Power Flow Cyber Attacks and Perturbation-Based Defense

Katherine R. Davis*, Kate L. Morrow†, Rakesh Bobba†, Erich Heine†
*PowerWorld Corporation
†Information Trust Institute, University of Illinois at Urbana-Champaign
kate@powerworld.com; {morrow4, rbobba, eheine}@illinois.edu

***Abstract*—In this paper, we present two contributions to false data injection attacks and mitigation in electric power systems. First, we introduce a method of creating unobservable attacks on the AC power flow equations. The attack strategy details how an adversary can launch a stealthy attack to achieve a goal. Then, we introduce a proactive defense strategy that is capable of detecting attacks. The defense strategy introduces known perturbations by deliberately probing the system in a specific, structured manner. We show that the proposed approach, under certain conditions, is able to detect the presence of false data injection attacks, as well the attack locations and information about the manipulated data values.**

I. Introduction

False data injection attacks in the power grid have recently begun to attract wide-spread interest. The role of data in power systems analysis, and hence the importance of studying false data injection attacks, is understood by considering the data pipeline from the measurement devices in the field to the point-of-end-use applications. Devices such as relays and remote terminal units (RTUs) perform the physical data acquisition and are connected through a Supervisory Control and Data Acquisition (SCADA) system. Communication networks then transmit this data which is ultimately processed by power applications, such as those in an energy management system (EMS) at a control center. A taxonomy of dependencies between the communications and the physical power infrastructures is given in [1].

At the application level, maliciously injected data can be a serious threat. The problem, from an operational reliability standpoint [2], [3], arises when the operator sees the effects of the injected false data (what the attacker wants him or her to see), yet is unaware that there is anything wrong. The operator is then in a position where he or she is unknowingly making decisions and taking control actions based on the false state of the system [4], [5], [6]. The importance of situational awareness is explicitly revealed in the reports of investigations following the August 14, 2003 blackout [7] and the more recent southwest outage [8].

A necessary enabler of designing protection and detection mechanisms is the ability to sufficiently characterize how attacks might look in a real power system and how they might be achieved from a practical standpoint.

In [9], a *probing* approach is proposed for detecting false data injection attacks. The approach relies on perturbing the power system by leveraging D-FACTS devices to change the impedance on a set of chosen lines in order to create observable changes in the system, which an adversary is unable to anticipate. The level of perturbation needed to thwart attacks and the deviation from optimal caused by the perturbations were also investigated.

TABLE I
Notation

$\mathbf{a}$	Attack vector, change in measurements
G	Set of attacker goals
$\overline{\mathbf{H}}$	Measurement Jacobian
$\mathbf{Y_{bus}} = \mathbf{G} + j\mathbf{B}$	System admittance matrix
$\mathbf{s_{\Theta,V}} = [\mathbf{\Theta}, \mathbf{V}]$	System state
$\mathbf{P}$	Real power injection
$\mathbf{Q}$	Reactive power injection
$\mathbf{f_{pq}}$	Mismatch of withdrawn and injected power
n	Number of buses
$\{PV\}$	Set of PV buses
$\{V\Theta\}$	Set of $V\Theta$ buses
$\mathbf{J}$	Power flow Jacobian
$\mathbf{m_i}$	Actual measurement state
$\tilde{\mathbf{m}}_\mathbf{i}$	Observed measurement state
$\mathbf{d_i}$	Probe to transition to $\mathbf{m_i}$
k_i	Controller change for $\mathbf{d_i}$

This work builds upon [9] in two directions. First, we introduce a simple algorithm for creating targeted unobservable false data injection attacks based on the AC power flow equations. Secondly, we introduce a perturbation-based detection strategy and show its ability to identify false data injection attacks by utilizing the value of $data-over-time$ (i.e., incorporating information from multiple snapshots).

The rest of this paper is organized as follows. Section II provides a brief background on false data injection attacks, with specific focus on unobservable attacks. A method of creating unobservable false data injection attacks based on the AC power flow equations is presented in Section III. The attack assumes the adversary has control over a limited set of meters. Furthermore, the adversary needs only local information to achieve the attack. Then, in Section IV, we formulate an intentional perturbation (probing) strategy for attack detection. The result is that by enacting a sequence of probes, it becomes possible under certain conditions to detect that an attack is occurring, identify which meters are involved, and identify changes in injected data values.

II. On Observability, Stealth Attacks, and Compromised Meters

Table I shows the notation used in this work.

A. Stealth Attacks

Unobservable or stealth attacks can be defined as data injection attacks which completely satisfy the system model equations. Thus, the observed measurements raise no alarms. For this to be possible in the power grid, it means that multiple solutions that satisfy the equations must exist for the system. Power systems are nonlinear systems which may have multiple or no solutions. The adversary must be able to compute (and inject) one of the solutions that is different from the true solution of the actual system. The surface of solutions for the DC power flow model follows a linear mapping, and this solution space has been fairly well characterized and explored in the literature (*e.g.,* [10], [11], [12]).

Due to their adherence to the power system measurement model equations, unobservable attacks are a notably severe class of false data injection attack. Typical residual-based bad data detection is not sufficient against stealth attacks.

A useful characterization of stealth attacks is provided in [11], where the general formulation can be summarized as an optimization problem to minimize the p-norm of an attack vector $\mathbf{a}$ such that the attack satisfies a set of goals and such that the attack is unobservable. This way of expressing the problem is general enough to encompass the false data injection attacks of interest.

In particular, in equation (1), repeated from [11], it is assumed that the attacker knows the measurement model $\overline{\mathbf{H}}$, the attacker wishes to find an attack vector $\mathbf{a}$ such that $\mathbf{a} \in Im(\overline{\mathbf{H}})$ (the attack is stealthy at a linearized level), and the attack is to satisfy a set of goals ($\mathbf{a} \in G$).

$$\begin{aligned} \min_{\mathbf{a}} \quad & \|\mathbf{a}\|_p \\ \text{s.t.} \quad & \mathbf{a} \in G, \mathbf{a} \in Im(\overline{\mathbf{H}}) \end{aligned} \tag{1}$$

More recently, [13] extends the characterization of unobservable attacks, expanding on the work in [14]. In particular, [14] characterizes the relationship between power system observability (i.e. the ability to estimate the state from a given set of measurements) and coordinated false data injection attack detectability. This concept, the basic correlation between state observability and attack detectability underlies the foundation of the entire notion of unobservable attacks that has developed into a topic of great interest to both the cyber and power communities.

Teixeira et. al. [15] arrive at an interesting practical result which is that under certain conditions, when attack vectors are chosen based on a linearized power flow model, (i.e. the DC model), and the attacks are chosen in accordance with a particular goal, the attack will succeed and be unobservable to a state estimator. The authors show this experimentally using the state estimator in MATPOWER [16], a toolbox for MATLAB. The unobservability of the attacks may hold even as the scaling of the attack vector is significantly increased.

More recently, Jia et. al. [17] demonstrate that unobservable attacks designed for the linear model are not necessarily effective when applied to a nonlinear system. The work compels the need to study attacks designed specifically for the nonlinear AC power flow equations and move away from the potentially misleading DC simplifications. Coincidentally, then, [17] provides motivation for the AC attacks presented in this paper.

Many countermeasures to false data injection attacks have been proposed (*e.g.,* [14], [10], [13], [9]) in the literature. While most such defenses rely on leveraging power sensor information, [18] fuses cyber intrusion detection system (IDS) alerts and power system measurements to detect false data injection and identify modified measurements.

As the grid becomes more 'intelligent' and more potential vulnerabilities are introduced, concerns regarding false data injection and their analysis will only grow more prominent. Thus, being able to model, study, and analyze false data injection attacks to understand the repercussions in a real system is of paramount importance. It remains an open question to ascertain the extent to which the threat of these attacks is real versus how much is perceived.

This work is a step towards providing an answer to that question. A better understanding of the attack space helps us as the defenders of the power grid work towards providing appropriate mitigation strategies based on the determined threat level and based on the characteristics, or signatures, of the types of attacks deemed of interest to pursue.

An advantage that the defenders of the system can expect is that the attacker has limited resources (access to meter data), whereas we can access measurements (even though they might be bad measurements) and control devices anywhere on the system.

III. Power Flow Attacks

This work extends [9] which alludes to the presence of the attacks described in this paper but does not actually present them. In [9], we merely assume that such attacks are possible. In the TCIPG testbed [19], we were able to demonstrate the attacks developed in this paper. This section provides the mathematical background and basis for the attacks.

Algorithm 1 summarizes the steps of the *attack power flow*. The attack power flow provides a recipe for a malicious intruder to modify measurements such that power balance is still satisfied. Such attacks are inherently unobservable in the AC power system model.

A. AC Power Flow Solution

The objective of the standard power flow problem is to solve for the state of the system such that the system satisfies real and reactive power balance. The power injected into each bus by aggregate generation and load must equal the power withdrawn by any connected lines and shunts. Solution of these equations is a well-known and well-studied nonlinear problem [20], [21]. Real and reactive power flows depend on the $state$, the voltage magnitude and angle, yet the state is unknown and must be found such that power balance is satisfied throughout the system. The power flow solution enforces the conservation of power.

Let $\mathbf{V}$ be the vector of voltage magnitudes and $\mathbf{\Theta}$ be the vector of voltage angles. The vector of real power net injections (generation minus load at each bus) is $\mathbf{P}$ and the vector of reactive power net injections is $\mathbf{Q}$. Each bus thus has four quantities: V_i, θ_i, P_i, and Q_i.

The power flow finds the set of voltages and angles that satisfy the conservation of power for a given set of net power injections at the buses. The system state $\mathbf{s}$ may be written as $\mathbf{s_{\Theta,V}} = [\mathbf{\Theta}, \mathbf{V}]^{\mathbf{T}}$, and the power flow or the power balance equations are given by the following:

$$P_i^{calc} = \sum_{k \in C} |V_i||V_k|(G_{ik}\cos\theta_{ik} + B_{ik}\sin\theta_{ik}) \quad (2)$$

$$Q_i^{calc} = \sum_{k \in C} |V_i||V_k|(G_{ik}\sin\theta_{ik} - B_{ik}\cos\theta_{ik}) \quad (3)$$

$$\Delta p_i = P_i^{calc} - P_i \quad (4)$$

$$\Delta q_i = Q_i^{calc} - Q_i \quad (5)$$

$$\mathbf{f_{pq}}(\mathbf{s_{\Theta,V}}) = [\Delta\mathbf{p}, \Delta\mathbf{q}] \quad (6)$$

where (6) is zero at solution and where the system admittance matrix is defined as $\mathbf{Y_{bus}} = \mathbf{G} + j\mathbf{B}$ [21].

The power flow problem may be written succinctly as

$$\mathbf{f_{pq}}(\mathbf{s_{\Theta,V}}) = \mathbf{0} \quad (7)$$

where finding its solution is at the heart of most power systems analysis. Its solution is typically found using the Newton-Raphson method, an iterative technique that requires multiple evaluations and factorizations of a large sparse matrix of sensitivities, the Jacobian matrix $\mathbf{J}$.

B. Slack, PV, and PQ buses

The above discussion is somewhat simplified and ignores a few details which are important to talk about here to help explain our attack recipe in the next section.

Consider that the only unknowns in the power flow equations might not be the system state, $\mathbf{s_{\Theta,V}}$. Often, load buses are represented in the power flow solution by constant power terms P and Q and are referred to as PQ buses. Generator buses often have voltage regulation to maintain some terminal voltage, V_{spec}. Thus, they are modeled as constant P and constant V buses, called PV buses. For PV buses, the voltage angle θ must still be solved for, but rather than solve for V which is specified, we solve for its Q such that the desired V is obtained. One bus is designated the $slack$ or the $swing$ bus. This accounts for the fact that absolute angles are not meaningful. Everywhere in the power flow equations, what actually appear are always only angle differences. Angles are only meaningful with respect to a reference.

The $slack$ bus is the bus which serves as an angle reference. In practice, the slack bus is chosen to be a large generator bus and its state variables θ and V are fixed values in the power flow. Its P and Q values are computed only after the power flow solution is complete. Hence, the $slack$ bus gets its name because it picks up the slack or the losses of the system. Line losses act as an unspecified load, where the value of the losses are not known until after solution for the state $\mathbf{s_{\Theta,V}}$.

TABLE II
TRUE DATA VALUES

Bus Num.	V(pu)	Theta(deg)	Pinj(pu)	Qinj(pu)
1	1.04	0	0.717	0.270
2	1.025	9.279	1.63	0.066
3	1.025	4.664	0.85	-0.109
4	1.026	-2.217	0.0	0.0
5	0.996	-3.989	-1.25	-0.5
6	1.013	-3.688	-0.9	-0.3
7	1.026	3.718	0.0	0.0
8	1.016	0.726	-1.0	-0.35
9	1.032	1.966	0.0	0.0

C. Attack Power Flow Recipe

The attack power flow solution determines the data injections needed to achieve a target while satisfying the underlying system model (power balance) and therefore making the attack unobservable.

In some ways, the attack power flow is the reverse of the normal power flow. The normal power flow designates one slack bus and solves for the states at all other buses. By contrast, the attack power flow assumes that states at all unmodifiable buses are fixed and solves for the value modifications necessary to make that true.

To perform an attack, we set all of the unmodifiable states to be fixed $V\theta$ buses. This holds V and θ fixed at that bus. The for loops (see lines 7 and 17) in Algorithm 1 calculate real and reactive power mismatch between injected and withdrawn power at a node. The mismatch is used in each Newton-Raphson iteration to update the value of the state. The algorithm exits when power balance is satisfied for the system (i.e., the norm of the mismatch is less than the tolerance; see line 30).

Consider an attack scenario where the attacker is targeting a single bus. In this case, all buses should be made $V\theta$ buses except the target bus. The attack is now described in context of a Western Systems Coordinating Council (WSCC) 9-bus test system [22]. In this example, the scenario is that the attacker modifies the data such that the system appears to be experiencing an overload. In reality, the system is fine, but the attacker is trying to entice us to shed load.

The system has the true data values shown in Table II. The attacker chooses false-data injection values for modification. In this example, the attacker's goal is to spoof the data to make it look like Q_6 increased by 100 MVAr. When this is accomplished via Algorithm 1, we (the operators) see the numbers in Table III.

The summary of the attack scenario is this. By manipulating the measurements at bus 6, the attacker is able to (1) cause the voltage at bus 6 to appear high (1.07 per unit instead of 1.01) thereby causing us to consider taking some control actions, like shedding load, and (2) the attacker's changes are consistent with the power flow solution, so it will be undetected using conventional residual-based bad data detection [23]. Note that the attacker also necessarily modifies a couple of measurements at bus 4 and bus 9; these are the boundary

TABLE III
ATTACK POWER FLOW VALUES

Bus Num.	V(pu)	Theta(deg)	Pinj(pu)	Qinj(pu)
1	1.04	0	0.716	0.270
2	1.025	9.280	1.63	0.066
3	1.025	4.664	0.850	-0.109
4	1.026	-2.217	-0.001	-0.640
5	0.996	-3.989	-1.25	-0.500
6	1.070	-4.173	-0.9	0.7
7	1.026	3.720	0.000	0
8	1.016	0.727	-1	-0.350
9	1.032	1.966	0.010	-0.345

buses to bus 6. This attack scenario is demonstrated by the authors in the video [19].

What this effect, of needing to manipulate the boundary bus values, represents is that the attacker must be able to defeat the defender's attempt to protect *basic measurements*. As proven in [14] and shown in various ways throughout the recent literature on defenses against false data injection, as long as a defender is able to protect a set of *basic measurements*, that is, those that are needed for observability of the system, unobservable attacks such as those presented here are not possible. Hence, as an attacker, the goal would be to violate this protection mechanism. This is a fundamental principle related to power system observability and the physical characteristics of the system: any attack which is truly unobservable (stealthy) $must$ cause the defender to violate the assumption of protected basic measurements. A state is observable if there is a measurement equation which provides enough information to estimate the state. Hence, if all measurements which reflect the value of the state variable are compromised, then the attacker has complete control over that state.

D. Attacker Knowledge and Discussion

The strategy presented creates unobservable attacks on the AC power flow equations and presents a specific way of implementing these attacks. The inputs to Algorithm 1 are the adversary's knowledge of the system topology $\mathbf{Y_{bus}}$, the system state $\mathbf{s_{\Theta,V}}$, and the power injections $\mathbf{P_0}$, $\mathbf{Q_0}$. The attacker is free to make assumptions about these values. Here, we examine the minimum knowledge required by an attacker to perform the attack.

Let us denote as *target buses* the buses where we focus our P and Q injection changes (i.e., bus 6 in our example). Let us denote as *boundary buses* all buses adjacent to the target buses (i.e., buses 4 and 9). All non-boundary and non-target buses are assumed to be unmodifiable.

Fixed $V\theta$ buses have no entries in the power flow Jacobian, and all buses are fixed $V\theta$ except the target buses. The only equations which must be solved are the P and Q equations at the target buses. The solution gives the state variables V and θ at the target buses. Then, by inspection of the power flow equations, the only buses whose P and Q equations depend on V and θ at the target buses are the boundary buses. For this reason, it is necessary that the attacker be able to modify the power injection values at the boundary buses.

In summary, the only measurements needed when calculating an unobservable attack are measurements at the buses adjacent to the target buses. Therefore, in order to perform an unobservable attack, the adversary does not need to know everything about the system, but only needs to know local information (both measurements and topology) only one layer out from the target buses.

Algorithm 1: Attack Power Flow Algorithm

Input: [$\mathbf{Y_{bus}}$, $\mathbf{V_0}$, $\mathbf{\Theta_0}$, $\mathbf{P_0}$, $\mathbf{Q_0}$, $\{V\Theta\}$, $\{PV\}$]
Output: [$\mathbf{V_f}$, $\mathbf{\Theta_f}$, $\mathbf{P_f}$, $\mathbf{Q_f}$]
1 nFixedThetas = size$\{V\Theta\}$;
2 nFixedVs = size$\{PV\}$;
3 nPQ = n - nFixedVs - nFixedThetas;
4 nThetaStates = n - nFixedThetas;
5 nVStates = n - nFixedVs;
6 **for** *iter = 1 : maxIter* **do**
7 **for** *i = 1 : nThetaStates* **do**
8 si = i + $nFixedThetas$;
9 pbr = 0;
10 **for** *j = 1 : n* **do**
11 sj = j;
12 pbr = pbr + CalcPflow($\mathbf{Y_{bus}}(si, sj), V_{si}, \theta_{si}$) , si $\neq$ sj ;
13 **end**
14 $\mathbf{pfcalcI}(i)$ = CalcPShunt($\mathbf{Y_{bus}}(si, si), V_{si}, \theta_{si}$) + pbr;
15 **end**
16 $\mathbf{pfI}$ = $\mathbf{pfcalcI}$ - $\mathbf{P_0}$;
17 **for** *i = 1 : nVStates* **do**
18 si = i + $nFixedThetas$ + $nFixedVs$;
19 qbr = 0;
20 **for** *j = 1 : n* **do**
21 sj = j;
22 qbr = qbr + CalcQflow($\mathbf{Y_{bus}}(si, sj), V_{si}, \theta_{si}$) , si $\neq$ sj ;
23 **end**
24 $\mathbf{qfcalcI}(i)$ = CalcQShunt($\mathbf{Y_{bus}}(si, sj), V_{si}, \theta_{si}$) + qbr;
25 **end**
26 $\mathbf{qfI}$ = $\mathbf{qfcalcI}$ - $\mathbf{Q_0}$;
27 $\mathbf{f_{pq}}$ = [$\mathbf{pfI}$; $\mathbf{qfI}$];
28 [$\mathbf{J(s)}$] = CalcJac($\mathbf{Y_{bus}}, \mathbf{s_{V\Theta}}$, nFixedVs, nFixedThetas);
29 $\mathbf{s_{V\Theta}}$ = $\mathbf{s_{V\Theta}} - [\mathbf{J(s)}]^{-}1 \cdot \mathbf{fpq}$;
30 **if** $||\mathbf{f_{pq}}|| < tol$ **then**
31 Attack power flow is solved;
32 $break$;
33 **end**
34 **end**
35 Calculate $\mathbf{P_f}$ and $\mathbf{Q_f}$ for all buses (see lines 7 - 26);
36 Return the results, [$\mathbf{V_f}$, $\mathbf{\Theta_f}$, $\mathbf{P_f}$, $\mathbf{Q_f}$];

IV. PROBING DEFENSE

The probing defense described in this paper uses an observe and perturb methodology to compare the expected results of a control action with the observed response of the system. The method uses only observed data values, and the expected results of the control actions can be determined using historical data, if available. The goal is to utilize our perceived but incorrect measurement state of the system to discern the spoof.

The approach in [9] is to change settings on Distributed Flexible AC Transmission System (D-FACTS) devices [24], [25] to cause impact on measurements, but any available control devices may be used (tap changing transformers, generators, switched shunts, etc.). We can obtain predictions of the impact of the probe on our measurements. The fundamental idea is that if the measurements do not change in the way we expect, there may be cause for alarm, and further investigation is warranted. The vector of settings for each device is termed

a key. This key is chosen from a set of acceptable keys. Keys, keyspaces, and key choice are explored further in [9].

A. Assumptions

The adversary is assumed to be modifying measurement values, but no assumptions are made about how this is done. We do not consider whether the attack occurs through a compromise of the data channel, the device, the control center, etc. The control devices are assumed to already be in place in the system for another purpose (e.g., loss minimization). The defenders are assumed to have control over the devices. We consider that the attack vector may be different at each time. The keys are pre-determined values, required to produce measurement changes that are large enough to be distinguishable from system noise.

We assume for simplicity that there are no other measurable events while conducting the series of probes, e.g., load and generation remain reasonably constant. Accounting for load or generation changes or other events while conducting a sequence is outside the scope of the current work, but is a future research direction.

B. Observations of Measurement States

In discussing the defense, a different definition of 'state' is needed from the power flow state described in Section III. We denote the *measurement state* of the system under key k_i as $\mathbf{m_i}$, which is a vector of measurements corresponding to a single snapshot of the system which includes a specific power flow state and topology. This can be a vector of real and reactive line flows. In the rest of the discussion, the measurement state is referred to only as the state.

Let the *true* initial state of the system be $\mathbf{m_0}$. The *observed* initial state of the system, $\tilde{\mathbf{m}}_\mathbf{0}$, may not be the same as the true initial state, since an attack may or may not be present. From $\mathbf{m_0}$, the first probe is enacted, causing the system to achieve its new true state

$$\mathbf{m_1} = \mathbf{d_1}(\mathbf{m_0}) \tag{8}$$

where $\mathbf{d_1}$ is the updated solution to the nonlinear power flow equations ((2), (3)) based on the application of key k_1. In this work, we use $\mathbf{d_i}$ to model the behavior of the true (and possibly unobservable) physical system under perturbation.

This true state $\mathbf{m_1}$ is unknown to us. The state we observe after applying the probe is given by:

$$\begin{aligned} \tilde{\mathbf{m}}_\mathbf{1} &= \mathbf{d_1}(\mathbf{m_0}) + \mathbf{a_1} \\ &= \mathbf{m_1} + \mathbf{a_1} \end{aligned} \tag{9}$$

where $\mathbf{a_1}$ denotes the attacker's spoof applied to the system while in state $\mathbf{m_1}$. Note that if an attack is present, the actual observed state after applying the probe is different from the expected state calculated with the probe.

Each spoofed vector $\mathbf{a_i}$ can only contain alterations to meters within the set of the attacker's controlled meters, M_α. Elements in the vector where the attacker does not have control are indicated by subscript j:

$$\mathbf{a_i}_j = 0 \ \forall \ j \notin M_\alpha \tag{10}$$

Our problem is to determine M_α (which set of meters the attacker is controlling), and to a lesser extent, the attack injections ($\mathbf{a_0}$, $\mathbf{a_1}$) (what was the attacker doing).

At each time snapshot, the system has a measurement state $\mathbf{m_i}$. The measurement state resulting from subsequent probe k_j is $\mathbf{m_j}$. The corresponding observed states are $\tilde{\mathbf{m}}_\mathbf{i}$ and $\tilde{\mathbf{m}}_\mathbf{j}$. This results in the following set of equalities:

$$\mathbf{m_j} = \mathbf{d_j}(\mathbf{m_i}), \ \forall \ j \tag{11}$$

$$\tilde{\mathbf{m}}_\mathbf{j} = \mathbf{d_j}(\mathbf{m_i}) + \mathbf{a_j} = \mathbf{m_j} + \mathbf{a_j}, \ \forall \ j \tag{12}$$

We expect that $\mathbf{d_j}(\mathbf{m_0}) = \mathbf{d_j}(\mathbf{m_i}) \ \forall \ i, j$, which relies on the assumption that the underlying state of the system is not changing. That is, we are not experiencing changes in load or generation, or topology other than our perturbations.

C. Attack Detection

If a probing sequence is enacted such that the system is made to return to a previously observed state, the difference between the expected result of the probe and the actual perceived effect of the probe will reveal the existence of bad data and its locations.

To see how this is possible, consider the system, its true values, and its perceived values at each time instant or probing snapshot. The examples for the first probe are given above (9), and the equations for each additional probe can be written in a similar way.

Now, consider the possibility that the attack vector is not solely dependent on the state, but could vary across different visits to a particular state. That is, the attack vectors for an observed state $\tilde{\mathbf{m}}_\mathbf{i}$ may be different when that state is observed at different times. This means that the attacker's injections may change from $\mathbf{a_i}_\alpha$ to $\mathbf{a_i}_\beta$, where α denotes one transition to state $\mathbf{m_i}$ and β denotes a subsequent transition. Then, let the actual perceived measurement state at transition α be $\tilde{\mathbf{m}}_{\mathbf{i},\alpha}$, and let the expected measurement state due to a previous transition β be $\tilde{\mathbf{m}}_{\mathbf{i},\beta}$. The difference can be written as follows,

$$\mathbf{\Delta}_{\tilde{\mathbf{i}}} = \tilde{\mathbf{m}}_{\mathbf{i},\alpha} - \tilde{\mathbf{m}}_{\mathbf{i},\beta} \tag{13}$$

which is equal to

$$\begin{aligned} \mathbf{\Delta}_{\tilde{\mathbf{i}}} &= \mathbf{m_i} + \mathbf{a}_{\mathbf{i},\alpha} - \mathbf{m_i} + \mathbf{a}_{\mathbf{i},\beta} \\ &= \mathbf{a}_{\mathbf{i},\alpha} - \mathbf{a}_{\mathbf{i},\beta} \end{aligned} \tag{14}$$

where $\mathbf{m_i}$ is a state which has previously been visited. Here, we take advantage of the fact that we are probing the system to states we have been to before. The observed quantity (14) directly tells us the difference in attack vectors at the returned-to states.

D. Discussion

We note a few observations from these results. First, based only on our observed quantities, after actuating a careful sequence of probes and observing the responses, it may be possible to isolate the source of the false data injection attack. A fundamental caveat is obvious from examination of equation (14). In this formulation, if $\mathbf{a}_{\mathbf{i},\alpha}$ and $\mathbf{a}_{\mathbf{i},\beta}$ are equal, the attack remains undetectable. This result seems to indicate that if it is possible for an attacker to always maintain a specific attack for a given state, this defense will not work. In that case, there is really no incentive for an attacker to change his or her attack vector unless unless it is necessary in order to launch stealthy attacks. However, it may still be possible for the defenders to unmask those attacks. For example, rather than using $\mathbf{a}_{\mathbf{i},\beta}$ (a single observation) as our expected value, we can use a weighted average over many previous observations. Multiple targeted probing sequences can be initiated to obtain more sets of data.

One important point is that the probes do not need to be from D-FACTS devices. Nowhere in this analysis is the presence of D-FACTS included or shown to be necessary. Any probing mechanism which is capable of producing observable changes in the system can be used as part of the probing defense strategy. The need for measurable perturbations resulting from a probe is examined in [9]. Finally, if an attack is not stealthy, this defense strategy may not be needed at all, as the attack could readily be detected by other more conventional means.

V. Conclusions

In this paper, we have shown how to launch unobservable attacks against the AC power flow equations using only local information. We have also proposed a particular probing defense strategy to detect attacks by enacting sequences of probes and keeping track of the observed quantities. The defense is independent of the type of probe as well as the type of attack. Future work will explore the effectiveness of the attacks and the probing defense, as the adversary's view of the system is varied.

Acknowledgement

This material is based upon work supported in part by the Department of Energy under Award Number DE-OE0000097, and by the Office of Naval Research under agreement Navy N00014-10-1-0818. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) alone.

References

[1] S. Sridhar, A. Hahn, and M. Govindarasu, "Cyber-physical system security for the electric power grid," *Proceedings of the IEEE*, vol. 100, no. 1, pp. 210 –224, Jan. 2012.

[2] North American Electric Reliability Corporation, "Reliability standards for the bulk power systems of North America," [Online]. Available: http://www.nerc.com/files/Reliability_Standards_Complete_Set_21Jul08.pdf, July 2008.

[3] N. Balu, T. Bertram, A. Bose, V. Brandwajn, G. Cauley, D. Curtice, A. Fouad, L. Fink, M. Lauby, B. Wollenberg, and J. Wrubel, "On-line power system security analysis," *Proceedings of the IEEE*, vol. 80, no. 2, pp. 262 –282, Feb 1992.

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

[5] L. Jia, R. J. Thomas, and L. Tong, "Impacts of malicious data on real-time price of electricity market operations," in *Hawaii International Conference on System Sciences*. IEEE Computer Society, 2012, pp. 1907–1914.

[6] A. Teixeira, H. Sandberg, G. Dan, and K.-H. Johansson, "Optimal Power Flow: Closing the Loop over Corrupted Data," in *American Control Conference (ACC)*, June 2012.

[7] U.S.-Canada Power System Outage Task Force, "Final Report on the August 14, 2003 Blackout in the United States and Canada: Causes and Recommendations," [Online]. Available: https://reports.energy.gov/BlackoutFinal-Web.pdf, April 2004.

[8] FERC/NERC Staff, "Arizona-Southern California Outages on September 8, 2011, Causes and Recommendations," [Online]. Available: http://www.nerc.com/fileUploads/File/News/AZOutage_Report_01MAY12.pdf, May 2012.

[9] K. L. Morrow, E. Heine, K. M. Rogers, R. B. Bobba, and T. J. Overbye, "Topology perturbation for detecting malicious data injection," in *Hawaii International Conference on System Sciences*. IEEE Computer Society, 2012, pp. 2104–2113.

[10] O. Kosut, L. Jia, R. J. Thomas, and L. Tong, "Malicious data attacks on smart grid state estimation: Attack strategies and countermeasures," in *First IEEE International Conference on Smart Grid Communications (SmartGridComm), 2010*. IEEE, 2010.

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

[12] Y. Liu, P. Ning, and M. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Transactions in Information and Systems Security (TISSEC), 2011*, vol. 14, no. 1, pp. 13:1–13:33, June 2011.

[13] A. Giani, E. Bitar, M. Garcia, M. McQueen, P. Khargonekar, and K. Poolla, "Smart grid data integrity attacks: characterizations and countermeasures," in *IEEE International Conference on Smart Grid Communications (SmartGridComm), 2011*, Oct. 2011, pp. 232 –237.

[14] 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 *1st Workshop on Secure Control Systems (SCS '10)*, 2010.

[15] A. Teixeira, G. Dan, 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*, August 2011.

[16] R. Zimmerman, C. Murillo-Sanchez, and R. Thomas, "Matpower: Steady-state operations, planning, and analysis tools for power systems research and education," *Power Systems, IEEE Transactions on*, vol. 26, no. 1, pp. 12 –19, feb. 2011.

[17] J. Liyan, R. J. Thomas, and L. Tong, "On the nonlinearity effects on malicious data attack on power system," in *To appear in 2012 Power and Energy Society general meeting*, July 2012.

[18] S. Zonouz, K. M. Rogers, R. Berthier, R. B. Bobba, W. H. Sanders, and T. J. Overbye, "SCPSE: Security-Oriented Cyber-Physical State Estimation for Power Grid Critical Infrastructures," *IEEE Trans. Smart Grid*, to appear.

[19] Trustworthy Cyber Infrastructure for the Power Grid, "Research Demos," [Online]. Available: http://tcipg.org/news/research-demos.

[20] A. Wood and B. Wollenberg, *Power Generation, Operation, and Control*, 2nd ed. John Wiley and Sons, 1996.

[21] J. Glover, M. Sarma, and T. Overbye, *Power system analysis and design*. Thomson, 2008.

[22] P. Sauer and M. Pai, *Power system dynamics and stability*. Champaign, IL: Stipes Publishing L.L.C., 1997.

[23] A. Monticelli, *State estimation in electric power systems*. Boston, Mass.: Kluwer Academic Publishers, 1999.

[24] D. Divan, W. Brumsickle, R. Schneider, B. Kranz, R. Gascoigne, D. Bradshaw, M. Ingram, and I. Grant, "A distributed static series compensator system for realizing active power flow control on existing power lines," *IEEE Transactions on Power Delivery*, vol. 22, pp. 642–649, Jan 2007.

[25] K. M. Rogers, "Power system control with distributed flexible ac transmission system devices," Master's thesis, University of Illinois at Urbana-Champaign, 2009.