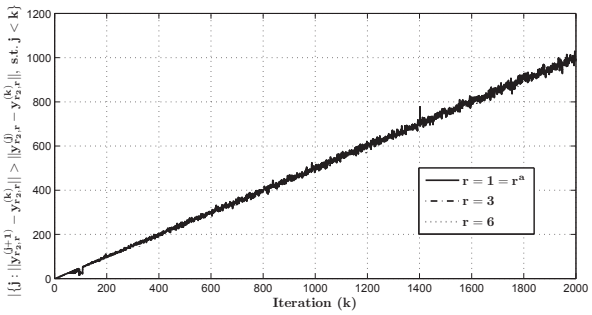

Figure 11: Evolution of the number of outlier state estimates based on $y^{(k)}_{r',r}$ in region $r' = 2$ vs. the number of rounds. FSV+MEAS attack at region $r^a = 1$ that does not admit convergence.

for region $r' = 2$ for three of its neighbors. The results are for the case without any attack. The results confirm that the number of outliers is small when the DSE converges, and also shows that the few outliers occur during the early iterations of the DSE. Fig. 10 shows results for a small FSV+MEAS attack that allows the DSE to converge, though with an estimation error (c.f. Figs 4 and 5). Initially, the number of outliers increases with the number of iterations, but it decreases as the DSE gets closer to convergence. Surprisingly, most outliers are detected based on $y^{(k)}_{2,6}$, although region 6 is not a neighbor of the attacked region ($r^a = 1$). Finally, Fig. 11 shows results for a FSV+MEAS attack that does not allow the DSE to converge. The number of outliers increases linearly with the number of iterations, and indicates the convergence problem immediately.

Fig. 10 and Fig. 11 show that outliers can be used to detect convergence problems due to, e.g., an attack. However, Fig. 11 also shows that localizing the point of the attack is not possible. One possible mitigation scheme could then be to disable the DSE, and let every region perform a local state estimation. Although power injections at border buses and the power flows on the tie lines cannot be estimated in this case, the resulting estimate is not affected by the attack.

6. CONCLUSION

We considered the vulnerability of distributed state estimation to targeted attacks against the exchanged data between operators. We described five attack strategies, and showed via simulations on an IEEE benchmark power system the effects of the attacks. The presented results led us to the following interesting conclusions. First, already a single compromised control center can cause convergence problems to the distributed state estimator. For small attacks the estimator converges but with significant errors, and the BDD algorithm cannot detect the attack location. For large attacks the estimator fails to converge and to provide a consistent state estimate. Second, it is important to protect the confidentiality of measurement data, since the attacker can perform strong attacks only if it knows the measurement data. Finally, the attacks could be detected by observing the number of outlier state estimates. Based on this detection scheme, we outlined a simple mitigation scheme. It is subject of our future work to extend the detection scheme such that it can localize the point of the attack, which could lead to an improved mitigation scheme.

7. REFERENCES

[1] A. Abur and A. G. Exposito. *Power System State Estimation: Theory and Implementation*. Marcel Dekker, Inc., 2004.

[2] S. d. T. Antonio J. Conejo and M. Canas. An optimization approach to multiarea state estimation. *IEEE Transactions on Power Systems*, 22(1), February 2007.

[3] R. B. Bobba, K. M. Rogers, Q. Wang, H. Khurana, K. Nahrstedt, and T. J. Overbye. Detecting false data injection attacks on dc state estimation. In *Preprints of the First Workshop on Secure Control Systems, CPSWEEK*, 2010.

[4] S. Boyd and V. Lieven. *Convex Optimization*. Cambridge University Press, 2004.

[5] D.-H. Choi and L. Xie. Fully distributed bad data processing for wide area state estimation. In *Proc. of IEEE SmartGridComm*, October 2011.

[6] G. Dán and H. Sandberg. Stealth attacks and protection schemes for state estimators in power systems. In *Proc. of IEEE SmartGridComm*, Oct. 2010.

[7] T. Dierks and E. Rescorla. RFC5246: The transport layer security (TLS) protocol version 1.2. http://tools.ietf.org/html/rfc5246, August 2008.

[8] A. Giani, E. Bitar, M. Garcia, M. McQueen, P. Khargonekar, and K. Poolla. Smart grid data integrity attacks: Characterizations and countermeasures. In *Proc. of IEEE SmartGridComm*, Oct. 2011.

[9] R. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 1990.

[10] T. T. Kim and H. V. Poor. Strategic protection against data injection attacks on power grids. *IEEE Trans. on Smart Grid*, 2:326–333, Jun. 2011.

[11] O. Kosut, L. Jia, R. Thomas, and L. Tong. Malicious data attacks on smart grid state estimation: Attack strategies and countermeasures. In *Proc. of IEEE SmartGridComm*, Oct. 2010.

[12] S. K. Le Xie, Dae-Hyun Choi and H. V. Poor. Fully distributed state estimation for wide-area monitoring systems. *IEEE Transactions on Smart Grid*, 3(3), 2012.

[13] Y. Liu, P. Ning, and M. Reiter. False data injection attacks against state estimation in electric power grids. In *Proc. of the 16th ACM conference on Computer and Communications Security (CCS)*, pages 21–32, 2009.

[14] A. Monticelli. Electric power system state estimation. *Proc. of the IEEE*, 88(2):262–282, 2000.

[15] L. Sankar, S. Kar, R. Tandon, and H. V. Poor. Competitive privacy in the smart grid: An information-theoretic approach. In *Proc. of IEEE SmartGridComm*, Oct. 2011.

[16] M. Shahidehpour and Y. Wang. *Communication and Control in Electric Power Systems*. John Wiley and Sons, 2003.

[17] Symantec Security Response. W32.duq: The precursor to the next stuxnet, November 2011.

[18] A. Teixeira, G. Dán, H. Sandberg, and K. H. Johansson. A cyber security study of a SCADA energy management system: Stealthy deception attacks on the state estimator. In *Proc. IFAC World Congress*, 2011.

[19] O. Vuković, K. C. Sou, G. Dán, and H. Sandberg. Network-aware mitigation of data integrity attacks on power system state estimation. *IEEE JSAC: Smart Grid Communications Series*, 30(6), 2012.