Abstract—We describe stochastic defense mechanisms designed to detect sophisticated grid attacks involving both physical actions (including load modification) and sensor output alteration. The attacks are undetectable under a full AC power flow model even assuming ubiquitous sensor placement, while hiding large line overloads. We first show that such attacks can be computed in the case of large transmission systems. The defensive techniques apply network control actions that change voltages in a random fashion in particular by introducing (random) low-rank corrections to covariance matrices.

Index Terms—Security, cyberphysical power grid attacks.

I. INTRODUCTION

Recent events and research efforts have highlighted the potential for powerful coordinated attacks on power grids that combine disruption or modification of sensor data with physical actions. Such attacks may succeed in hiding from operators undesirable system conditions, long enough that physical damage or automatic shutdown of equipment takes place, an undesirable and potentially risky outcome.

We propose defensive techniques to be deployed when a high-fidelity attack on a power grid is suspected. The attack is assumed to be partial in the sense that only a subset of buses and lines are attacked, but this subset is unknown. These techniques involve two ideas:

(a) using network resources to randomly change power flow quantities, especially voltages and, in particular

(b) changing the covariance structure of e.g. voltages in a manner unpredictable by the attacker. The specific version of this idea that we analyze introduces a low-rank adjustment to the covariance of phase angles.

These defensive techniques focus on the phase immediately following the initial attack, and aim to expose inconsistencies in the data dynamically output by the attacker. We describe conditions under which the defense succeeds in discovering the boundary of the attacked zone.

A second focus is on the feasibility of computing very high-fidelity attacks with a sparse signature. The starting point of this work is the currently used “State Estimation” procedure whereby sensor readings are used together with a linearized model of power flow in order to estimate other system parameters. In its simplest form, this procedure uses the DC power flow model,

\[
B\theta = P^g - P^d,
\]

where \(B\) is the bus susceptance matrix, \(\theta\) is the vector of phase angles, and \(P^g\) and \(P^d\) are (respectively) the vectors of active power generation and load. Sensors, which may not be ubiquitous, report phase angles and statistical estimation procedures can be used to recover missing readings as well as other operational data. As discussed in the above works an attacker that is able to modify sensor output may be able to alter the true phase angles \(\theta\) through a perturbation \(\delta\) in the null space of \(B\), the vector \(\theta' = \theta + \delta\) is thus consistent with the equations (1). The resulting attack is thus considered undetectable. As shown in [1] such attacks may be sparse (i.e. \(\delta\) has small support); computation of an optimally sparse attack is considered in [2]. In the “cyber-physical” version of an attack the attacker additionally modifies the topology of the network; in that case the data component of the attack is designed to hide this fact. Cyber attacks that modify admittances are considered in [17], which also uses inconsistencies in AC the current-voltage relationship to pinpoint the location of an attack. [9] performs PCA (principal component analysis) on the covariance of power flows to discover anomalies by inspecting changes in the smaller eigenvalue modes, also see [18], [10], [11], [12], [19], [20].

When the attacker does not have unhindered access to sensors, or if e.g. the result of the attack is that sensors stop reporting, sophisticated techniques may still be brought to bear in order to identify, for example, the topology modification. See [13]–[16]. Most of the above work relies on the DC
(linear) model of power flows; the model in [15] (as far as we can tell) was the first to use the nonlinear AC power flows model. Under the assumption that the physical attack disconnects lines, that defender knows the data-attacked zone of the grid, and structural assumptions about the attack, the techniques in [15] recover the attack.

II. NOTATION

We represent AC power flows using the polar representation. The voltage at a bus $k$ is of the form $V_k = |V_k|e^{j\theta_k}$ where $j = \sqrt{-1}$. A line $km$ is described by using the standard “\pi” model which includes series impedance, line charging and transformer attributes. See, e.g. [21], [22]. Under this model, the complex currents injected into line $km$ at bus $k$ and $m$ (resp.) are given by the formula

$$ I_{km} = Y_{km} (V_k, V_m), \tag{2} $$

where $Y_{km}$ is the branch admittance matrix for line $km$; the complex power injected into line $km$ at $k$ equals $p_{km} + jq_{km} = V_k I_{km}$. Thus $p_{km} = p_{km}(V_k, V_m)$ and $q_{km} = q_{km}(V_k, V_m)$ are quadratic functions of the voltages at $k$ and $m$, which can be summarized in the form

$$ p_{km} + jq_{km} = S_{km}(|V_k|, |V_m|, \theta_k, \theta_m). \tag{3} $$

The complex power flow and angle limits on a line $km$ are denoted by $S_{km}^{\text{max}}$ and $\theta_{km}^{\text{max}}$ (respectively), the voltage limits at a bus are given by $V_k^{\text{min}}$ and $V_k^{\text{max}}$, and the active and reactive limits at a generator bus $k$ are indicated by $P_k^{g,\text{min}}, P_k^{g,\text{max}}$ and $Q_k^{g,\text{min}}, Q_k^{g,\text{max}}$ (resp.).

Given a bus $k$ we denote by $\delta(k)$ the set of all lines of the form $km$, $\mathcal{N}$ is the set of buses (we write $n = |\mathcal{N}|$) and $\mathcal{G}$ is the set of generator buses, given a set of buses $S$ we denote by $\partial S$ (the boundary of $S$) the subset of buses of $S$ that are incident with a line with an end not in $S$.

We model AGC (Automatic Generation Control) as follows. There is a selected subset of generators $\mathcal{R} \subset \mathcal{G}$ (the participating generators) and parameters $\alpha_k \geq 0$ for $k \in \mathcal{R}$ (the participation factors) with $\sum_{k \in \mathcal{R}} \alpha_k = 1$. If aggregate net active power generation changes by some value $\Delta$, with generator $k \in \mathcal{R}$ changing its output by $\alpha_k \Delta$.

The susceptance matrix $B$ of the DC power flow model (1) is defined by $B_{kk} = \sum_{km \in \delta(k)} 1/x_{km}$ for any bus $k$, $B_{km} = -1/x_{km}$ for any line $km$, and $B_{km} = 0$ otherwise; where $x_{km} > 0$ is the reactance of line $km$.

III. IDEAL ATTACKS

In this section we will focus on a precise formulation for a high-fidelity attack problem. We use the term ‘initial’ to indicate that the attack comprises actions taken at one point in time. Later we will discuss a ‘dynamic’ phase that follows the initial attack. The following template applies to this problem:

1For simplicity we assume at most one generator per bus.

Template III.1. Initial Attack

(a) It is assumed that at each bus $k$ there is a sensor measuring voltage at $k$ and current at each line $km \in \delta(k)$.
(b) The attacker has selected a (sparse) subset $\mathcal{A}$ of buses, as well as a target line $uv$ within $\mathcal{A}$ that will be overloaded.
(c) For any bus $k \in \mathcal{A}$, the attacker can modify data provided by a sensor located at $k$.
(d) The attacker’s physical actions are of two types. First, the attacker can modify loads at buses in $\mathcal{A}$. Additionally the attacker can disconnect lines with both ends in $\mathcal{A}$.
(e) Actions (c)-(d) are performed in a single step.
(f) The data received by the control center satisfies complete fidelity as per AC power flow equations and shows all system limits being satisfied, while in actuality line $uv$ is overloaded.
(g) When the attack includes load changes, secondary response (i.e. AGC response) is taken into account by the attacker.

Conditions (a), (f) and (g) amount to a strong form of undetectability. Nevertheless, we provide examples of large scale systems that are susceptible to attacks of the form (a)-(g).

Note that we allow loads to be modified, but not generation. In our numerical examples we enforce that $\mathcal{G} \cap \mathcal{A} = \emptyset$, out of a perception that generator sites are more carefully protected.

We next present conditions that we will impose so as to guarantee undetectability. True data will be the true physical data. In contrast, reported data is that which is actually received by the control center and includes the attacker’s modifications. The true data will be given by the (voltage, current) pair of vectors $(V^T, I^T)$ whereas the reported data will be given by $(V_R^T, I_R^T)$.

An important requirement for the reported data is current-voltage consistency:

$$ I_R = Y_k(V_R, V_m), \tag{4} $$

(i.e.) equation (2). This condition will be enforced in the computation given below in an indirect fashion (also see [17] for a different use of this requirement). In general, of course, an attacker might only seek approximate consistency, using ambient noise to hide errors. Additionally:

(s.1) On a bus $k \notin \mathcal{A}$ the true and reported data agree (no data modification outside $\mathcal{A}$, by definition).
(s.2) At a bus $k \in \partial \mathcal{A}$ the attacker is constrained by the condition $V_k^R = V_k^T$. This condition is necessary to avoid attack detection, given (a) and the second equation in (4) applied to a line $km$ where $m \notin \mathcal{A}$.
(s.3) On buses $k \in \mathcal{A} - \partial \mathcal{A}$ we may have $V_k^R \neq V_k^T$ and on lines with both ends in $\mathcal{A} - \partial \mathcal{A}$ the true and reported currents may also differ.
(s.4) The reported voltages and currents must be consistent with meaningful (complex) power injections. Specifically, consider a bus $k$. Then $\sum_{km \in \delta(k)} V^R_k I^R_{km}$ equals the power injected into the system at bus $k$, according to
If the attack causes a net change in the sum of loads,
where “T” indicates true and “R”, reported:

This formulation uses the following real-valued variables,

\[
\forall u,v \in A \subseteq N \setminus G
\]

A. Formulation of the initial attack problem

Formally, as an input to the initial attack problem we have a set \( A \subset N \setminus G \) of buses, a set of lines \( L \subset E \) and a line \( uw \in E - L \), with \( u,v \in A \). Write \( A^C = N \setminus A \). Let \( (S^p_k, jQ^q_k)_{k \in N} \) and \( (S^\hat{p}_k, jQ^\hat{q}_k)_{k \in N} \) be (resp.) the estimated complex power generation and loads at the time
of the attack. We assume that the attacker observes all these quantities.

The initial attack problem is given by the following formulation.
An explanation of the variables and constraints will be provided below.

\[
\begin{align*}
\text{Max } & \quad (p_{uv}^T)^2 + (q_{uv})^2 \\
\text{s.t. } & \quad \forall k \in A^C \cup \partial A, \ |V_k^T| = |V_k^R|, \ \theta_k^T = \theta_k^R \quad (5a) \\
& \quad \forall k \in A, \ - (P_k^{d,R} + jQ_k^{d,R}) = \sum_{k \in \delta(k)} (P_{km}^R + jQ_{km}^R), \quad (5c) \\
& \quad - (P_k^{d,T} + jQ_k^{d,T}) = \sum_{k \in \delta(k)} (P_{km}^T + jQ_{km}^T), \quad (5d) \\
& \quad P_k^{d,T} \geq 0, \ P_k^{d,R} \geq 0 \quad (5e) \\
& \quad \forall k \in A^C \setminus R : \quad P_k^g - P_k^{\hat{g}} + j(Q_k^g - Q_k^{\hat{g}}) = \sum_{k \in \delta(k)} (P_{km}^T + jQ_{km}^T), \quad (5f) \\
& \quad \forall k \in R : \quad P_k^g - P_k^{\hat{g}} = \alpha_k \Delta \quad (5g) \\
& \quad \forall k \in G : \quad P_k^{g,\min} \leq P_k^g \leq P_k^{g,\max}, \ Q_k^{g,\min} \leq Q_k^g \leq Q_k^{g,\max} \quad (5i) \\
& \quad \forall k \in N : \quad V_k^T \leq |V_k^R|, |V_k^R| \leq V_k^{max} \quad (5j) \\
& \quad \forall k \in E : \quad |\theta_k^T - \theta_k^R| \leq \theta_{k,\max} \quad (5k) \\
& \quad \forall k \in E : \quad \max\{\|p_{km}^R - Q_{km}^R\|, \|p_{mk}^R - Q_{mk}^R\|\} \leq S_{km}^{\max} \quad (5l) \\
& \quad \forall k \in E : \quad P_k^T + jQ_k^T = S_k \left(|V_k^R|, |V_k^R|, \theta_k^T, \theta_k^R\right), \quad \forall k \in L \quad (5m) \\
& \quad \forall k \in E : \quad P_k^R + jQ_k^R = S_k \left(|V_k^R|, |V_k^R|, \theta_k^R, \theta_k^R\right), \quad \forall k \in L \quad (5n) \\
& \quad \forall k \in E : \quad P_k^R + jQ_k^R = S_k \left(|V_k^R|, |V_k^R|, \theta_k^R, \theta_k^R\right), \quad \forall k \in L \quad (5p)
\end{align*}
\]

This formulation uses the following real-valued variables, where “\( T \)” indicates true and “\( R \)” reported:

\[
\begin{align*}
& |V_k^T|, \theta_k^T, |V_k^R|, \theta_k^R \forall \text{ bus } k \in N \quad (true \text{ and reported voltage magnitudes and angles}) \\
& |P_k^{d,T}, Q_k^{d,T}, P_k^{d,R}, Q_k^{d,R} \forall \text{ bus } k \in A \quad (active \text{ and reactive, true and reported loads in } A) \\
& |P_k^g, Q_k^g \forall \text{ bus } k \in R \quad (generation \text{ at participating buses}) \\
& \forall \text{ line } km \in E, p_{km}, q_{km}, \text{ and also } p_{km}, q_{km}, \text{ if } km \notin L \quad (active \text{ and reactive, true and reported power flows}) \\
& \Delta \text{ (net change in active power generation)}
\end{align*}
\]

In this formulation, power flows are represented through the quadratics \( P_k^g, Q_k^g, P_{mk}, Q_{mk} \) (see eqs. (3)) which appear in the formulation as \( (5m)-(5p) \). Also note that the problem includes voltage variables but no current variables. However, having solved the above optimization problem, the attacker reports, for each line \( km \) with both ends in \( A \), a current pair \( I_{km}^R, I_{mk}^R \) computed using the formula

\[
(i_{km}^R) = Y_{km} \left( |V_k^R|, e^{i\theta_k^R}, |V_{km}^R|, e^{i\theta_{km}^R}\right),
\]

thereby attaining current-voltage consistency. Note that if either \( k \in \partial A \) or \( m \in \partial A \) the true and reported voltage values are identical – see Lemma 3 below.

**Lemma 1.** Consider a feasible solution to problem (5). Let \( H \)
\( \text{ denote either } \emptyset \text{ or } R \) (i.e. true or reported). Then the voltages
\( |V_k^H|^e^{i\theta_k^H} \) for all \( k \in N \) yield a solution to the power flow problem where

1. Bus \( k \) has load \( P_k^T + jQ_k^T \) for \( k \in A \) and \( P_k^r + jQ_k^r \) if \( k \in A^C \).
2. Bus \( k \in G \) has generation \( P_k^g + jQ_k^g \) if \( g \in R \) and \( P_k^g + jQ_k^g \) if \( g \in G \setminus R \).
3. Line \( km \) has power flow \( \hat{P}_{km}^T + \hat{Q}_{km}^T \) when \( H = R \) and also when \( H = T \) and \( km \notin L \).
4. When \( H = R \) (reported data) the solution is fully feasible, i.e. it satisfies voltage, generation, phase angle and power flow limits.
5. When \( H = T \) (true data) the solution satisfies voltage, generator and phase angle limits, but only satisfies power flow limits on lines \( km \) with both \( k, m \in A \cap \partial A \). Furthermore, the solution is consistent with lines in \( L \) being cut.

**Proof.** Property (3) follows from constraints (5m)-(5p). Hence, (1) and (2) follow from constraints (5c)-(5f). Properties (4)-(5) follow from constraints (5i)-(5j).

As a corollary to (1)-(2) of Lemma 1, a feasible solution to problem (5) satisfies, exactly, power-injection consistency, i.e. condition (s.5) above:

\[
\sum_{km \in \delta(k)} V_{km}^R \text{ is net injection at } k \quad \forall k,
\]

where ‘net injection’ is the reported net injection on buses in \( A \) and the true net injection for buses not in \( A \).

**Lemma 2.** Consider a feasible solution to problem (5). The solution is consistent with a secondary-response adjustment of active power generator amounting to \( \Delta \) units.

**Proof.** Follows from constraint (5g).
Lemma 3. Consider a feasible solution to problem (5). Then (a) the true and reported voltages agree on \( \mathcal{A}^C \cup \partial \mathcal{A} \). Further, (b) the true and reported currents on a line \( km \) are identical if \( k, m \in \mathcal{A}^C \cup \partial \mathcal{A} \).

Proof. (a) Follows from constraint (5b), and (b) is a consequence of (a). □

Corollary 4. Suppose we compute a feasible solution to problem (5) whose objective value is strictly greater than \( (S_{uv}^{\max})^2 \). Then the reported solution amounts to an undetectable attack that hides an overload on line \( uv \).

B. Computational viability

Above we have presented a mathematically correct version of the initial attack problem that would lead to an (initially) undetectable attack. However, would such an attack be realistic? How would an attacker identify the set \( \mathcal{A} \) of attacked buses?

Problem (5) is a nonlinear, nonconvex optimization problem, and thus, in principle, a challenging computational task. Nevertheless this problem is similar to the standard ACOPF or PF problem and (at least) a local optimum should be efficiently computable; this expectation is borne out by our experiments. Strict maximization in (5) is not required for an attack to be successful (all that is needed is an overload of the line \( uv \)).

A broader issue concerns the selection of the sets \( \mathcal{A} \) and \( \mathcal{L} \). This is a combinatorial problem which is bound to be intractable. In fact [16] describes a number of strong NP-hardness results in the DC setting, e.g. given vectors of phase angles \( \theta \) and \( \theta' \) it is NP-hard to compute a set \( \mathcal{L} \) such that \( B' \theta' = B \theta \) where \( B' \) is the bus susceptance matrix of the network with \( \mathcal{L} \) removed.

Nevertheless, as discussed in the literature, an attacker may be willing to incur significant computational costs in order to compute a successful attack. While it is reasonable to assume that an attacker’s ability to take physical action or to modify data is limited (see the discussion in [1], [2]), not assuming computational intelligence on the part of an attacker amounts to a limitation on the part of the defender.

We separate two distinct issues here: first the identification of the set \( \mathcal{A} \), which is done in advance and may be computationally intensive, and second, the solution to problem (5) which only requires a few seconds. Let us assume that the attacker has had (undetected) access to system and sensor data long enough to identify a weak sector of the transmission system, i.e. the set \( \mathcal{A} \). In this task the attacker would rely on the fact that typical (time- and day-dependent) load and generation profiles for transmission systems are statistically predictable with some accuracy. This fact would help the attacker in the computation of a target set \( \mathcal{A} \), perhaps using enumeration, using load estimates in problem (5).

Having identified a particular set \( \mathcal{A} \), problem (5) would be run once again just prior to the attack, now using close estimates of the loads obtained from ambient conditions. Assuming that the attack is perpetrated during a period of slowly changing loads, and not close in time to a generator redispatch, the attack will likely be sufficiently numerically accurate so as to become difficult to detect.

IV. Example

In the following instance we consider the case2746wp (that has 2746 buses) from the Matpower case library [23]. The adversary attacks the set of buses \( \mathcal{A} = \{1137, 1138, 1139, 1141, 1361, 1491\} \) with \( \mathcal{A} - \partial \mathcal{A} = \{1137, 1138, 1141, 1491\} \). See Figure 1. In this attack the quantity \( \Delta \) in (5g) equals 135.09. We also have \( \mathcal{L} = \emptyset \) (no lines are cut). The set of generators participating in secondary response is \( \mathcal{R} = \{17, 18, 55, 57, 150, 383, 803, 804, 1996\} \) with participating factors \( \alpha_k = 1/9 \) for all \( k \in \mathcal{R} \).

Table I shows the true and reported flow for lines where the solutions differ, with a strong overload on lines (1361, 1141) and (1138, 1141).

| bus \( k \) | bus \( m \) | \( P_{km}^T \) | \( P_{km}^R \) | \( q_{km}^T \) | \( q_{km}^R \) | \( \|P_{km}^T - P_{km}^R\| \) | \( \|q_{km}^T - q_{km}^R\| \) | \( S_{km}^{\max} \) |
|---|---|---|---|---|---|---|---|---|
| 1139 | 1137 | 5.36 | 3.36 | 2.66 | 2.66 | 4.29 | 4.28 | 114.00 |
| 1361 | 1141 | 108.51 | 108.51 | 10.49 | 10.49 | 109.02 | 109.02 | 114.00 |
| 1141 | 1491 | 13.46 | 6.20 | 2.41 | 2.39 | 13.68 | 6.64 | 114.00 |
| 1141 | 1138 | 209.25 | 98.06 | 4.44 | 5.24 | 209.29 | 98.20 | 114.00 |

Table II displays the load and generation of the buses involved in the attack.

| bus \( k \) | \( P_{gk}^T \) | \( Q_{gk}^T \) | \( P_{gk}^R \) | \( Q_{gk}^R \) | \( \alpha_k \Delta \) |
|---|---|---|---|---|---|
| 1137 | 0 | 0 | 3.36 | 2.68 | 14.74 | 1.37 |
| 1138 | 103.29 | 29.84 | 208.91 | 3.32 | 20.82 | 1.46 |
| 1139 | 0 | 0 | 3.36 | 2.68 | 15.80 | 1.37 |
| 1141 | 0 | 0 | 5.97 | 2.64 | 24.29 | 1.45 |
| 1361 | 0 | 0 | 1.58 | 2.39 | 91.90 | 1.45 |
| 1491 | 4.76 | 1.12 | 13.45 | 2.65 | 20.58 | 1.48 |

Fig. 1. Attacked zone \( \mathcal{A} \) and its neighborhood. Generators are shown in blue.
V. THE DYNAMIC PHASE

Following the initial attack, the attacker needs to dynamically perturb the attack data so as to produce a realistic data stream that is both consistent and continues to hide the overload.

In analogy to our notation for the initial problem, at time \( t > 0 \) following the attack, we denote by \( V_k^R(t) \) and \( V_k^I(t) \) the observed (or computed) pair of time series \( V \). Here the attacker supplies a previously data replay our analyses below apply to either form. Reported data for \( A \) will be manufactured by the attacker and (from the attacker’s perspective) need to approximately satisfy current-voltage consistency (4) and power-injection consistency (6).

In addition, in this work we assume that the attack is perpetrated when ambient conditions (in particular loads) are, approximately, constant. Let us denote by \( V_k^R(0) \) the voltage at \( k \) computed by the initial attack, i.e.

\[
V_k^R(0) = |V_k^R| e^{j\phi_k^R}
\]

and likewise define the current \( I_{km}(0) \). The statement that ambient conditions are approximately constant, post-attack, can be informally rephrased as

\[
V_k^R(t) \approx V_k^R(0) \forall k, \quad I_{km}(t) \approx I_{km}(0) \forall km.
\]

If indeed ambient conditions are approximately constant (7) will hold (statistically) for any bus or line not in the attacked zone \( A \) but are a requirement for the attacker.

Two types of attack have been used in the literature. First, the “noisy data” attack in our setting works as follows:

**Template VI.1. Noisy Data Attack**

At time \( t > 0 \) the attacker reports at each bus \( k \in A \) a voltage \( V_k^R(t) = V_k^R(0) + \nu_k(t) \). Here \( \nu_k(t) \) is a random phasor drawn from small variance, zero mean distribution. Likewise the attacker reports for each line \( km \) with both ends in \( A \), currents

\[
\begin{bmatrix} I_{km}^R(t) \\ I_{km}^I(t) \end{bmatrix} = \begin{bmatrix} I_{km}^R(0) \\ I_{km}^I(0) \end{bmatrix} + \begin{bmatrix} \mu_{km}(t) \\ \mu_{mk}(t) \end{bmatrix}
\]

where \( \mu_{km}(t), \mu_{mk}(t) \) are drawn from zero mean distributions with small variance.

Note that these definitions satisfy requirement (7), and approximately satisfy current-voltage consistency. As a functionally equivalent alternative to (8) the attacker could simply set

\[
\begin{bmatrix} I_{km}^R(t) \\ I_{km}^I(t) \end{bmatrix} = Y_{km} \begin{bmatrix} V_k^R(t) \\ V_{km}^I(t) \end{bmatrix},
\]

our analyses below apply to either form.

A second form of attack that has been considered is the data replay attack. Here the attacker supplies a previously observed (or computed) pair of time series \( V_k^R(t) \) and \( I_k^R(t) \) for buses and lines within the set \( A \).

1) Discussion: The reader may recall that in the initial attack computation we enforced that reported voltages in \( \partial A \) are exact, i.e. equal to the true voltages. In the dynamic phase this condition is necessarily relaxed by the attacker, though this action carries the risk (to the attacker) that current-voltage consistency will not hold, statistically, for some line \( km \) with \( k \in \partial A \) and \( m \notin A \). Thus e.g. in the noisy data attack template given above the distributions for the \( \nu_k(t), \mu_{km}(t) \) and \( \mu_{km}(t) \) should have sufficiently small variance relative to the variance of ambient conditions. Further requirements on such variances will be discussed in Section VII.1. In any case, when ambient conditions (e.g. loads) are nearly constant, the noisy-data attack may continue to approximately satisfy current-voltage and power-injection consistency and thus remain numerically undetectable. The same holds for the data replay version provided the replayed voltages in \( \partial A \) closely approximate ambient conditions.

In the next section we present defensive mechanisms that dynamically change voltages in a way that is unpredictable to an attacker. The key observation is that a substantial change to voltages in \( \partial A \) will cause the noisy-data attack, applied verbatim as in Template VI.1, to fail, because of large current-voltage inconsistencies on lines \( km \) with \( k \in \partial A \) and \( m \notin A \). Of course, the template need not be applied verbatim, and in particular the attacker may seek to leverage the possibility of sensor error. We will consider this point in the next section.

In [15] current-voltage consistency is used in a different setting: (i) the attacked zone \( A \) is known by the defender, (ii) the attacker only disconnects lines. Under a number of assumptions, in particular that there is a matching between \( A^C \) and \( A \) that covers all buses in \( A \) it is shown that the attack can be accurately recovered.

VI. DEFENSE

In the above sections we showed that, conceptually at least, it is possible to compute high-fidelity attacks that disguise dangerous network conditions. Other attacks are also potentially conceivable, e.g. impedance changes, transformer tap changes, etc. In this section we describe a generic randomized defense strategy that can deployed when a complex attack is suspected. We will assume that the attack impacts a proper subset \( A \) of the system that is unknown to the control center, as was the case above, though the generic defense strategy applies under more general attacks as well. The strategy can be summarized by the following template:

**Procedure VI.1. Random Defense.**

Iterate:

1. **D1:** Choose, for each \( k \in G \) a (random) value \( \delta_k \) such that \( \sum_{g \in G} \delta_k \approx 0 \). Command each generator \( k \in G \) to change its output to \( P_k^9 + \delta_k \).
2. **D2:** Following the generation change in step **D1** identify inconsistencies in the observed sensor readings.

Each iteration would last several seconds, and statistically significant inconsistencies identified by this scheme would be flagged as potential evidence of an attack. The above strategy could be AGC-like if only generators \( k \in R \) (the responding generators) are allowed to have \( \delta_k \neq 0 \) and in general it amounts to a generator redispatch. An additional point is that an implementation of step **D1** should guarantee safe system operation; this consideration leads to computation of the \( \delta_k \) in step **D1** by means of an OPF-like problem.

We note that there is an existing literature on using network resources so as to change power flow physics in order to detect
structure or faults. See [24], [25], [26], [27], [28]. Indeed, even though the description of our random defense focuses on power injections, one could also consider other random probing strategies that change power flows, such as adjusting transformer settings, controlled line tripping, and the use of DERs, storage and FACTS devices.

There are several implementations of the generic strategy. Generally the defender wants to make the $| \delta_k |$ large because to first order changes in voltage angles are proportional to $\| \delta \|_2$, and a large change in phase angles is likely to give rise to a significant current-voltage or power-injection inconsistencies in sensor readings in $\partial A$, as discussed above. This idea forms the basis for a simple, current-consistency based version of Template VI.1 given in Section VI-A.

An attacker aware that the random defense strategy is applied may try to replace e.g. the noisy data attack with a more careful manipulation of reported data. For example, the attacker could react to a significant change to voltages in $\partial A$ by solving a nonlinear, nonconvex system of inequalities designed to guarantee approximate current-voltage and power-injection consistency. In addition, any implied load change within $A$ must be very small (or it would contradict observed frequencies). Finally the attacker would need to perform this computation very quickly, and repeatedly (because the defense will be applied repeatedly).

A more practicable alternative (for the attacker) would be to consider arbitrary changes to voltages in buses in $\partial A$, with the remaining voltages in $A$ obtained as in Template VI.1. We will term this the enhanced noisy data attack. In Section VI-A we will show that when the random defense causes large-enough voltage changes in $\partial A$, the enhanced noisy-data attack fails. See Lemma 5.

A more sophisticated defensive idea, given in Section VII, changes the stochastics of power flow data, in particular voltage covariance, and probes the corresponding properties of the reported data.

Our defensive strategies can be easily adjusted if sensors are not available throughout the system. Of course, the fewer the sensors the more limited the impact of the defense. Indeed, some interesting work (using the standard, DC-equation state estimation) precisely seeks to perform system identification post-attack when only limited sensor information is available [13], [14], [15], [16].

### A. Current-voltage defense

Next we describe a particular implementation of step D2 of the generic defensive strategy which relies on the current-voltage consistency condition (4). This implementation will take into account the possibility of sensor error. Whereas above a phasor $\phi$ had a true value $\phi^T$ (the physical value) and a reported value $\phi^R$ (the value received by the control center), now we will have the sensed value $\phi^S$ which is the value actually produced by the sensor. Reported and sensed data arising from sensors not controlled by the attacker are equal, but may differ at sensors within $A$.

In addition, sensed and true data may differ. For a phasor $\phi$ define $\text{err}(\phi) \triangleq \phi^S - \phi^T$. In the PMU setting, the TVE (total vector error) criterion [29], [30] guarantees that

$$|\text{err}(\phi)| < \tau |\phi^T|,$$

where $0 < \tau < 1$ is a tolerance. Standards enforce $\tau = 1\%$, though experimental testing of PMUs shows far smaller errors [31]. From (10) we obtain

$$|\text{err}(\phi)| < \tau |1 - \tau|^{-1} |\phi^S|.$$  \hspace{1cm} (11a) \hspace{1cm} (11b)

We will describe three voltage-current consistency criteria. An important point is that the current-voltage consistency condition, combined with estimations of possible sensor error, yield a nonlinear relationship, and appropriately reformulation of this relationship can render useful benefits. To simplify notation we will drop the “$(t)$” from phasors though it should be understood throughout. For a line $km$ write

$$Y_{km} = \begin{pmatrix} y_{km}^{(1)} & y_{km}^{(2)} \\ y_{km}^{(3)} & y_{km}^{(4)} \end{pmatrix}$$

**Criterion 1.** We have that $I_{mk}^T = Y_{km}^T V_{km} - Y_{km}^T V_{m}$. Write $Z_{km}^{(3)} = [Y_{km}^{(3)}]^{-1}$. Hence

$$V_{k}^S - Z_{km}^{(3)}(I_{mk}^S - Y_{km}^S V_{km}) = \text{err}(V_k) - Z_{km}^{(3)}(\text{err}(I_{mk}) - Y_{km}^S \text{err}(V_m))$$

which yields, using (10), (11b), and the triangle inequality

$$|V_k^S - Z_{km}^{(3)}(I_{mk}^S - Y_{km}^S V_{km})| < \tau |V_k^T| + \frac{\tau |Z_{km}^{(3)}||I_{mk}^S - Y_{km}^S V_{km}|}{1 - \tau} =$$

$$\begin{aligned} 
\tau |Z_{km}^{(3)}(I_{mk}^T - Y_{km}^T V_{m})| + \frac{\tau |Z_{km}^{(3)}||I_{mk}^T - Y_{km}^T V_{m}|}{1 - \tau} &\leq \frac{2\tau |Z_{km}^{(3)}||I_{mk}^T - Y_{km}^T V_{m}|}{1 - \tau} \end{aligned}$$

(13)

Under Criterion 1, if, statistically, the reported phasors $V_k^R$, $V_m^R$, $I_{mk}^R$ fail to satisfy (13) line $km$ is flagged as suspicious. A similar analysis concerns $V_k^R$, $V_m^R$, $I_{mk}^R$. **Remark:** By construction, if $k, m \notin A$ then line $km$ will not be flagged.

**Criteria 2 and 3.** Proceeding as above we have

$$|I_{km}^{(1)} - Y_{km}^S V_{km}^S| < \frac{\tau}{1 - \tau} (|I_{km}^{(1)}| + |Y_{km}^{(1)}||V_{km}^S||V_{m}^S|),$$

and similarly with $I_{mk}$, and when line $km$ is a pure impedance line (no transformer) with charging susceptance $b_{km}$,

$$|I_{km}^{S} + I_{mk}^{S}| < \frac{|b_{km}|}{2(1 - \tau)} (|V_{k}^{S}| + |V_{m}^{S}|) + \frac{\tau}{1 - \tau} (|I_{km}^{S}| + |I_{mk}^{S}|).$$

(14)

(15)

If the reported phasors do not satisfy (14) or (15) then the line is flagged.
1) Discussion: Note that a line not attacked will not be flagged, as per the TVE condition. Additional criteria can be developed to handle power-injection consistency. However, Criterion 1 alone can suffice to defeat the attacker (i.e. uncover inconsistencies) when voltage angles are sufficiently changed under our random defense. To understand this point, consider a bus $k \in \partial \mathcal{A}$ such that there is a line $km$ with $m \notin \mathcal{A}$ and also a line $ka$ where $a \in \mathcal{A} - \partial \mathcal{A}$.

We study an iteration of the random defense which (to simplify notation) we assume begins at time $t = 0$. Consider line $ka$ first. To avoid having line $ak$ flagged, the attacker $t$ will need to manufacture a time series $V^n_k(t)$, $V^n_a(t)$ and $I^n_{ak}(t)$ that (statistically) satisfy (13). But under the noisy data attack, on average $V^n_a(t) = V^R_a(0)$ and $I^n_{ak}(t) = I^R_{ak}(0)$. Hence the attacker needs (on average) that

$$2\tau |Z_{ka}^{(3)}|\left(|I^n_{ak}(0)| + |Y^{(4)}_{ka}||V^R(0)|\right) > |V^R_k(t) - Z_{ka}^{(3)}(I^n_{ak}(0) - Y^{(4)}_{ka}V^R_a(0))| = |V^R_k(t) - V^R_k(0)|$$

Now consider line $km$. Since $m \notin \mathcal{A}$, $V^n_m(t) = V^S_m(t)$ and $I^n_{mk}(t) = I^S_{mk}(t)$. Also, denote:

- $V^S_k(\ast)$: the true voltage at $k$ at the start of the current iteration of the random defense, i.e. the voltage resulting from the injection change in step D1. Then, assuming unbiased sensor errors and zero-mean ambient noise, $V^S_m(\ast)$ will equal the expectation of $V^S_m(t)$ during the iteration.

- Likewise define the current $I^S_{mk}(\ast)$.

Hence the attacker needs (on average) that

$$2\tau |Z_{km}^{(3)}|\left(|I^n_{mk}(\ast)| + |Y^{(4)}_{km}||V^S_m(\ast)|\right) > |V^R_k(t) - Z_{km}^{(3)}(I^n_{mk}(\ast) - Y^{(4)}_{km}V^S_m(\ast))| = |V^R_k(t) - V^S_k(\ast)|$$

As a result of these observations we have:

**Lemma 5.** Consider buses $k, a, m$ as described above. Suppose that

$$|V^S_k(\ast) - V^S_m(\ast)| > \frac{2\tau |Z_{ka}^{(3)}|\left(|I^n_{ak}(0)| + |Y^{(4)}_{ka}||V^R(0)|\right)}{1 - \tau} + \frac{2\tau |Z_{km}^{(3)}|\left(|I^n_{mk}(\ast)| + |Y^{(4)}_{km}||V^S_m(\ast)|\right)}{1 - \tau}$$

Then it is impossible for the enhanced noisy data attacker to statistically satisfy Criterion 1 on both lines $ka$ and $km$. □

Comment: This lemma highlights how large changes in voltages caused by the random defense challenge the attacker.

2) Experiment: Next we describe a set of experiments involving the current-voltage defense applied to the attack given in Section IV. The current defense was implemented as follows:

- For any generator bus $k \notin \mathcal{R}$, $|\delta_k| \leq \epsilon P^n_k$. We used values $\epsilon = 0.01, 0.05$.
- The set of responding generators, $\mathcal{R}$, was of cardinality 200. For $k \in \mathcal{R} |\delta_k|$ can be arbitrarily large. We chose $\delta_k > 0$ with probability 1/2.

- No generator may exceed its limits (voltage or generation), but subject to all these conditions we maximize $\sum_{k \in \mathcal{G}} |\delta_k|$.

In Table III, we perform the above analysis on the lines ($k = 1139, a = 1137$) and ($k = 1139, m = 1110$) with $\tau = 0.01$. “Ratio” is the ratio of the left-hand side to the right-hand side of expression (16). We see that the condition for Lemma 5 is amply satisfied. A similar analysis pertains to line (1141, 1361), the other line connecting $\mathcal{A}$ to its complement.

| Table III | CURRENT-VOLTAGE DEFENSE. |
|-----------|--------------------------|
| $\epsilon$ | Experiment 1 | Experiment 2 |
| $\sum_{k \in \mathcal{G}} \delta^+_k$ | 289.01 | 964.77 |
| $\sum_{k \in \mathcal{G}} \delta^-_k$ | 174.47 | 256.04 |

| Line ($k = 1139, a = 1137$) | $|V^R_k(0)| > \theta^R_{ka}(0)$ | $I^R_{ak}(0)$ |
|-----------------------------|-----------------------------|---------------|
| $V^R_k(0)$ | $0.0919 \angle -6.993^o$ | $-0.0275 + 0.0281j$ |
| $I^R_{ak}(0)$ | $0.0275 + 0.0281j$ | $-0.0275 + 0.0281j$ |

| Line ($k = 1139, m = 1110$) | $|V^R_k(0)| > \theta^R_{km}(\ast)$ | $I^R_{mk}(\ast)$ |
|-----------------------------|-----------------------------|---------------|
| $V^R_k(0)$ | $0.0919 \angle -7.822^o$ | $0.0905 - 0.4976j$ |
| $I^R_{mk}(\ast)$ | $0.0905 + 0.4976j$ | $0.1289 - 0.4901j$ |

| Voltages at $k = 1139$ | $V^R_{mk}(0)$ | $\theta^R_{mk}(\ast)$ |
|------------------------|--------------|---------------------|
| $V^R_{mk}(0)$ | $0.0919 \angle -6.991^o$ | $0.0919 \angle -6.991^o$ |
| $V^R_{mk}(\ast)$ | $0.1014 \angle -7.822^o$ | $0.1014 \angle -7.822^o$ |

| Lemma 5 applied to bus $k = 1139$ | $\frac{V^R_k(0) - V_S^R(k)}{V^R_k(0) - V_S^R(k)}$ |
|---------------------|---------------------|
| Ratio | 1.913 | 1.732 |

B. Controlling voltages through generation changes

From a basic perspective, an application of the current-voltage version of Procedure VI.1 will succeed when voltages in $\partial \mathcal{A}$ are sufficiently changed. The defender, of course, does not know the set $\mathcal{A}$ and thus it is of interest to understand when the voltage at any given bus can be changed by appropriately choosing the injections $\delta$.

We will provide a deterministic, positive answer to this issue by considering the DC power flow approximation (1), and for simplicity focus on cases where $|D| = 2$. Thus, let $s$ and $t$ be two buses in $\mathcal{R}$, let $\Delta > 0$, and let $P^g$ be defined by

$$P^s = P^g + \Delta, \quad P^t = P^g - \Delta, \quad \hat{P}^g = P^g_k \text{ otherwise.}$$

**Lemma 6.** Suppose $B\theta = P^g - P^d$, and $B\hat{\theta} = \hat{P}^g - P^d$. Let $k \neq t$ be a bus such that the network contains a path between $s$ and $k$ that does not include $t$. Then

$$\hat{\theta}_k - \hat{\theta}_t > \theta_k - \theta_t. \quad (17)$$

**Proof.** Equation (17) does not change if we subtract from every $\theta$ any constant, and likewise with the $\hat{\theta}$. Thus, without loss of generality $\hat{\theta}_k = \theta_k = 0$. Under this assumption (17) reads:

$$\hat{\theta}_k > \theta_t > 0. \quad (18)$$

Let $M$ be the set of buses $p \neq t$ such that

1. The network contains a path from $s$ to $p$ that avoids $t$, and
2. Subject to (1), $\hat{\theta}_p - \theta_p$ is minimum.

Aiming for a contradiction, we will assume that

$$\hat{\theta}_p - \theta_p \leq 0 \quad \text{for} \quad p \in M. \quad (19)$$
Showing that (19) is false yields (18). For any line $km$ define the flow value $f_{km} = (\hat{\theta}_k - \hat{\theta}_m - \theta_k + \theta_m)/x_{km}$. Since $B(\hat{\theta} - \theta) = \hat{P}^g - P^g$, the flow vector $f$ corresponds (under the DC power flow model) to a power flow with $\Delta$ units of generation at $s$, $\Delta$ units of load at $t$, and zero generation and load elsewhere. Note that for any line $km$, $f_{km} > 0$ iff
\[ \hat{\theta}_k - \theta_k > \hat{\theta}_m - \theta_m. \] (20)

This observation implies
\[ \hat{\theta}_s - \theta_s > 0. \] (21)

[To obtain this fact, decompose the flow vector $f$ into a set of path flows from $s$ to $t$ and telescope (20) along any such path.] Pick any $p \in M$ and let $P$ be a path from $s$ to $p$ that avoids $t$. Say $P = v_0, v_1, \ldots, v_i$ where $v_0 = s$ and $v_i = p$, and let $h$ be smallest such that $v_h \in M$. By (21) $v_h \notin M$, i.e., $h > 0$. Then by definition of $h$, $\hat{\theta}_{v_{i-1}} - \theta_{v_{i-1}} > \hat{\theta}_{v_i} - \theta_{v_i}$, i.e., $f_{v_{i-1},v_i} > 0$. But by assumption $v_h \neq t$. So there exists some line $v_h,m$ such that $f_{v_h,m} > 0$. Therefore using the assumption $\hat{\theta}_h - \theta_h \leq 0$ for all $k \in M$, $v_h \in M$, and (20),
\[ 0 \geq \hat{\theta}_{v_h} - \theta_{v_h} > \hat{\theta}_m - \theta_m. \] (22)

So $m \neq t$, and as a result by construction there is a path from $s$ to $m$ that avoids $t$. But then (22) contradicts the fact that $v_h \in M$.

Lemma 6 guarantees a nontrivial voltage angle at any bus $k$ assuming a structural relationship between $k$, $s$ and $t$. Furthermore,

Lemma 7. Suppose $k$ is any bus and that $|R| \geq 2$. Then a pair $s,t$ satisfying the assumptions of Lemma 6 exists.

Proof. Choose $s \in R$ such that $s$ is closest to $k$.

C. Attack reconstruction

In the previous sections we described defensive procedures that correctly identify a set of lines and or buses as compromised by the attack, under fairly general conditions. In this section we will present a fast algorithm that to broaden the classification of buses and lines (as safe, or not) to encompass the entire network. A central idea in this approach is the concept of a trusted bus. A bus $k$ is trusted if the control center can expect that all sensors located at $k$ provide truthful output; we assume that at least one trusted bus exists. Without such an assumption the entire suite of signals received by the control centers could be falsified and it is questionable whether any meaningful attack reconstruction can be performed.

To simplify the discussion, we will also use the following: Assumption. For any bus $k \in \partial A$ there is a line $km$ with $m \in A - \partial A$.

This assumption can be justified as follows. For any bus $h \in \partial A$, as we have discussed above sensors at $h$ report correct or approximate correct voltage values. If $k \in \partial A$ does not satisfy the assumption, then every line $km$ is such that $m \notin A$ or $m \in \partial A$. Then, assuming the initial attack satisfied current-voltage consistency for every line, all current readings provided by sensors at $k$ are also correct. In summary, all readings by sensors at $k$ are correct and the attacker may simply remove $k$ from the attacked set without impacting the attack itself. Our main reconstruction technique is given in the next template.

**Procedure VI.2. Attack reconstruction.**

**Initialization:** Every trusted bus is labeled safe and every other bus is labeled unsafe. Some subset of the lines are labeled compromised.

**Iterate:**

1: If we can find a line $km$ such that $k$ is safe, $m$ is unsafe and no line incident with $m$ is compromised, then we relabel $m$ as safe.

2: If we can find a line $km$ such that $k$ is safe, $m$ is unsafe and there is a compromised line $mh$, then we relabel $m$ as attacked.

3: If there are no lines $km$ of the type described in steps 1 or 2 then STOP.

In each step 1 or 2 one bus is relabeled from unsafe to safe or attacked; hence the procedure must stop after a finite number of iterations. The impact (and correctness) of this procedure is described by the following results.

**Lemma 8.** At termination of the procedure, any bus $m$ that is labeled safe satisfies $m \notin A$.

Proof. By assumption, no trusted bus has been attacked. Thus, aiming for a contradiction, consider the first iteration of the procedure where in Step 1 a line $km$ is used to label bus $m$ as safe, and yet $m \in A$. Since bus $k$ was previously labeled as safe, by induction it follows that $k \notin A$ and therefore it follows that $m \in \partial A$. But then there is a line of the form $mh$ that is compromised as per the current test. This fact contradicts the choice of $km$ in Step 1.

**Corollary 9.** Suppose that at termination a bus $m$ is labeled safe. Then there is a path $P$ from a trusted bus $s$, to $m$, such that $k \notin A$ for every bus $k \in P$.

**Lemma 10.** At termination of the procedure, any bus $m$ that is labeled attacked satisfies $m \in \partial A$.

Proof. Any bus $m$ that is labeled attacked at termination, was relabeled as such when considering a line $km$ in some execution of Step 2. The corresponding bus $k$, by Lemma 8, satisfies $k \notin A$. Then clearly $m \in \partial A$.

**Definition:** Let $T$ be the set of connected components of $\mathcal{N} - A$ that include trusted buses.

**Lemma 11.** Let $K$ be any of the components of $\mathcal{N} - A$. Suppose $K$ contains a trusted bus. Then at termination of the procedure all buses of $K$ are safe.

Proof. At the initialization of the procedure all trusted buses are safe and hence $K$ includes at least one safe bus. We need to prove that at termination of the procedure, all buses in $K$ are safe. If that were not the case it would follow (since $K$ is connected) that at termination there is a line $km \in K$ with $k$ safe and $m$ unsafe. But that is a contradiction given Step 1 of the procedure.
Lemma 12. Let \( K \) be any of the components of \( N - A \). Suppose \( K \) does not contain a trusted bus. Then at termination of the procedure all buses of \( K \) are unsafe.

Proof. Follows from Corollary 9.

Lemma 12 points out the implications resulting from a weak set of trusted buses. A careful choice of \( A \) can result on components of \( N - A \) without trusted buses, which are still unsafe at the conclusion of the procedure. If we can assume, for example, that all participating generators (or all generators) are located in trusted buses, current grid economics guarantees that such buses will be located throughout the network, which works to minimize the impact of Lemma 12.

As a final remark, Procedure VII.2 is a special case of the standard reachability algorithm and thus can be implemented to run in time linear in the number of buses and lines.

1) Numerical experiments using AC power flows: The above discussion concerns DC power flows. In order to investigate how voltages change under injection changes, under AC power flows, we perform an experiments using examples from the Matpower library. For each system we perform ten experiments. In each experiment we compute an AC power flow which is constrained to satisfying the given voltage bounds at all generator buses, but not at load buses, as well as power injection constraints and generator limits, while allowing large injection changes in a random subset of generators. For a non-generator bus \( k \), let \( V^g_k \) be its voltage in the base case (i.e. the Matower case), and let \( V_{i,k} \) be its voltage in experiment \( i = 1, \ldots, 10 \). Finally, define

\[
score(k) = \max_{1 \leq i \leq 10} \frac{|V_{i,k} - V^g_k|}{|V^g_k|}.
\]

In Table IV, “Min Score” is the minimum score across all non-generator buses. Thus the table provides experimental verification for substantial AC voltage changes under random generator injections.

| Case           | Min Score | Average Score |
|----------------|-----------|---------------|
| case118        | 11.61%    | 32.77%        |
| case1354pegase | 7.62%     | 51.00%        |
| case2746wp     | 5.00%     | 10.09%        |

VII. Covariance Defense

In this section we describe a different defense mechanism motivated by the fact that real-world PMU data streams exhibit non-generic stochastic structure in (for example) voltage angles [18], [32]. In particular, covariance matrices across several time scales have very low rank, and “noise” (data deviations from averages) is strongly response devices, or appropriate participating generators. We additionally assume that the buses in the set \( \mathcal{F} \) are trusted, as in Section VI-C.

Let us consider, as an example, the post-attack voltage angle vector \( \theta = (\theta_k : k \in N) \). Following the attack we obtain a time series \( \hat{\theta}(t) \) of measurements of this vector (including data corrupted by the attacker). The control center can efficiently and approximately learn statistics of \( \theta \), in particular closely estimate its covariance matrix [33], which we denote by \( \Omega \).

Under the covariance defense we choose a (random) vector \( v \in \mathbb{R}^n \). Then we perform a number of substeps (“epochs”, each lasting a second or less); in each substep the output of generators in \( \mathcal{F} \) is perturbed by a random vector \( \delta \) calculated so that \( \Omega \) is additively altered by a rank-1 correction. This perturbation in stochastics may (eventually) be understood by the attacker, but in the intervening period the attacker will generate inconsistent data. More discussion is given below.

In the template given next, \( \lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_r > 0 \) are the set of eigenvalues of \( \Omega \) that are larger than a certain threshold \( \epsilon > 0 \), and let \( w_1, w_2, \ldots, w_r \) be the corresponding eigenvectors. Let \( \bar{B} \) be a pseudoinverse of the bus susceptance matrix \( B \). Finally let \( \Delta > 0 \) be large compared to \( \epsilon \).

Procedure VII.1. Covariance Defense.

Input:

A zero-mean distribution \( \mathcal{P} \) with support in \([-1, 1]\).

Iterate:

V1: Choose a nonzero vector \( v \in \mathbb{R}^n \) so that
(a) \( \langle Bv \rangle_k = 0 \) \( \forall k \notin \mathcal{F} \).
(b) \( w_i^T v = 0 \) for \( 1 \leq i \leq r \).
(c) for each \( k \in \mathcal{F} \), the output value \( P^g_k + \Delta(Bv)_k \) is feasible for generator \( k \).

V2: For \( s = 1, 2, \ldots \) perform epoch \( s \):
(a) Draw \( x \) from \( \mathcal{P} \).
(b) Alter power injection at each \( k \in \mathcal{F} \) by \( x \Delta(Bv)_k \).

Next we comment on this procedure. V1(a)-(b) is a linear system on \( n \) variables and \( |\mathcal{F}| - r \) degrees of freedom, and thus has a nonzero solution when \( |\mathcal{F}| > r \). In addition V1(c) can be attained with nonzero \( v \) if we further assume that the buses in \( \mathcal{F} \) can all “ramp up” and “ramp down” (if necessary by scaling \( \Delta \)).

Next we analyze the procedure. We assume that the attacker does not react to the variance defense; this assumption will be discussed below. Consider an execution of step V2(b), and define \( \delta = x \Delta v \). Using V1(a) and V2(b) the \( n \)-vector of power injection changes is given by \( B \delta \).

\[
E(\delta) = 0 \quad \text{and} \quad \sigma_\delta^2 = \lambda vv^T \quad \text{where} \quad \lambda \doteq \sigma_\delta^2 \Delta^2. \tag{23}
\]

Let us denote by \( \bar{B} \) the bus admittance matrix of the network, post-attack (unknown to the control center). Let \( \theta^T \) be the (random) vector of true voltage phase angles during an execution of Step V2, and let \( \theta^T \) be the (also random) true vector of voltage phase angles prior to the defense (here randomness is driven primarily by ambient noise in loads and resulting generation adjustments). Under the DC model,

\[
\hat{B}\theta^T = P^g - P^d + B \delta. \tag{24}
\]

Let \( \bar{B} \) denote a (left) pseudoinverse of \( \bar{B} \). We then have

\[
\hat{\theta}^T = \bar{B}(P^g - P^d + B \delta) = \theta^T + \bar{B}B \delta. \tag{25}
\]

As a result we have, where \( u \doteq \bar{B}Bv \):

Lemma 13. Suppose that the random variable \( x \) in the covariance defense is stochastically independent of the attacker’s actions. Then \( \sigma^2 = \sigma_\delta^2 \).
Proof sketch. The absence of a cross-term is justified by the independence assumption.

Thus the covariance of (true) phase angles experiences a rank-1 correction; this observation applies under the DC model and only approximately under the AC model. In summary:

(a) The defender does not observe $\sigma^2_{\theta^2}$ (or $\sigma^2_{\theta T}$). Indeed, the $(\mathcal{A}, \mathcal{A})$-submatrix of the covariance matrix actually observed by the defender equals $\Omega_{\mathcal{A}, \mathcal{A}}$ both before and after the defense. The $(\mathcal{A}^c, \mathcal{A}^c)$-submatrix of $\sigma^2_{\theta^2}$ is observed.

(b) If, in an execution of step V2, the defender expects that $w_k \neq 0$ for all $k$ then the defender, aware of Lemma 13, expects to see a nonzero change in every entry of the covariance matrix.

In summary if the defender expects (b) to hold then the attacked zone $\mathcal{A}$ will be discovered. In this regard, note that $BBv$ is the change of phase angles resulting from injections $Bw$. Hence, using e.g. Lemma 6, if $v$ is sufficiently dense we would expect (b) to hold.

The result is stronger when $B = B^e$ (e.g. when the attack only involves load modifications). In that case Lemma 13 reduces to $\sigma^2_{\theta^2} = \sigma^2_{\theta^2} + \lambda v v^T$ and, by construction (b) of Step V1 the rank-1 correction is approximately orthogonal to the (observed) covariance matrix $\Omega$.

The above discussion assumed that the attacker does not react to the covariance defense. In principle the attacker could, for example, try to modify the covariance of voltage data restricted to $\mathcal{A}$. The challenge to the attacker is that this action must also preserve voltage-current (and power-injection) consistency. Additionally the attacker might first want to identify the new covariance structure that the defender has imposed. Such “learning” would require observations, i.e. time, during which the attacker is still expected to produce data radings, producing an error trail; the fact that the covariance defense as it iterates through steps V1 repeatedly changes the covariance structure is an additional challenge.

As argued, the effectiveness of the covariance defense hinges on finding vectors $v$ with large support, so that the term $\lambda v v^T$ is likely to correlate buses inside and outside the (unknown) attacked zone $\mathcal{A}$. To this effect, a vector $v \in \mathbb{R}^n$ satisfying V1(a)-(b) will be called useful. We first have:

**Lemma 14.** Suppose $v^1$ and $v^2$ are useful. Then there exist infinitely many scalars $\lambda$ such that $v^1 + \lambda v^2$ is useful and with support equal to the union of supports of $v^1$ and $v^2$.

**Proof.** Suppose $v^1_k \neq 0$ or $v^2_k \neq 0$. Then $v^1_k + \lambda v^2_k = 0$ for at most one value $\lambda$.

Second, given parameters $w_1, \ldots, w_n$, consider the LP:

$$\max \sum_{k} w_k v_k, \quad \text{s.t. } v \text{ satisfies V1(a)-(b), } |v_k| \leq 1, \forall k$$

Appropriate choices of parameters $w_k$ will encourage dense (large support) vectors; in particular by iterating through several choices of the $w_k$ one can encourage different indices $k$ to appear in the support of $v$. In the above attack example using case2746wp, this idea yields a family of 10 useful vectors whose support covers the set of all buses. Using Lemma 14 we obtain one useful vector $v$ with $v_k \neq 0$ for all $k$.

**VIII. ACKNOWLEDGMENT**

This work was supported by DOE award GMLC77; grants HDTAR1-13-1-0021, ONR N00014-16-1-2889, a DARPA RADICS award, and a DARPA Lagrange award.

**REFERENCES**

[1] Y. Liu, P. Ning, and M. K. Reiter, “False data injection attacks against state estimation in electric power grids,” in *Proceedings of the 16th ACM Conference on Computer and Communications Security*, ser. CCS ’09. New York, NY, USA: ACM, 2009, pp. 21–32. [Online]. Available: http://doi.acm.org/10.1145/1653662.1653666

[2] 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, June 2011.

[3] O. Vukovic, K. C. Sou, G. Dan, and H. Sandberg, “Network-aware mitigation of data integrity attacks on power system state estimation,” *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 6, pp. 1108–1118, July 2012.

[4] S. Li, Y. Yilmaz, and X. Wang, “Quickset detection of false data injection attack in wide-area smart grids,” *IEEE Transactions on Smart Grid*, vol. 6, no. 2, pp. 2725–2735, Nov 2015.

[5] J. Kim and L. Tong, “On topology attack of a smart grid: Undetectable attacks and countermeasures,” *IEEE Journal on Selected Areas in Communications*, vol. 31, pp. 1294 – 1305, 07/2013 2013.

[6] J. Kim, L. Tong, and R. J. Thomas, “Subspace methods for data attack on state estimation: A data driven approach,” *IEEE Transactions on Signal Processing*, vol. 63, no. 5, pp. 1102–1114, March 2015.

[7] G. Dan and H. Sandberg, “Stealth attacks and protection schemes for state estimators in power systems,” in *2010 First IEEE International Conference on Smart Grid Communications*, Oct 2010, pp. 214–219.

[8] D. Deka, R. Baldick, and S. Vishwanath, “Data attacks on power grids: Leveraging detection,” in *2015 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (IGT)*, Feb 2015, pp. 1–5.

[9] J. Valenzuela, J. Wang, and N. Bissinger, “Real-time intrusion detection in power system operations,” *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 1052–1062, May 2013.

[10] A. Anwar, A. N. Mahmood, and Z. Tari, “Ensuring data integrity of opf module and energy database by detecting changes in power flow patterns in smart grids,” *IEEE Transactions on Industrial Informatics*, vol. 13, no. 6, pp. 3299–3311, Dec 2017.

[11] R. Moslemi, A. Mesbahi, and J. M. Velni, “A fast, decentralized covariance selection-based approach to detect cyber attacks in smart grids,” *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 4930–4941, Sept 2018.

[12] S. Mousavian, J. Valenzuela, and J. Wang, “Real-time data reacquisition in electrical power systems based on artificial neural networks,” *Electric Power Systems Research*, vol. 96, pp. 285 – 295, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0378779612003483

[13] S. Soltan, M. Yannakakis, and G. Zussman, “Power Grid State Estimation Following a Joint Cyber and Physical Attack,” *IEEE Transactions on Control of Network Systems*, vol. PP, no. 99, pp. 1–1, 2016.

[14] S. Soltan and G. Zussman, “Power Grid State Estimation after a cyber-physical attack under the AC power flow model,” in *Proc. IEEE PES-GM’17*, 2017.

[15] ——, “EXPOSE the Line Failures following a Cyber-Physical Attack on the Power Grid,” to appear, *IEEE Transactions on Control of Network Systems*, 2018.

[16] S. Soltan, M. Yannakakis, and G. Zussman, “REACT to Cyber Attacks on Power Grids,” to appear, *IEEE Transactions on Network Science and Engineering*, 2018.

[17] D. K. Molzahn and J. Wang, “Detection and characterization of intrusions to network parameter data in electric power systems,” *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 1942–1946, Oct 2018.

[18] L. Xie, Y. Chen, and P. R. Kumar, “Dimensionality reduction of synchrophasor data for early event detection: Linearized analysis,” *IEEE Transactions on Power Systems*, vol. 29, no. 6, pp. 7284–7294, Nov 2014.

[19] X. A. Liu, D. Laverty, and R. Best, “Islanding detection based on probabilistic pca with missing values in pmu data,” in *2014 IEEE PES General Meeting — Conference Exposition*, July 2014, pp. 1–6.

[20] Z. Wang, Y. Zhang, and J. Zhang, “Principal components fault location based on wams/pmu measure system,” in *2011 IEEE Power and Energy Society General Meeting*, July 2011, pp. 1–5.
