On Topology Attack of a Smart Grid: Undetectable Attacks and Countermeasures

Jinsub Kim and Lang Tong, *Fellow, IEEE*

***Abstract*—Covert data attacks on the network topology of a smart grid is considered. In a so-called man-in-the-middle attack, an adversary alters data from certain meters and network switches to mislead the control center with an incorrect network topology while avoiding detections by the control center. A necessary and sufficient condition for the existence of an undetectable attack is obtained for strong adversaries who can observe all meter and network data. For weak adversaries with only local information, a heuristic method of undetectable attack is proposed. Countermeasures to prevent undetectable attacks are also considered. It is shown that undetectable attacks do not exist if a set of meters satisfying a certain branch covering property are protected. The proposed attacks are tested with IEEE 14-bus and IEEE 118-bus system, and their effect on real-time locational marginal pricing is examined.**

***Index Terms*—Malicious data attack, cyber physical system security, SCADA system, power system state and topology estimation, bad data detection.**

I. Introduction

A DEFINING feature of a smart grid is its abilities to monitor the state of a large power grid, to adapt to changing operating conditions, and to react intelligently to contingencies, all of which depend critically on a reliable and secure cyber-infrastructure. It has been widely recognized that the heavy reliance on a wide area communication network for grid monitoring and real-time operation comes with increasing security risks of cyber-attacks. See [1] for a vulnerability analysis of energy delivery control systems.

While *information* security has been a major focus of research for over half a century, the mechanisms and the impacts of attack on *cyber physical systems* such as the power grid are not yet well understood, and effective countermeasures are still lacking.

We consider in this paper a form of "man-in-the-middle" (MiM) attack [2] on the topology of a power grid. An MiM attack exploits the lack of authentication in a system, which allows an adversary to impersonate a legitimate participant. In the context of monitoring a transmission grid, sophisticated authentications are typically not implemented due to the need of reducing communication delay and the presence of legacy communication equipment. If an adversary is able to gain access to remote terminal units (RTUs) or local data concentrators, it is possible for the adversary to replace actual data packets with carefully constructed malicious data packets and impersonate a valid data source.

MiM attacks on a power grid may have severe consequences. The adversary can mislead the control center that the grid is operating under a topology different from that in reality. Such an attack, if launched successfully and undetected by the control center, will have serious implications: a grid that is under stress may appear to be normal to the operator thereby delaying the deployment of necessary measures to ensure stability. Similarly, a grid operating normally may appear to be under stress to the operator, potentially causing load shedding and other costly remedial actions by the operator.

Launching a topology attack, fortunately, is not easy; a modern energy management system is equipped with relatively sophisticated bad data and topology error detectors, which alerts the operator that either the data in use are suspicious or there may indeed be changes in the network topology. When there are inconsistencies between the estimated network topology (estimated mostly using switch and breaker states) and the meter data (*e.g.*, there is significant amount of power flow on a line disconnected in the estimated topology,) the operator takes actions to validate the data in use. Only if data and the estimated topology pass the bad data test, will the topology change be accepted and updates be made for subsequent actions.

The attacks that are perhaps the most dangerous are those that pass the bad data detection so that the control center accepts the change (or the lack of change) of network topology. To launch such attacks, the adversary needs to modify simultaneously the meter data and the network data (switch and breaker states) in such a way that the estimated topology is consistent with the data. Such attacks are referred to as *undetectable attacks*; they are the main focus of this paper.

A. Summary of results

Results of this paper aim to achieve two objectives. First, we characterize conditions under which undetectable attacks are possible, given a set of vulnerable meters that may be controlled by an adversary. To this end, we consider two attack regimes based on the *information set* available to the attacker. The more information the attacker has, the stronger its ability to launch a sophisticated attack that is hard to detect.

The *global information* regime is where the attacker can observe all meter and network data before altering the adversary-controlled part of them. Although it is unlikely in practice that an adversary is able to operate in such a regime, in analyzing the impact of attacks, it is typical to consider the worst case

Manuscript received October 8, 2012; revised March 20, 2013. This work is supported in part by the National Science Foundation under Grant CNS-1135844 and the DoE CERTS program. Part of this work was presented at IEEE PES Innovative Smart Grid Technologies Conference, Washington, D.C., February, 2013.

J. Kim and L. Tong are with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853 USA (e-mail: jk752@cornell.edu, ltong@ece.cornell.edu).

Digital Object Identifier 10.1109/JSAC.2013.130712.

by granting the adversary additional power. In Section III, we present a necessary and sufficient algebraic condition under which, given a set of adversary controlled meters, there exists an undetectable attack that misleads the control center with an incorrect "target" topology. This algebraic condition provides not only numerical ways to check if the grid is vulnerable to undetectable attacks but also insights into which meters to protect to defend against topology attacks. We also provide specific constructions of attacks and show certain optimality of the proposed attacks.

A more practically significant situation is the *local information* regime where the attacker has only local information from those meters it has gained control. Under certain conditions, undetectable attacks exist and can be implemented easily based on simple heuristics. We present in Section IV intuitions behind such simple attacks and implementation details.

The second objective is to provide conditions under which topology attack cannot be made undetectable. Such a condition, even if it may not be the tightest, provides insights into defense mechanisms against topology attacks. In Section V, we show that if a set of meters satisfying a certain branch covering property are protected, then topology attacks can always be detected. In practice, protecting a meter may be carried out at multiple levels, from physical protection measures to software protection schemes using more sophisticated authentication protocols.

B. Related works and organization

Liu, Ning, and Reiter [3] appear to be the first to introduce the concept of data injection attack (also referred to as malicious data attack) of a smart grid. Assuming that the attacker is capable of altering data from a set of meters, a similar scenario assumed in this paper, the authors of [3] show that the adversary can perturb the network state by an arbitrarily large amount without being detected by any detector. In other words, the data attack considered in [3] is undetectable. The main difference between [3] and the current paper is that the attacks considered in [3] perturb only the network state, not the network topology. It is thus most appropriate to refer to attacks in [3] and many follow-ups as *state attack*, in distinguishing the *topology attack* considered in this paper.

The work in [3] is influential; it has inspired many further developments, *e.g.*, [4]–[7] and references therein, all focusing on state attacks. A key observation is made by Kosut *et al.* in [8], [9], showing that the condition of non-existence of an undetectable attack is equivalent to that of network observability [10], [11]. This observation leads to graph theoretic techniques that characterize network vulnerability [9]. The condition presented in the current paper on the non-existence of an undetectable topology attack mirrors the state attack counterpart in [9].

The problem of adding protection on a set of meters to prevent undetectable state attacks was considered by Bobba *et al.* [4]. The current paper considers the same problem in the context of topology attack. While meter protection problem for state attacks is equivalent to protecting a sufficient number of meters to ensure observability [4], [9], the corresponding problem for topology attacks is somewhat different and more challenging. See Section V for discussions.

The problem of detecting topology error from meter data is in fact a classical problem, casted as part of the bad data detection problem [12]–[14]. Monticelli [15] pioneers the so-called generalized state estimation approach where, once the state estimate fails the bad data test, modifications of topology that best represent the meter data are considered. Abur *et al.* [16] and Mili *et al.* [17] apply the idea to various state estimation methods. Extensive works followed to improve computational efficiency, estimation accuracy, and convergence property over the aforementioned methods (*e.g.*, see [18]–[20] and references therein).

Finally, there is a limited discussion on the impact of a malicious data attack on power system operations. Should state estimates be used in closed-loop control of the power grid, such an attack may cause serious stability problems. The current state of the art, however, uses state estimates for real-time dispatch only in a limited fashion. However, state estimates are used extensively in calculating real-time locational marginal price (LMP) [21]. Thus, attacks that affect state estimates will affect the real-time LMP calculation [22]–[24]. The way that a topology attack affects LMP is significantly different from that of a state attack. In Section VI, we demonstrate that a topology attack has significant impact on real-time LMP.

The rest of the paper is organized as follows. Section II presents mathematical models of state estimation, bad data test, and topology attacks. In Section III, we study topology attacks in the global information regime. The algebraic condition for an undetectable attack is presented, and construction of a cost-effective undetectable attack is provided. Section IV presents a heuristic attack for the attacker with local information. Based on the algebraic condition presented in Section III, Section V provides a graph theoretical strategy to add protection to a subset of meters to prevent undetectable attacks. Section VI presents simulation results to demonstrate practical uses of our analysis and feasibility of the proposed attacks, and Section VII finishes the paper with concluding remarks.

II. Preliminaries

In this section, we present models for the power network, measurements, and adversary attacks. We also summarize essential operations such as state estimation and bad data detection that are targets of data attacks.

A. Network and measurement models

The control center receives two types of data from meters and sensors deployed throughout the grid. One is the digital network data $\mathbf{s} \in \{0,1\}^d$, which can be represented as a string of binary bits indicating the on and off states of various switches and line breakers. The second type is the analog meter data $\mathbf{z}$, which is a vector of bus injection and line flow measurements.

Without an attack or a sensing error, $\mathbf{s}$ gives the true breaker states. Each $\mathbf{s} \in \{0,1\}^d$ corresponds to a system topology, which is represented by a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of buses and $\mathcal{E}$ is the set of *connected* transmission lines. For each physical transmission line between two buses (*e.g.*, i and j), we assign an arbitrary direction for the line (*e.g.*,

(i,j)), and (i,j) is in $\mathcal{E}$ if and only if the line is connected. In addition, $\mathcal{E}_0$ denotes the set of all lines (with the assigned directions), both connected and disconnected. Assigning arbitrary directions for lines is not intended to deliver any physical meaning, but only for ease of presentation.

The state of a power system is defined as the vector $\mathbf{x}$ of voltage phasors on all buses. In the absence of attacks and measurement noise, the meter data $\mathbf{z}$ collected by the SCADA system are related to the system state $\mathbf{x}$ and the system topology $\mathcal{G}$ via the AC power flow model [25]:

$$\mathbf{z} = h(\mathbf{x}, \mathcal{G}) + \mathbf{e} \tag{1}$$

where $\mathbf{z}$ typically includes real and reactive parts of bus injection and line flow measurements, h is the nonlinear measurement function of $\mathbf{x}$ and $\mathcal{G}$, and $\mathbf{e}$ the additive noise.

A simplified model, one that is often used in real-time operations such as the computation of real-time LMP, is the so-called DC model [25] where the nonlinear function h is linearized near the operating point. In particular, the DC model is given by

$$\mathbf{z} = H\mathbf{x} + \mathbf{e} \tag{2}$$

where $\mathbf{z} \in \mathbb{R}^m$ consists of only the real parts of injection and line flow measurements, $H \in \mathbb{R}^{m\times n}$ is the measurement matrix, $\mathbf{x} \in \mathbb{R}^n$ is the state vector consisting of voltage phase angles at all buses except the slack bus, and $\mathbf{e} \in \mathbb{R}^m$ is the Gaussian measurement noise with a diagonal covariance matrix Σ.

The fact that the measurement matrix H depends on the network topology $\mathcal{G}$ is important, although we use the notation H without explicit association with its topology $\mathcal{G}$ for notational convenience. For ease of presentation, consider the noiseless measurement $\mathbf{z} = H\mathbf{x}$. If an entry z_k of $\mathbf{z}$ is the measurement of the line flow from i to j of a *connected* line in $\mathcal{G}$, z_k is $B_{ij}(x_i - x_j)$ where B_{ij} is the line susceptance and x_i is the voltage phase angle at bus i. The corresponding row of H is equal to

$$\mathbf{h}_{(i,j)} \triangleq [0\cdots 0 \quad \underbrace{B_{ij}}_{i\text{th entry}} \quad 0\cdots 0 \quad \underbrace{-B_{ij}}_{j\text{th entry}} \quad 0\cdots 0]. \tag{3}$$

On the other hand, if z_k is the measurement of the line flow through a *disconnected* line in $\mathcal{G}$, z_k is zero, and the corresponding row of H consists of all zero entries. If z_k is the measurement of bus injection at i, it is the sum of all the outgoing line flows from i, and the corresponding row of H is the sum of the row vectors corresponding to all the outgoing line flows.

In this paper, we consider both AC and DC power flow models. The DC model allows us to obtain a succinct characterization of undetectable attacks as described in Section III. However, these results hold only locally around the operating point, because the results are obtained from the linearized model. See [26] for a more detailed discussion. General results for the more realistic (nonlinear) AC model are difficult to obtain. We present in Section IV a heuristic attack that are undetectable for both AC and DC models.

It was shown in [24] that using the DC model and linear state estimator in numerical analysis of an attack tends to exaggerate the impact of the attack. Hence, for accurate analysis, we use the AC model and nonlinear state estimator in the numerical simulations presented in Section VI.

B. Adversary model

The adversary aims at modifying the topology estimate from $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ to a different "target" topology $\bar{\mathcal{G}} = (\mathcal{V}, \bar{\mathcal{E}})$. Note that $\mathcal{G}$ and $\bar{\mathcal{G}}$ have the same set of vertices. In other words, we only consider the attacks aimed at perturbing transmission line connectivities[1]. In addition, we assume that the power system is observable regardless an attack is present or not: *i.e.*, the measurement matrix in the DC model always has full rank. This means that the adversary avoids misleading the control center with drastic system changes (*e.g.*, division into two diconnected parts) that may draw too much attention of the control center[2]. We call the lines not common to both $\bar{\mathcal{E}}$ and $\mathcal{E}$ (*i.e.*, lines in $\bar{\mathcal{E}} \triangle \mathcal{E} \triangleq (\bar{\mathcal{E}} \setminus \mathcal{E}) \cup (\mathcal{E} \setminus \bar{\mathcal{E}})$) *target lines* and the buses at the ends of the target lines *target buses*.

To alter the network topology, the adversary launches a man-in-the-middle attack as described in Fig. 1: it intercepts $(\mathbf{s}, \mathbf{z})$ from RTUs, modifies part of them, and forwards the modified version $(\bar{\mathbf{s}}, \bar{\mathbf{z}})$ to the control center.

Throughout the paper, except in Section IV, we assume that the adversary has global information, *i.e.*, it knows network parameters and observes all entries of $(\mathbf{s}, \mathbf{z})$ before launching the attack, although it may modify only the entries it gained control of. Such an unlimited access to network parameters and data is a huge advantage to the attacker. In Section V, countermeasures are designed under this assumption so that they can be robust to such worst case attacks.

The mathematical model of an attack to modify $\mathcal{G}$ to $\bar{\mathcal{G}}$ is as follows (the notation that a bar is on a variable denotes the value modified by the adversary):

$$\begin{aligned} \bar{\mathbf{s}} &= \mathbf{s} + \mathbf{b} \pmod 2, \\ \bar{\mathbf{z}} &= \mathbf{z} + \mathbf{a}(\mathbf{z}), \quad \mathbf{a}(\mathbf{z}) \in \mathcal{A}, \end{aligned} \tag{4}$$

where $\bar{\mathbf{s}}$ is the modified network data corresponding to $\bar{\mathcal{G}}$, $\mathbf{b} \in \{0,1\}^d$ represents the modifications on the network data $\mathbf{s}$, $\mathbf{a}(\mathbf{z}) \in \mathbb{R}^m$ denotes the attack vector added to the meter data $\mathbf{z}$, and $\mathcal{A} \subset \mathbb{R}^m$ denotes the subspace of feasible attack vectors.

We assume that the adversary can modify the network data accordingly for any target topology that deems to be valid to the control center. This is the opposite of the assumption employed by most existing studies on *state* attacks where network data that specify the topology are not under attack.

For the attack on analog meter data, we use the notation $\mathbf{a}(\mathbf{z})$ to emphasize that the adversary can design the attack vector based on the whole meter data $\mathbf{z}$. This assumption will be relaxed in Section IV to study an attack with local information. In addition, $\mathcal{A}$ has a form of $\{\mathbf{c} \in \mathbb{R}^m : \ c_i = 0, \ i \in \mathcal{I}_S\}$ where $\mathcal{I}_S$ is the set of indices of secure meter data entries that

[1]The attacks aiming to split or combine buses are out of scope of this paper. Such attacks require modifying the measurements of breaker states *inside* substations. If the control center employs generalized state estimation [27], such modification invokes substation-level state estimation which leads to a robust bad data test. Hence, such attacks are harder to avoid detection.

[2]In fact, the results to be presented in this paper also hold for the general case where the target topology can be anything (*e.g.*, the system may be divided into several disconnected parts), if the control center employs the same bad data test even when the network is unobservable.

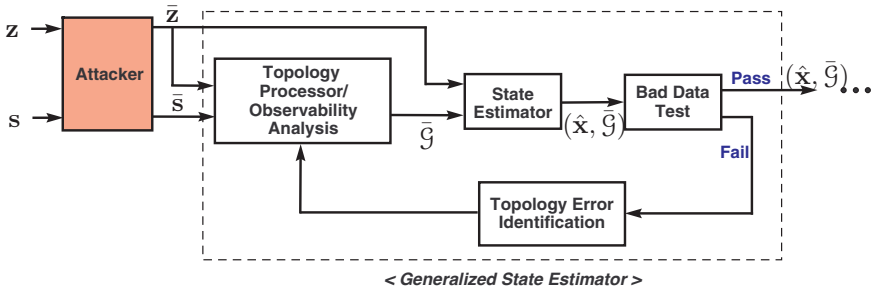


Fig. 1. Attack model with generalized state estimation

the adversary cannot alter and $\{1,\dots,m\}\setminus\mathcal{I}_S$ represents the adversary-controlled entries. Note that $\mathcal{A}$ fully characterizes the power of the adversary, and the mapping $\mathbf{a}:\mathbb{R}^m\to\mathcal{A}$ fully defines the attack strategy.

C. State estimation, bad data test, and undetectable attacks

As illustrated in Fig. 1, the control center executes generalized state estimation (GSE) [27] with network and meter data as inputs; the inputs are $(\mathbf{s},\mathbf{z})$ in the absence of an attack and $(\bar{\mathbf{s}},\bar{\mathbf{z}})$ if there is an attack. GSE regards both network and meter data as possibly erroneous. Once the bad data test detects inconsistency among data and estimates, GSE filters out the outliers from the data and searches for a new *pair* of topology and state estimates that fit the data best. Our focus is on the attacks that can pass the bad data test such that no alarm is raised by GSE.

Under the general AC model (1), if $(\mathbf{s},\mathbf{z})$ is the input to GSE, and $\hat{\mathcal{G}}$ is the topology corresponding to $\mathbf{s}$, the control center obtains the weighted least squares (WLS) estimate of the state $\mathbf{x}$:

$$\hat{\mathbf{x}} \quad = \arg\min_{\mathbf{y}} (\mathbf{z}-h(\mathbf{y},\hat{\mathcal{G}}))^t\Sigma^{-1}(\mathbf{z}-h(\mathbf{y},\hat{\mathcal{G}})).$$

Note that $\hat{\mathcal{G}}=\mathcal{G}$ in the absence of an attack while $\hat{\mathcal{G}}=\bar{\mathcal{G}}$ in the presence of an attack. In practice, nonlinear WLS estimation is implemented numerically [25].

Under the DC model (2), the WLS state estimator is a linear estimator with a closed form expression

$$\begin{aligned}\hat{\mathbf{x}} &= \arg\min_{\mathbf{y}} (\mathbf{z}-\hat{H}\mathbf{y})^t\Sigma^{-1}(\mathbf{z}-\hat{H}\mathbf{y})\\ &= (\hat{H}^t\Sigma^{-1}\hat{H})^{-1}\hat{H}^t\Sigma^{-1}\mathbf{z},\end{aligned}$$

where $\hat{H}$ is the measurement matrix for $\hat{\mathcal{G}}$. The linear estimator is sometimes used as part of an iterative procedure to obtain the nonlinear WLS solution.

The residue error is often used at the control center for bad data detection [25]. In the so-called $J(\hat{\mathbf{x}})$ test [28], the weighted least squares error

$$J(\hat{\mathbf{x}}) = (\mathbf{z}-h(\hat{\mathbf{x}},\hat{\mathcal{G}}))^t\Sigma^{-1}(\mathbf{z}-h(\hat{\mathbf{x}},\hat{\mathcal{G}}))$$

is used in a threshold test:

$$\begin{cases}\text{bad data} & \text{if } J(\hat{\mathbf{x}})>\tau,\\ \text{good data} & \text{if } J(\hat{\mathbf{x}})\le\tau,\end{cases} \tag{5}$$

where τ is the detection threshold, and it is determined to satisfy a certain false alarm constraint α.

We define that an attack is *undetectable* if its detection probability is as low as the false alarm rate of the detector. In this paper, we use the $J(\hat{\mathbf{x}})$ test as the bad data detector.

Definition 2.1: An attack $\mathbf{a}$ to modify $\mathcal{G}$ to $\bar{\mathcal{G}}$ is said to be *undetectable* if, for any true state $\mathbf{x}$, the $J(\hat{\mathbf{x}})$-test with any false alarm constraint detects the attack with the detection probability no greater than its false alarm rate.

In the absence of noise, the only source of bad data is, presumably, an attack. In this case, the probabilistic statement of undetectability becomes a deterministic one. A data attack $(\mathbf{z}+a(\mathbf{z}),\bar{\mathbf{s}})$ that modifies the topology from $\mathcal{G}$ to $\bar{\mathcal{G}}$ is undetectable if for every noiseless measurement $\mathbf{z}$, there exists a state vector $\bar{\mathbf{x}}$ such that $\mathbf{z}+a(\mathbf{z})=h(\bar{\mathbf{x}},\bar{\mathcal{G}})$. Unfortunately, such a nonlinear condition is difficult to check.

Under the DC model, however, the undetectability condition has a simple algebraic form. Let $(\mathbf{s},\mathbf{z})$ be the input to GSE and H is the measurement matrix for the topology corresponding to $\mathbf{s}$. In the presence of an attack, GSE receives $(\bar{\mathbf{s}},\bar{\mathbf{z}})$ instead of $(\mathbf{s},\mathbf{z})$, and $\bar{H}$–the measurement matrix for the target topology $\bar{\mathcal{G}}$–replaces H. In the absence of noise, the $J(\hat{\mathbf{x}})$-detector is equivalent to checking whether the received meter data is in the column space of the valid measurement matrix. Thus, the equivalent undetectable topology attack can be defined by the following easily checkable form:

Definition 2.2: An attack to modify $\mathcal{G}$ to $\bar{\mathcal{G}}$ with the attack vector $\mathbf{a}$ is said to be *undetectable* if

$$\mathbf{z}+\mathbf{a}(\mathbf{z})\in\mathrm{Col}(\bar{H}),\quad \forall\mathbf{z}\in\mathrm{Col}(H), \tag{6}$$

where H and $\bar{H}$ are the measurement matrices for $\mathcal{G}$ and $\bar{\mathcal{G}}$ respectively, and $\mathrm{Col}(H)$ is the column space of H and $\mathrm{Col}(\bar{H})$ the column space of $\bar{H}$.

III. Topology Attack with Global Information

We assume the DC model (2) and present the result for the existence of undetectable topology attacks.

A. Condition for an undetectable attack

We first derive a necessary and sufficient algebraic condition for existence of an undetectable attack that modifies $\mathcal{G}$ to $\bar{\mathcal{G}}$ with the subspace $\mathcal{A}$ of feasible attack vectors. To motivate the general result, consider first the noiseless case.

1) Noiseless measurement case: Suppose there is an undetectable attack $\mathbf{a}$ with $\mathbf{a}(\mathbf{z}) \in \mathcal{A}$, $\forall \mathbf{z} \in \text{Col}(H)$. Then, undetectability implies that $\mathbf{z}+\mathbf{a}(\mathbf{z}) \in \text{Col}(\bar{H})$, $\forall \mathbf{z} \in \text{Col}(H)$, and thus, $\text{Col}(H) \subset \text{Col}(\bar{H}, \mathcal{A})$.[3]

Now suppose $\text{Col}(H) \subset \text{Col}(\bar{H}, \mathcal{A})$. There exists a basis $\{\mathbf{c}_1, \ldots, \mathbf{c}_p, \mathbf{d}_1, \ldots, \mathbf{d}_q\}$ of $\text{Col}(\bar{H}, \mathcal{A})$ such that $\{\mathbf{c}_1, \ldots, \mathbf{c}_p\}$ is a subset of columns of $\bar{H}$ and $\{\mathbf{d}_1, \ldots, \mathbf{d}_q\}$ is a set of linearly independent vectors in $\mathcal{A}$. For any $\mathbf{z} \in \text{Col}(H)$, since $\text{Col}(H) \subset \text{Col}(\bar{H}, \mathcal{A})$, there exist unique $(\alpha_i)_{i=1}^p \in \mathbb{R}^p$ and $(\beta_j)_{j=1}^q$ such that $\mathbf{z} = \sum_{i=1}^p \alpha_i \mathbf{c}_i + \sum_{j=1}^q \beta_j \mathbf{d}_j$. If we set $\mathbf{a}(\mathbf{z}) = -\sum_{j=1}^q \beta_j \mathbf{d}_j$, $\mathbf{z} + \mathbf{a}(\mathbf{z}) = \sum_{i=1}^p \alpha_i \mathbf{c}_i \in \text{Col}(\bar{H})$. In addition, $\mathbf{a}(\mathbf{z}) \in \mathcal{A}$ for all $\mathbf{z}$. Hence, there exists an undetectable attack with the subspace $\mathcal{A}$ of feasible attack vectors.

The above arguments lead to the following theorem.

Theorem 3.1: There exists an undetectable attack to modify $\mathcal{G}$ to $\bar{\mathcal{G}}$ with the subspace $\mathcal{A}$ of feasible attack vectors if and only if $\text{Col}(H) \subset \text{Col}(\bar{H}, \mathcal{A})$.

2) Noisy measurement case: The following theorem states that the algebraic condition in Theorem 3.1 can also be used in the noisy measurement case.

Theorem 3.2: There exists an undetectable attack to modify $\mathcal{G}$ to $\bar{\mathcal{G}}$ with the subspace $\mathcal{A}$ of feasible attack vectors if and only if $\text{Col}(H) \subset \text{Col}(\bar{H}, \mathcal{A})$.

In addition, if an attack $\mathbf{a}$ is such that $\text{Col}(H) \not\subseteq \text{Col}(\bar{H}, \mathcal{A})$, then for almost every[4] $\mathbf{x} \in \mathbb{R}^n$, when $\mathbf{x}$ is the true state, the detection probability for the attack approaches 1 as the noise variances uniformly decrease to 0.

Proof: See Appendix. ■

Note that when the algebraic condition is not met, the attack can be detected with high probability if the noise variances are sufficiently small. With this algebraic condition, we can check whether the adversary can launch an undetectable attack with $\mathcal{A}$ for the target $\bar{\mathcal{G}}$. The condition will be used in Section V to construct a meter protection strategy to disable undetectable attacks for any target topology.

By finding the smallest dimension of $\mathcal{A}$ satisfying the condition, we can also characterize the minimum cost of undetectable attacks for $\bar{\mathcal{G}}$; in the adversary's point of view, a smaller dimension of $\mathcal{A}$ is preferred, because increasing the dimension of $\mathcal{A}$ necessitates compromising more RTUs or communication devices. In the following section, we present an undetectable attack requiring a small number of data modifications and prove its optimality for a class of targets by utilizing the algebraic condition.

B. State-preserving attack

This section presents a simple undetectable attack, referred to as *state-preserving attack*. As the name suggests, the attack intentionally preserves the state in order to have a sparse attack vector. We again motivate our result by considering first the noiseless case.

[3] $\text{Col}(\bar{H}, \mathcal{A})$ denotes the space spanned by the columns of $\bar{H}$ and a basis of $\mathcal{A}$.

[4] This means "for all $\mathbf{x} \in \mathbb{R}^n \setminus \mathcal{S}$, for some $\mathcal{S} \subset \mathbb{R}^n$ with a zero Lebesgue measure".

1) Noiseless measurement case: Given $\mathbf{z} = H\mathbf{x} \in \text{Col}(H)$, the state-preserving attack sets $\mathbf{a}(\mathbf{z})$ equal to $(\bar{H}-H)\mathbf{x}$. Then, $\mathbf{z} + \mathbf{a}(\mathbf{z}) = \bar{H}\mathbf{x} \in \text{Col}(\bar{H})$; the attack is *undetectable*. Note that the state $\mathbf{x}$ remains the same after the attack. Since H has full column rank, $\mathbf{a}(\mathbf{z})$ can be simply calculated as

$$\mathbf{a}(\mathbf{z}) = (\bar{H} - H)\mathbf{x} = (\bar{H} - H)(H^tH)^{-1}H^t\mathbf{z}. \tag{7}$$

For $\mathbf{a}(\mathbf{z})$ above to be a valid attack vector, it is necessary to be a sparse vector constrained by the meters, the data of which can be altered by the adversary.

To see an intuitive reason why $\bar{H}\mathbf{x} - H\mathbf{x}$ is sparse, consider the simple case that a line is removed from the topology while the state is *preserved*. In this case, the line flows through all the lines, except the removed line, stay the same. Because, the line flow from i to j is determined by (i) (x_i, x_j) and (ii) whether i and j are connected, and for most lines, these two factors remain the same. Hence, only few entries are different between $\bar{H}\mathbf{x}$ and $H\mathbf{x}$. Below, we will show that, for all state $\mathbf{x} \in \mathbb{R}^n$, all entries of $(\bar{H} - H)\mathbf{x}$ are zeros except those associated with the target lines.

As noted in [11], H can be decomposed as $H = MBA^t$, where $M \in \mathbb{R}^{m \times l}$ is the measurement-to-line incidence matrix with $l \triangleq |\mathcal{E}_0|$, $B \in \mathbb{R}^{l \times l}$ is a diagonal matrix with the line susceptances in the diagonal entries, and $A^t \in \mathbb{R}^{l \times n}$ is the line-to-bus incidence matrix. Each column of M (each row of A^t) corresponds to a distinct line in $\mathcal{E}_0$. For $1 \le j \le l$, if the jth column of M corresponds to $(a, b) \in \mathcal{E}_0$, let $v_j^+ \triangleq a$ and $v_j^- \triangleq b$. Then, M is defined such that $M_{ij} = \pm 1$ if the ith meter (the meter corresponding to the ith row of M) measures (i) the line flow from $v_j^\pm$ to $v_j^\mp$ or (ii) the injection at bus $v_j^\pm$; otherwise, $M_{ij} = 0$. For A^t, $(A^t)_{ji} = \pm 1$ if $v_j^\pm = i$, and the line corresponding to the jth row of A^t (or equivalently the jth column of M) is *connected* in $\mathcal{G}$; otherwise, $(A^t)_{ji} = 0$. Note that M and B are independent of the topology, but A^t does depend on $\mathcal{G}$. Fig. 2 provides an example to illustrate the structures of M, B, and A^t. Similarly, $\bar{H}$ is decomposed as $\bar{H} = MB\bar{A}^t$.

As illustrated in Fig. 2, the entries of $BA^t\mathbf{x} \in \mathbb{R}^{l \times 1}$ correspond to the line flows of all the lines in $\mathcal{E}_0$ when the state is $\mathbf{x}$ and the topology is $\mathcal{G}$. Similarly, $B\bar{A}^t\mathbf{x}$ is the vector of line flows when the state is $\mathbf{x}$ and the topology is $\bar{\mathcal{G}}$. Since the states are the same, the kth entry of $BA^t\mathbf{x}$ and that of $B\bar{A}^t\mathbf{x}$ are different only if the corresponding line is connected in one of $\mathcal{G}$ and $\bar{\mathcal{G}}$ while disconnected in the other. Therefore, $(B\bar{A}^t - BA^t)\mathbf{x}$ has all zero entries except the entries corresponding to the lines in $\bar{\mathcal{E}} \triangle \mathcal{E}$. Specifically, the entry corresponding to $(i, j) \in \bar{\mathcal{E}} \setminus \mathcal{E}$ assumes $f_{ij}(\mathbf{x}) \triangleq B_{ij}(x_i - x_j)$, and the entry corresponding to $(i, j) \in \mathcal{E} \setminus \bar{\mathcal{E}}$ assumes $-f_{ij}(\mathbf{x})$. Hence, $(\bar{H} - H)\mathbf{x} = M(B\bar{A}^t - BA^t)\mathbf{x}$ is equal to

$$\sum_{(i,j) \in \bar{\mathcal{E}} \setminus \mathcal{E}} f_{ij}(\mathbf{x})\mathbf{m}_{(i,j)} - \sum_{(i,j) \in \mathcal{E} \setminus \bar{\mathcal{E}}} f_{ij}(\mathbf{x})\mathbf{m}_{(i,j)} \tag{8}$$

where $\mathbf{m}_{(i,j)}$ is the column vector of M corresponding to (i, j). Note that $\mathbf{m}_{(i,j)}$ is a sparse vector that has nonzero entries only at the rows corresponding to the line flow meters on the line (i, j) and the injection meters at i and j.

From (8), for any state $\mathbf{x} \in \mathbb{R}^n$, $(\bar{H} - H)\mathbf{x}$ is a linear combination of elements in $\{\mathbf{m}_{(i,j)} : (i, j) \in \bar{\mathcal{E}} \triangle \mathcal{E}\}$. Hence,

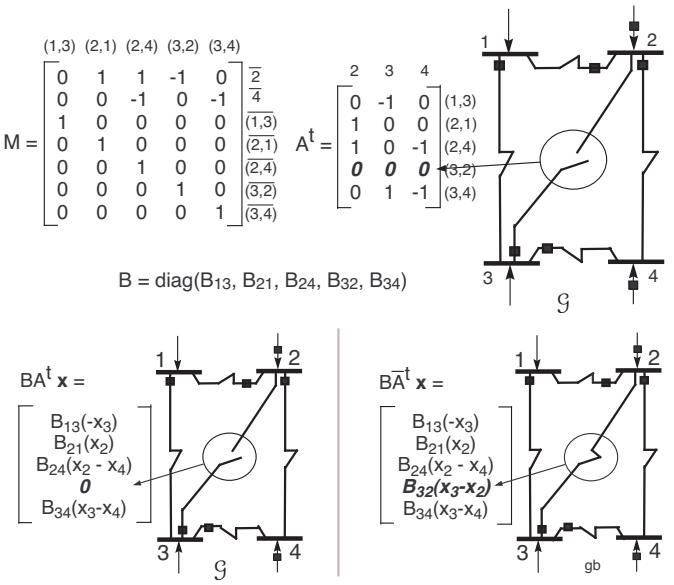


Fig. 2. The measurement, line, or bus corresponding to each row or column is labled. Bus 1 is the slack bus. For the rows of M, $\bar{i}$ denotes the injection meter at bus i, and $\overline{(i,j)}$ the meter for the line flow from i to j.

the state-preserving attack, which sets $\mathbf{a}(\mathbf{z}) = (\bar{H} - H)\mathbf{x}$, modifies at most the line flow meters on the target lines and the injection meters at the target buses.

We now show in the next two theorems that, under certain conditions, the state-preserving attack has the least cost in the sense that it requires the adversary to modify the smallest number of meter data (*i.e.*, the smallest dimension for $\mathcal{A}$).

Theorem 3.3: Assume that (i) the actual and target topologies differ by only one line, *i.e.*, $|\bar{\mathcal{E}} \triangle \mathcal{E}| = 1$, and (ii) every line in $\bar{\mathcal{E}}$, incident[5] from or to any target bus with an injection meter, has at least one line flow meter on it. Then, among all undetectable attacks, the state-preserving attack modifies the smallest number of meters, which is the total number of line flow and injection meters located on the target line and target buses.

Proof: See Appendix ■

Another scenario that the state-preserving attack has the minimum cost is when the adversary aims to delete lines from the actual topology.

Theorem 3.4: Let $\mathcal{G}^*$ and $\bar{\mathcal{G}}^*$ denote the undirected versions of $\mathcal{G}$ and $\bar{\mathcal{G}}$ respectively. Suppose that the adversary aims to remove lines from $\mathcal{G}$, *i.e.*, $\bar{\mathcal{E}} \subsetneq \mathcal{E}$, and the following hold:

- Every line in $\bar{\mathcal{E}}$, incident from or to a target bus with an injection meter, has at least one line flow meter on it.
- In $\mathcal{G}^*$, target lines do not form a closed path.
- $\bar{\mathcal{G}}^*$ does *not* include a tree $\mathcal{T}$ satisfying the following:
  1) (number of nodes in $\mathcal{T}$) ≥ 4, and
  2) every node in $\mathcal{T}$ is a target bus with an injection meter.

Then, among all undetectable attacks, the state-preserving attack modifies the smallest number of meters, which is the total number of line flow and injection meters located on the target lines and target buses.

Proof: See Appendix ■

[5]A line (i,j) is said to be incident from i and incident to j.

Roughly speaking, the assumptions in Theorem 3.4 hold when target lines are far from each other such that there is no big tree in $\bar{\mathcal{G}}$ consisting solely of target buses.

The main advantage of the state-preserving attack is that by preserving the system state during the attack, the attack can be launched by perturbing only *local* meters around the target lines; hence, only few data entries need to be modified. Theorem 3.3 and Theorem 3.4 supports the claim by stating the optimality of the state-preserving attack under the mild assumptions. The theorems also imply that the minimum cost of an undetectable attack can be easily characterized if the target topology satisfies the theorem assumptions.

2) Noisy measurement case: Following the intuition behind the state-preserving attack in the noiseless case, we will construct its counterpart for the noisy measurement case. Recall the relation (8):

$$(\bar{H} - H)\mathbf{x} = \sum_{(i,j)\in\bar{\mathcal{E}}\setminus\mathcal{E}} f_{ij}(\mathbf{x})\mathbf{m}_{(i,j)} - \sum_{(i,j)\in\mathcal{E}\setminus\bar{\mathcal{E}}} f_{ij}(\mathbf{x})\mathbf{m}_{(i,j)}.$$

The above implies that

$$(\bar{H} - H)\mathbf{x} \in \mathcal{M} \triangleq \text{span}\{\mathbf{m}_{(i,j)} : \ (i,j) \in \bar{\mathcal{E}} \triangle \mathcal{E}\} \qquad (9)$$

We set $\mathbf{a}(\mathbf{z})$ as a minimizer of the $J(\hat{\mathbf{x}})$-test statistic[6]:

$$\mathbf{a}(\mathbf{z}) \triangleq \arg\min_{\mathbf{d}\in\mathcal{M}} \|(\mathbf{z}+\mathbf{d}) - \bar{H}\hat{\mathbf{x}}_{\text{WLS}}[\mathbf{z}+\mathbf{d}]\|^2_{\Sigma^{-1}} \qquad (10)$$

where $\hat{\mathbf{x}}_{\text{WLS}}[\mathbf{z}+\mathbf{d}]$ denotes the WLS state estimate when the topology estimate is $\bar{\mathcal{G}}$, and $\mathbf{z}+\mathbf{d}$ is observed at the control center. Note that, since $\mathbf{a}(\mathbf{z}) \in \mathcal{M}$, the attack with $\mathbf{a}$ modifies at most the line flow measurements of the target lines and the injection measurements of the target buses.

Now, suppose that the adversary modifies breaker state measurements such that the topology estimate becomes $\bar{\mathcal{G}}$ and simultaneously modifies the meter data with $\mathbf{a}(\mathbf{z})$. Then, the $J(\hat{\mathbf{x}})$-test statistic at the control center is upper bounded as

$$\begin{aligned}&\|(\mathbf{z}+\mathbf{a}(\mathbf{z})) - \bar{H}\hat{\mathbf{x}}_{\text{WLS}}[\mathbf{z}+\mathbf{a}(\mathbf{z})]\|^2_{\Sigma^{-1}}\\ &\leq \|(\bar{H}\mathbf{x}+\mathbf{e}) - \bar{H}\hat{\mathbf{x}}_{\text{WLS}}[\bar{H}\mathbf{x}+\mathbf{e}]\|^2_{\Sigma^{-1}},\end{aligned}$$

because $(\bar{H} - H)\mathbf{x}$ is an element of $\mathcal{M}$. Note that the right hand side is the $J(\hat{\mathbf{x}})$-test statistic when the meter data are consistent with the topology estimate $\bar{\mathcal{G}}$. Hence, it has χ^2_{m-n} distribution, the same as the distribution of the $J(\hat{\mathbf{x}})$-test statistic under the absence of bad data [28]. This argument leads to the following theorem stating that this attack is undetectable.

Theorem 3.5: The state-preserving attack $\mathbf{a}$, defined in (10), is undetectable.

Note that $\hat{\mathbf{x}}_{\text{WLS}}[\mathbf{z}+\mathbf{d}]$ in (10) is a linear function of $\mathbf{z}+\mathbf{d}$, so $\mathbf{a}(\mathbf{z})$ can be obtained as a linear weighted least squares solution. Specifically, $\mathbf{a}(\mathbf{z})$ has a form of $\mathbf{a}(\mathbf{z}) = D\mathbf{z}$ where $D \in \mathbb{R}^{m\times m}$ depends on $\mathcal{G}$, $\bar{\mathcal{G}}$, and Σ, but not on $\mathbf{z}$. Hence, D can be obtained off-line before observing $\mathbf{z}$.

Note also that the state-preserving attacks in the noiseless and noisy cases modify the same set of meters. In addition, recall that the condition for existence of an undetectable attack is the same for both noiseless and noisy cases. The optimality

[6]We use $\|\mathbf{r}\|^2_{\Sigma^{-1}}$ to denote the quadratic form $\mathbf{r}^t\Sigma^{-1}\mathbf{r}$.

statements for the state-preserving attack in Theorem 3.3 and Theorem 3.4 were derived purely based on the condition for undetectability. Hence, the same optimality statements hold for the noisy measurement case, as stated in the following corollary, and the same interpretation can be made.

Corollary 3.5.1: For the noisy measurement DC model, suppose that the condition in Theorem 3.3 or the condition in Theorem 3.4 hold. Then, among all undetectable attacks, the state-preserving attack modifies the smallest number of meters, which is the total number of line flow and injection meters located on the target lines and target buses.

IV. Undetectable Topology Attacks with Local Information

In this section, we consider the more realistic scenario of a weak attacker who does not have the measurement data of the entire network; it only has access to a few meters. The information available to the adversary is local. We also generalize the linear (DC) measurement model to the nonlinear (AC) model. The resulting undetectable attacks, however, are limited to line removal attacks, *i.e.*, the adversary only tries to remove lines from the actual network topology.

We first consider the noiseless measurement case under the DC model. Since we are restricted to line-removal attacks, $\bar{\mathcal{E}}$ is a strict subset of $\mathcal{E}$. Therefore, recalling (8), we have

$$(\bar{H}-H)\mathbf{x} = -\sum_{(i,j)\in\mathcal{E}\setminus\bar{\mathcal{E}}} f_{ij}(\mathbf{x})\mathbf{m}_{(i,j)} \tag{11}$$

where $f_{ij}(\mathbf{x})$, as defined in Section III, denotes the line flow from i to j when the line is connected, and the state is $\mathbf{x}$.

Let z_{ij} denote the measurement of the line flow from i to j. Due to the absence of noise, $z_{ij} = f_{ij}(\mathbf{x}) = -f_{ji}(\mathbf{x}) = -z_{ji}$. With this observation and (11), we have

$$(\bar{H}-H)\mathbf{x} = -\sum_{(i,j)\in\mathcal{E}\setminus\bar{\mathcal{E}}} z_{ij}\mathbf{m}_{(i,j)} \tag{12}$$

Therefore, setting $\mathbf{a}(\mathbf{z}) = (\bar{H}-H)\mathbf{x}$, which is the state-preserving attack, is *equivalent* to setting

$$\mathbf{a}(\mathbf{z}) = -\sum_{(i,j)\in\mathcal{E}\setminus\bar{\mathcal{E}}} z_{ij}\mathbf{m}_{(i,j)} \tag{13}$$

From (13), one can see that adding the above $\mathbf{a}(\mathbf{z})$ to $\mathbf{z}$ is equivalent to the following heuristic described in Fig. 3:

1) For every target line (i,j), subtract z_{ij} and z_{ji} from the injection measurements at i and j respectively.
2) For every target line (i,j), modify z_{ij} and z_{ji} to 0.

This heuristic simply forces the line flows through the target lines, which are disconnected in $\bar{\mathcal{G}}$, to be zeros, while adjusting the injections at the target buses to satisfy the power balance equations [25]. If a target line (i,j) has only one line flow meter (*e.g.*, z_{ji}), we can use $-z_{ji}$ in the place of z_{ij}. But, if some target line has no line flow meter, this heuristic is not applicable. Note that the heuristic only requires the ability to observe and modify the line flow measurements of the target lines and the injection measurements at the target buses. The adversary can launch it without knowing the topology or network parameters (*i.e.*, H and $\bar{H}$ are not necessary). Since the heuristic is equivalent to the state-preserving attack, it is undetectable.

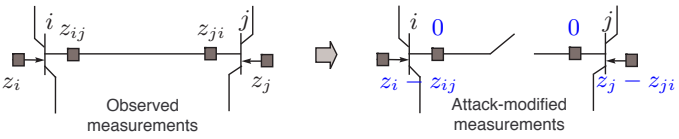


Fig. 3. Heuristic operations around the target line (i,j)

The same heuristic is applicable to the noisy measurements $\mathbf{z} = H\mathbf{x} + \mathbf{e}$. To avoid detection, the adversary can make $\mathbf{a}(\mathbf{z})$ approximate $\bar{H}\mathbf{x} - H\mathbf{x}$ such that $\mathbf{z} + \mathbf{a}(\mathbf{z})$ is close to $\bar{H}\mathbf{x} + \mathbf{e}$. Because $z_{ij} = f_{ij}(\mathbf{x}) + e_{ij}$, z_{ij} is an unbiased estimate of $f_{ij}(\mathbf{x})$. Similarly, $-\sum_{(i,j)\in\mathcal{E}\setminus\bar{\mathcal{E}}} z_{ij}\mathbf{m}_{(i,j)}$ is an unbiased estimate of $-\sum_{(i,j)\in\mathcal{E}\setminus\bar{\mathcal{E}}} f_{ij}(\mathbf{x})\mathbf{m}_{(i,j)}$, which is equal to $\bar{H}\mathbf{x} - H\mathbf{x}$. Hence, it is reasonable to set $\mathbf{a}(\mathbf{z}) = -\sum_{(i,j)\in\mathcal{E}\setminus\bar{\mathcal{E}}} z_{ij}\mathbf{m}_{(i,j)}$ even in the noisy measurement case.

The same idea is applicable to the AC power flow model with the nonlinear state estimator. Suppose that $\mathbf{z}$ is the real power measurement from the AC power flow model: $\mathbf{z} = h(\mathbf{x}) + \mathbf{e}$, where $\mathbf{x}$ is the vector of the voltage phasors at all buses, and h is the nonlinear measurement function for $\mathcal{G}$. Let $\bar{h}$ denote the measurement function for $\bar{\mathcal{G}}$. If $\mathbf{a}(\mathbf{z})$ is equal to $\bar{h}(\mathbf{x}) - h(\mathbf{x})$,

$$\bar{\mathbf{z}} = (h(\mathbf{x}) + \mathbf{e}) + \mathbf{a}(\mathbf{z}) = \bar{h}(\mathbf{x}) + \mathbf{e}, \tag{14}$$

which is consistent with $\bar{\mathcal{G}}$, so the attack cannot be detected. We will show that the attack vector of the heuristic approximates $\bar{h}(\mathbf{x}) - h(\mathbf{x})$.

For simplicity, assume that the attacker aims at removing a single line (i,j) from $\mathcal{G}$. Then, $h(\mathbf{x})$ and $\bar{h}(\mathbf{x})$ are different only in the entries corresponding to the injections at i and j and the line flows through (i,j). Specifically, $\bar{h}(\mathbf{x}) - h(\mathbf{x})$ has all zero entries except $-h_{ij}(\mathbf{x})$ at the rows corresponding to the injection at i and the line flow from i to j, and $-h_{ji}(\mathbf{x})$ at the rows corresponding to the injection at j and the line flow from j to i, where $h_{ij}(\mathbf{x})$ denotes the entry of $h(\mathbf{x})$ corresponding to the line flow from i to j. Since $z_{ij} = h_{ij}(\mathbf{x}) + e_{ij}$ and $z_{ji} = h_{ji}(\mathbf{x}) + e_{ji}$, z_{ij} and z_{ji} can be considered as unbiased estimates of $h_{ij}(\mathbf{x})$ and $h_{ji}(\mathbf{x})$ respectively. Hence, the attacker can use z_{ij} and z_{ji} to construct an unbiased estimate of $\bar{h}(\mathbf{x}) - h(\mathbf{x})$. Adding this estimate to $\mathbf{z}$ is equivalent to the heuristic operation of Fig. 3, which subtracts z_{ij} and z_{ji} from z_i and z_j respectively, and sets z_{ij} and z_{ji} to zeros. The same argument holds for the reactive measurement part and multiple-line removal attacks. In practice, the heuristic attack should be executed twice separately, once for real measurements and second for reactive measurements. In Section VI, numerical simulations demonstrate that the heuristic attack on the AC power flow model with the nonlinear state estimation has a very low detection probability.

V. Countermeasure for Topology Attacks

In this section, we consider countermeasures that prevent attacks by a strong adversary with global information. In particular, we assume that a subset of meters can be secured so that the adversary cannot modify data from these meters.

In practice, this can be accomplished by implementing more sophisticated authentication protocols. We present a so-called cover-up protection that identifies the set of meters that need to be secured.

The algebraic condition in Theorems 3.1-3.2 provides a way to check whether a set of adversary-controlled meters is enough to launch an undetectable attack. Restating the algebraic condition, there exists an undetectable attack with the subspace $\mathcal{A}$ of feasible attack vectors, if and only if $\text{Col}(H) \subset \text{Col}(\bar{H}, \mathcal{A})$ for some $\bar{\mathcal{G}}$ (different from $\mathcal{G}$).

Let $\mathcal{I}_S$ denote the set of indices for the entries of $\mathbf{z}$ corresponding to the protected meters. Then, $\mathcal{A}$ is $\{\mathbf{c} \in \mathbb{R}^m : \ c_i = 0, \ i \in \mathcal{I}_S\}$. The objective of the control center is to make any undetectable attack infeasible while minimizing the cost of protection (*i.e.*, minimizing $|\mathcal{I}_S|$ or equivalently, maximizing the dimension of $\mathcal{A}$).

To achieve the protection goal, $\mathcal{A}$ should satisfy that for any target topology $\bar{\mathcal{G}}$, $\text{Col}(H) \not\subset \text{Col}(\bar{H}, \mathcal{A})$. However, finding such $\mathcal{A}$ by checking the conditions for all possible targets is computationally infeasible. To avoid computational burden, the following theorem gives a simple graph-theoretical strategy.

Theorem 5.1 (Cover-up strategy): Let $\tilde{\mathcal{E}}$ and $\tilde{\mathcal{E}}_0$ denote the undirected counterparts of $\mathcal{E}$ and $\mathcal{E}_0$ respectively. For $i \in \mathcal{V}$, let $\mathcal{L}_i$ denote the set of edges in $(\mathcal{V}, \tilde{\mathcal{E}}_0)$ that are incident to i.

Suppose there is a spanning tree $\mathcal{T} = (\mathcal{V}, \mathcal{E}_\mathcal{T})$ of $(\mathcal{V}, \tilde{\mathcal{E}})$ (the current topology) and a vertex subset $\mathcal{B}$ ($\mathcal{B} \subset \mathcal{V}$) that satisfies

$$\mathcal{E}_\mathcal{T} \cup (\cup_{b \in \mathcal{B}} \mathcal{L}_b) = \tilde{\mathcal{E}}_0. \tag{15}$$

Then, if we protect (i) one line flow meter for each line in $\mathcal{E}_\mathcal{T}$ and (ii) the injection meters at all buses in $\mathcal{B}$, an undetectable attack does not exist for any target topology.

Proof: See Appendix. ■

The condition (15) means that the edges of $\mathcal{T}$ and the edges incident to vertices in $\mathcal{B}$ can cover all the lines (both connected and disconnected) of the grid. One can easily find such $\mathcal{T}$ and $\mathcal{B}$ using available graph algorithms.

Fig. 4 describes a cover-up strategy for IEEE 14-bus system. The strategy used the spanning tree $\mathcal{T}$ marked by red dash lines, and $\mathcal{B} = \{1, 4, 13\}$. The unprotected meters and protected meters are marked by black rectangles and blue circles respectively. In this example, the strategy requires protection of 30% of meters. In addition, numerically checking the algebraic condition showed that if the control center removes *any* of the protections, the grid becomes vulnerable to undetectable topology attacks. This suggests that the strategy does not require protection of an excessive number of meters. For IEEE 118-bus system, a cover-up strategy required protection of 31% of meters.

The cover-up strategy also prevents undetectable state attacks [3]. It follows from Theorem 1 in [9], which states that an undetectable state attack does not exist if and only if the secure meters, protected by the control center, make the system state observable. Because the strategy protects one line meter for each line in the spanning tree $\mathcal{T}$, the system state is always observable with the protected meters [11].

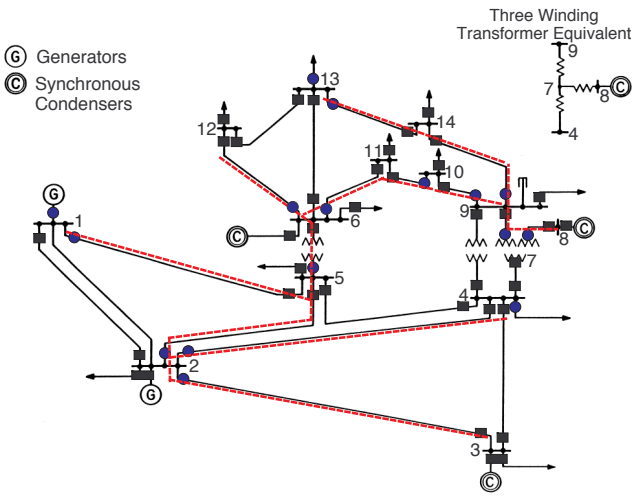


Fig. 4. The cover-up strategy for IEEE 14-bus system: Rectangles (or circles) on buses and lines represent injection meters and line flow meters respectively. We assume that $\mathcal{E} = \mathcal{E}_0$. The attacker may attempt to remove lines from $\mathcal{G}$.

VI. Numerical Results

We first present practical uses of the algebraic condition for undetectable attacks. Then, we test the proposed attacks with IEEE 14-bus and 118-bus systems, and present their effect on real-time LMPs.

A. Application of undetectability condition

In Section III-A, the necessary and sufficient algebraic condition is given to check whether an adversary can launch an undetectable attack for a target $\bar{\mathcal{G}}$ with a subspace $\mathcal{A}$ of feasible attack vectors. Here, we provide examples of how the condition can be used by both attackers and the control center.

Suppose that an attacker with global information aims to remove a specific set of lines from the topology. In Section III-A, we have shown that the state-preserving attack requires the smallest dimension of $\mathcal{A}$ among undetectable attacks under mild conditions. If the conditions are met and the attacker can perform the necessary meter modifications, the state-preserving attack can be launched with the guaranteed optimality. However, if the attacker cannot perform some meter modification required by the state-preserving attack, it should search for an undetectable alternative with a reasonably small dimension for $\mathcal{A}$. The algebraic condition can be used to find such an alternative[7]. For instance, for a line-removal attack on the IEEE 14-bus network in Fig. 4, Table I shows some alternatives to the state-preserving attack when the attacker cannot modify some injection meter.

When the set of adversary-controlled meters is fixed, the algebraic condition can be exploited to find the target topologies, for which the attacker can launch undetectable attacks.

[7]One heuristic way to find an alternative, which we employed, is to begin with a large set $\mathcal{K}$ of adversary-controlled meters that satisfies the algebraic condition and the constraint (*e.g.*, exclude a certain injection meter) and remove meters from $\mathcal{K}$ one by one such that after each removal of a meter, $\mathcal{K}$ still satisfies the algebraic condition. If no more meter can be removed, we take $\mathcal{K}$ as an alternative. The final set depends on the initial $\mathcal{K}$ and the sequence of removed elements. One can try this procedure multiple times with different initial $\mathcal{K}$s and removal sequences, and pick the one with the smallest size.

TABLE I
THE ADVERSARY-CONTROLLED METERS FOR THE ATTACKS TO REMOVE LINES $(2,4)$ AND $(12,13)$: $i \to j$ DENOTES THE METER FOR THE LINE FLOW FROM BUS i TO BUS j. i DENOTES THE INJECTION METER AT BUS i.

	Adversary-controlled meters
State-preserving attack	$2 \to 4$, $4 \to 2$, $12 \to 13$, $13 \to 12$, 2, 4, 12, 13
Alternative 1 (not modifying 12)	$2 \to 4$, $4 \to 2$, $12 \to 13$, $13 \to 12$, $6 \to 12$, $12 \to 6$, 2, 4, 6, 13
Alternative 2 (not modifying 4)	$2 \to 4$, $4 \to 2$, $12 \to 13$, $13 \to 12$, $2 \to 3$, $3 \to 2$, $3 \to 4$, $4 \to 3$, 2, 3, 12, 13

TABLE II
THE SETS OF LINES THAT CAN BE REMOVED BY UNDETECTABLE ATTACKS

$\lvert\hat{\mathcal{E}} \triangle \mathcal{E}\rvert$	$\hat{\mathcal{E}} \triangle \mathcal{E}$ (lines to be removed by the attack)
1	$\{(6,12)\}$, $\{(6,11)\}$, $\{(10,11)\}$, $\{(9,10)\}$, $\{(9,14)\}$, $\{(13,14)\}$, $\{(12,13)\}$
2	$\{(10,11),(13,14)\}$, $\{(9,14),(12,13)\}$, $\{(9,10),(13,14)\}$, $\{(6,12),(13,14)\}$, $\{(6,12),(10,11)\}$, $\{(6,12),(9,10)\}$, $\{(6,11),(12,13)\}$, $\{(6,11),(9,14)\}$
3	$\{(6,11),(9,14),(12,13)\}$, $\{(6,12),(9,10),(13,14)\}$, $\{(6,12),(10,11),(13,14)\}$

For instance, in the IEEE 14-bus network in Fig. 4, assume that the attacker can modify the data from the injection meters at 11, 12, and 14, and all the line flow meters on $(6,12)$, $(6,11)$, $(10,11)$, $(9,10)$, $(9,14)$, and $(13,14)$. Then, numerically checking the algebraic condition show that the attacker cannot launch an undetectable attack for any target. However, if the attacker can additionally control the line flow meters on $(12,13)$, it can launch an undetectable attack to remove any set of lines listed in Table II from the current topology.

The control center can also utilize the algebraic condition to decide which meters to put more security measures on. For instance, in the IEEE 14-bus network, suppose that the control center protects all the injection meter. In the worst case, the attacker may be able to modify all the line flow measurements. In this case, checking the algebraic condition shows that the attacker can launch an undetectable line-removal attack for any target topology, as long as the system with the target topology is observable. However, checking the algebraic condition also shows that if the control center can additionally protect any line flow meter, an undetectable attack does not exist for any target. Therefore, it is worthwhile for the control center to make an effort to secure one more line flow meter.

B. Undetectability and effects on real-time LMP

We tested the state-preserving attack with global information and the heuristic with local information on IEEE 14-bus and IEEE 118-bus system, and investigated their effect on real-time LMPs. The AC power flow model and nonlinear state estimation were used to emulate the real-world power grid.

For simulations, we first assigned the line capacities, generation limits, and estimated loads, and obtained the day-ahead dispatch. Then, we modeled the voltage magnitudes and phases of buses as Gaussian random variables centered at the system state for the day-ahead dispatch, with small variances. In each Monte Carlo run, we generated a state vector from the distribution and used the nonlinear AC power flow model[8] with Gaussian measurement noise to generate the noisy measurements. The attacker observed the noisy measurements, added the corresponding attack vector to them, and passed the corrupt measurements to the control center. The control center employed the nonlinear state estimator to obtain the residue and performed the $J(\hat{\mathbf{x}})$-test with the residue. If $J(\hat{\mathbf{x}})$-test failed to detect the attack, the real-time LMPs were calculated based on the state estimate.

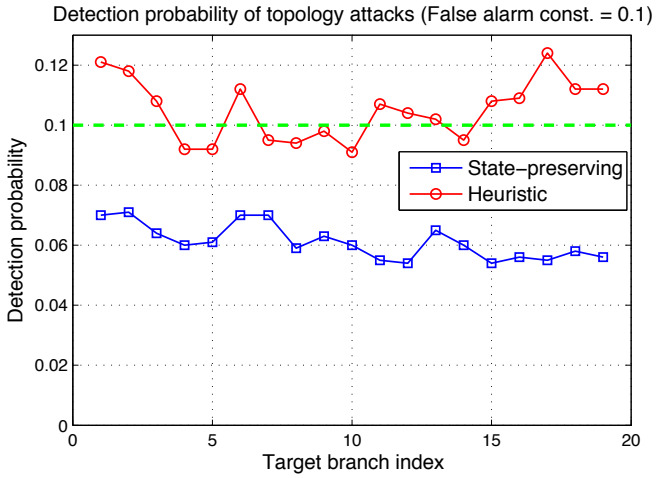


Fig. 5. Detection probability (1000 Monte Carlo runs): the x-axis is for the index of the target line. Measurement noise standard deviation is 0.5 p.u.

TABLE III
DETECTION PROBABILITIES (1000 MONTE CARLO RUNS)

	14-bus		118-bus	
false alarm const. α	$\alpha = 0.1$	$\alpha = 0.01$	$\alpha = 0.1$	$\alpha = 0.01$
state-preserving	0.061	0.009	0.075	0.005
heuristic	0.105	0.019	0.095	0.009

In simulations, we assumed that the attacker aims to remove a single line from the topology. Fig. 5 presents the detection probability of the proposed attacks on IEEE 14-bus system, for different target lines. The attacks on most target lines succeeded with low detection probabilities, close to the false alarm constraint 0.1. Table III shows the detection probability averaged over all possible single-line removal attacks. In both IEEE 14-bus and 118-bus systems, the proposed attacks were hardly detected. In most cases, detection probabilities were as low as the false alarm rates. The performance of the heuristic was remarkably good, considering that it only requires to observe and control few local data.

We also examined the absolute perturbation of the real-time LMPs (see [21] for real-time LMP). The parameters in the real-time LMP calculation include the estimated set of congested lines and the shift-factor matrix; both depend on the topology estimate. Hence, we expect that topology attacks would disturb the real-time LMP calculation. In our simulations, both the state-preserving attack and the heuristic perturbed the real-time LMPs by 10% on average for IEEE

[8]In simulations, we have reactive measurements, which were not considered in our analysis of the state-preserving attack. We simply applied the same analysis for the reactive components of the linearlized decoupled model [25] and derived the reactive counterpart of the state-preserving attack.

14-bus system and 3.3% for IEEE 118-bus system. In the 118-bus system, attacks on some target lines had effects on only the buses near the target lines, so the average perturbation was lower than the 14-bus case.

VII. Conclusion

In this paper, we have considered undetectable malicious data attack aimed at creating a false topology at the control center. We obtain a necessary and sufficient condition for an attack launched by a strong attacker to be undetectable. We also present a class of undetectable line removal attacks that can be launched by weak attackers with only local information. Finally, we present a countermeasure against strong attackers by protecting a subset of meters.

Some of the results presented in this paper are obtained under strong conditions. Here, we mention several of such limitations as pointers for further study. First, the DC model assumed in Section III makes the results valid only near the operating point. It has been demonstrated in [24] that the DC model tends to exaggerate the effect of state attacks, and the nonlinear state estimator has the ability to significantly reduce the attacks' impact on the state estimate. Obtaining conditions for undetectable topology attacks under the AC model is of considerable interest. See [26] for a further discussion.

Second, we have so far focused mostly on state-preserving topology attacks. Even though such attacks are optimal under certain scenarios, to understand the full implication of topology attacks, it is necessary to consider attacks that affect both topology and states.

Finally, we study in this paper only one particular form of countermeasure, namely implementing authentication at a subset of meters. Other mechanisms should be studied, including one with more sophisticated bad data detection and those taking into accounts of system dynamics.

Appendix

A. Proof of Theorem 3.2

The *if* statement can be proved by constructing an undetectable attack following the arguments used to prove Theorem 3.1 and Theorem 3.5. Due to the space limit, we only provide the proof of the *only if* statement.

Let $\mathbf{a}$ be any attack with $\mathrm{Col}(H) \nsubseteq \mathrm{Col}(\bar{H}, \mathcal{U})$ where $\mathcal{U} \triangleq \{\mathbf{u}_1, \ldots, \mathbf{u}_K\}$ denotes the basis of $\mathcal{A}$ consisting of unit vectors in $\mathbb{R}^m$ and $U \in \mathbb{R}^{m \times K}$ is the matrix having the vectors in $\mathcal{U}$ as its columns. Without loss of generality, we assume that the columns of $\bar{H}$ and the unit vectors in $\mathcal{U}$ are linearly independent; if not, we can just work with a smaller set of $\mathcal{U}$ satisfying the independence condition.

Because $\mathrm{Col}(H) \nsubseteq \mathrm{Col}(\bar{H}, \mathcal{U})$, $\mathrm{Col}(H) \cap \mathrm{Col}(\bar{H}, \mathcal{U})$ is a subspace of $\mathrm{Col}(H)$ with a strictly smaller dimension. Hence, $\mathcal{S} \triangleq \{\mathbf{x} \in \mathbb{R}^n : H\mathbf{x} \in \mathrm{Col}(H) \cap \mathrm{Col}(\bar{H}, \mathcal{U})\}$ has the dimension less than n and thus a zero Lebesgue measure in $\mathbb{R}^n$. Let $\mathbf{x}$ be an arbitrary element of $\mathbb{R}^n \setminus \mathcal{S}$. Then, $\mathbf{y} \triangleq H\mathbf{x} \notin \mathrm{Col}(\bar{H}, \mathcal{U})$. When $\mathbf{x}$ is the true state, $\mathbf{z} = \mathbf{y} + \mathbf{e}$, and the $J(\hat{\mathbf{x}})$-test statistic under the attack $\mathbf{a}$ is

$$J = \|W(\mathbf{y} + \mathbf{e} + \mathbf{a}(\mathbf{y} + \mathbf{e}))\|_{\Sigma^{-1}}$$

where $W = I - \bar{H}(\bar{H}^t\Sigma^{-1}\bar{H})^{-1}\bar{H}^t\Sigma^{-1}$. Since $\mathbf{a}(\mathbf{z}) \in \mathrm{Col}(\mathcal{U})$ for all $\mathbf{z}$, J is lower bounded by

$$L \triangleq \min_{(a_k)_{k=1}^K} \|W(\mathbf{y} + \mathbf{e} + \sum_{k=1}^{K} a_k \mathbf{u}_k)\|_{\Sigma^{-1}}.$$

The minimization in L is achieved by the linear WLS solution, and one can show that $L = (\hat{W}(\mathbf{y}+\mathbf{e}))^t\Sigma^{-1}\hat{W}(\mathbf{y}+\mathbf{e})$ where $\hat{W} \triangleq W - (WU)[(WU)^t\Sigma^{-1}(WU)]^{-1}(WU)^t\Sigma^{-1}W$. W and $\hat{W}$ are idempotent and $\Sigma^{-1}W$ is symmetric. Using these properties, one may derive that

$$L = (\Sigma^{-\frac{1}{2}}(\mathbf{y} + \mathbf{e}))^t\Sigma^{\frac{1}{2}}\hat{W}^t\Sigma^{-\frac{1}{2}}(\Sigma^{-\frac{1}{2}}(\mathbf{y} + \mathbf{e})).$$

The above quadratic form has the following properties: (i) $\Sigma^{\frac{1}{2}}\hat{W}^t\Sigma^{-\frac{1}{2}}$ is idempotent and symmetric, (ii) $\Sigma^{-\frac{1}{2}}(\mathbf{y}+\mathbf{e}) \sim \mathcal{N}(\Sigma^{-\frac{1}{2}}\mathbf{y}, I_m)$, and (iii) $\mathrm{rank}(\Sigma^{\frac{1}{2}}\hat{W}^t\Sigma^{-\frac{1}{2}}) = m - n - K$. With these three properties, Theorem B.33 and Theorem 1.3.3 in [29] imply that L has the noncentral chi-squared distribution with the $(m - n - K)$ degree of freedom and the noncentral parameter $\lambda \triangleq (\hat{W}\mathbf{y})^t\Sigma^{-1}(\hat{W}\mathbf{y})$.

It can be shown that $\mathbf{y} \notin \mathrm{Col}(\bar{H}, \mathcal{U})$ implies $\hat{W}\mathbf{y} \neq \mathbf{0}$. Hence, if the diagonal entries of Σ (denoted by σ_{ii}^2, $1 \leq i \leq m$) uniformly decrease to 0, then $\lambda = \sum_{i=1}^m \sigma_{ii}^{-2}(\hat{W}\mathbf{y})_i^2$ grows to infinity. Suppose that the $J(\hat{\mathbf{x}})$-test uses a threshold τ; note that under the DC model, τ depends on the false alarm constraint α, but not on Σ [28]. The detection probability of the attack is $\Pr(J > \tau)$, and it is lower bounded by $\Pr(L > \tau)$. And, $\Pr(L > \tau)$ approaches 1 as the noncentral parameter λ grows to infinity. Therefore, if the diagonal entries of Σ (*i.e.*, noise variances) uniformly decreases to 0, then λ grows to infinity and $\Pr(J > \tau)$ approaches 1. Hence, the only if statement and the additional statement are proved. ■

B. Proof of Theorem 3.3

Let $\bar{\mathcal{E}} \triangle \mathcal{E} = \{(a, b)\}$. We prove the statement for the case that the attack removes (a, b), and there are two line flow meters on (a, b) (one for each direction) and injection meters at both a and b. For the line addition attack and other meter availabilities, the similar argument can be made.

Suppose there exists an undetectable attack with $\mathcal{A}$, and let $\mathcal{U} = \{\mathbf{u}_1, \ldots, \mathbf{u}_K\}$ denote the basis of $\mathcal{A}$ consisting of unit vectors in $\mathbb{R}^m$. Theorem 3.1 implies $\mathrm{Col}(H) \subset \mathrm{Col}(\bar{H}, \mathcal{A})$. It can be easily verified that $\mathbf{m}_{(a,b)} \in \mathrm{Col}(\bar{H}, \mathcal{A})$, and this implies $\mathbf{m}_{(a,b)} = \bar{H}\mathbf{x} + \sum_{k=1}^K \alpha_k \mathbf{u}_k$ for some $\mathbf{x} \in \mathbb{R}^n$ and $(\alpha_k)_{k=1}^K \in \mathbb{R}^K$. Then, $\bar{\mathbf{m}} \triangleq \mathbf{m}_{(a,b)} - \sum_{k=1}^K \alpha_k \mathbf{u}_k \in \mathrm{Col}(\bar{H})$.

Let $\bar{m}^{ij}$ ($\bar{m}^i$) denote the row entry of $\bar{\mathbf{m}}$ corresponding to the line flow from i to j (the injection at i) and $\mathbf{u}_{(i,j)}$ ($\mathbf{u}_{(i)}$) denote the m-dimensional unit vector with 1 at the row corresponding to the line flow from i to j (the injection at i). Physically, $\bar{\mathbf{m}} \in \mathrm{Col}(\bar{H})$ means that $\bar{\mathbf{m}}$ is a vector of meter data consistent with the topology $\bar{\mathcal{G}}$. It implies that (i) $\bar{m}^{ab}$ and $\bar{m}^{ba}$ are zeros, since (a, b) is disconnected in $\bar{\mathcal{G}}$, and (ii) the Kirchhoff's current laws (KCL) should hold at bus a and b in $\bar{\mathcal{G}}$, *i.e.*, the sum of all outgoing line flows from a should be equal to the injection amount at a. Using the special structure of $\mathbf{m}_{(a,b)}$ and $\bar{\mathbf{m}}$, the following can be proved. From (i), one can prove that $\mathbf{u}_{(a,b)}$, $\mathbf{u}_{(b,a)} \in \mathcal{U}$. From (ii), one can show that $\mathcal{U}$ should include $\mathbf{u}_{(a)}$ or some $\mathbf{u}_{(a,k)}$ (or $\mathbf{u}_{(k,a)}$) with a and

k connected in $\mathcal{G}$. Similarly, $\mathcal{U}$ should include $\mathbf{u}_{(b)}$ or some $\mathbf{u}_{(b,l)}$ (or $\mathbf{u}_{(l,b)}$) with b and l connected in $\mathcal{G}$. Hence, $|\mathcal{U}|$ is no less than the total number of meters located on the target line (a, b) and the target buses a and b. ■

C. Proof of Theorem 3.4

Suppose $\mathbf{a}$ is an undetectable attack with $\mathcal{A}$ for the target topology $\bar{\mathcal{G}}$ satisfying the theorem conditions. Let $\mathcal{U} = \{\mathbf{u}_1, \ldots, \mathbf{u}_K\}$ be the basis of $\mathcal{A}$ consisting of unit vectors in $\mathbb{R}^m$, and $\mathcal{J} \subset \mathcal{V}$ denote the set of target buses with injection meters. For ease of presentation, we assume that each target line (i, j) has two line flow meters, one for each direction. For other meter availabilities, the similar argument can be made.

Theorem 3.1 implies that $\text{Col}(H) \subset \text{Col}(\bar{H}, \mathcal{U})$. It can be easily shown that if the target lines do not form a closed path in $\mathcal{G}$, then $\text{Col}(H) \subset \text{Col}(\bar{H}, \mathcal{U})$ implies that $\mathbf{m}_{(i,j)} \in \text{Col}(\bar{H}, \mathcal{U})$ for all target lines $(i, j) \in \mathcal{E} \setminus \bar{\mathcal{E}}$.

$\mathbf{m}_{(i,j)} \in \text{Col}(\bar{H}, \mathcal{U})$ means that it is possible to find a linear combination of vectors in $\mathcal{U}$, $\sum_{k=1}^{K} \alpha_k \mathbf{u}_k$, such that $\bar{\mathbf{m}}_{(i,j)} \triangleq \mathbf{m}_{(i,j)} + \sum_{k=1}^{K} \alpha_k \mathbf{u}_k \in \text{Col}(\bar{H})$. $\bar{\mathbf{m}}_{(i,j)} \in \text{Col}(\bar{H})$ implies that (i) the row entries of $\bar{\mathbf{m}}_{(i,j)}$ corresponding to the line flows of the disconnected lines in $\bar{\mathcal{G}}$ are zeros, and (ii) the entries of $\bar{\mathbf{m}}_{(i,j)}$ satisfy KCLs at all buses in $\bar{\mathcal{G}}$.

For each $(i, j) \in \mathcal{E} \setminus \bar{\mathcal{E}}$, since (i, j) is disconnected in $\bar{\mathcal{G}}$, $\bar{m}^{ij}_{(i,j)} = \bar{m}^{ji}_{(i,j)} = 0$. On the other hand, $m^{ij}_{(i,j)} = 1$ and $m^{ji}_{(i,j)} = -1$. Hence, $\mathcal{U}$ should include $\mathbf{u}_{(i,j)}$ and $\mathbf{u}_{(j,i)}$. Therefore, $\mathcal{U}$ should contain $\{\mathbf{u}_{(i,j)}, \mathbf{u}_{(j,i)} : (i, j) \in \mathcal{E} \setminus \bar{\mathcal{E}}\}$.

For each $i \in \mathcal{J}$, the assumptions imply that each line adjacent to i in $\bar{\mathcal{G}}$ has at least one line flow meter. We let $\mathbf{n}_i$ denote the set of the line flow meters on the lines incident to i in $\bar{\mathcal{G}}$, and $\mathbf{m}^{\mathbf{n}_i}_{(i,j)}$ denote the vector of the corresponding entries in $\mathbf{m}_{(i,j)}$. Because $\mathbf{m}_{(i,j)}$ has nonzero entries only for the injections at i and j and the line flows through (i, j), $\mathbf{m}^{\mathbf{n}_i}_{(i,j)}$ has all zero entries. On the other hand, $m^{i}_{(i,j)} = 1$. Hence, for $\bar{\mathbf{m}}_{(i,j)}$ to satisfy the KCL at bus i in $\bar{\mathcal{G}}$, at least one of $m^{i}_{(i,j)}$ or entries of $\mathbf{m}^{\mathbf{n}_i}_{(i,j)}$ has to be modified by $\sum_{k=1}^{K} \alpha_k \mathbf{u}_k$. Thus, $\mathcal{U}$ should contain $\mathbf{u}_{(i)}$ or $\mathbf{u}_{(a,b)}$ for some $(a, b) \in \mathbf{n}_i$.

In case that $\mathbf{u}_{(i)} \notin \mathcal{U}$, for $\bar{\mathbf{m}}_{(i,j)}$ to satisfy the KCL at bus i in $\bar{\mathcal{G}}$, at least one entry of $\bar{\mathbf{m}}^{\mathbf{n}_i}_{(i,j)}$ should have a nonzero value: suppose $\bar{m}^{ik}_{(i,j)}$ takes a nonzero value. If $k \in \mathcal{J}$, we can make a similar argument based on the KCL at k: $\mathcal{U}$ should contain $\mathbf{u}_{(k)}$ or $\mathbf{u}_{(a,b)}$ for some $(a, b) \in \mathbf{n}(k) \setminus \{(i, k), (k, i)\}$. Following this line of argument, we can derive that for each $i \in \mathcal{J}$, $\mathcal{U}$ should contain unit vectors corresponding to at least one of the following sets: (i) injection meter at i, (ii) line flow meters on all the lines in some path $(i, v_2, \ldots, v_n)$ in $\bar{\mathcal{G}}^*$ and injection meter at v_n where $v_2, \ldots, v_n \in \mathcal{J}$, or (iii) line flow meters on all the lines in some path $(i, v_2, \ldots, v_n)$ in $\bar{\mathcal{G}}^*$ where $v_2, \ldots, v_{n-1} \in \mathcal{J}$ and v_n is either equal to one of $\{v_2, \ldots, v_{n-1}\}$ or not in $\mathcal{J}$. For each $i \in \mathcal{J}$, $\mathcal{U}$ should contain at least one set of unit vectors corresponding to any of the above three cases: we let $\mathcal{S}_i$ to denote an arbitrary one of such sets.

Note that $\{\mathbf{u}_{(i,j)}, \mathbf{u}_{(j,i)} : (i, j) \in \mathcal{E} \setminus \bar{\mathcal{E}}\}$ does not overlap with $\cup_{i \in \mathcal{J}} \mathcal{S}_i$. Hence, $|\mathcal{U}| \geq |\cup_{i \in \mathcal{J}} \mathcal{S}_i| + |\{\mathbf{u}_{(i,j)}, \mathbf{u}_{(j,i)} : (i, j) \in \mathcal{E} \setminus \bar{\mathcal{E}}\}|$. Proving $|\cup_{i \in \mathcal{J}} \mathcal{S}_i| \geq |\mathcal{J}|$ gives us the theorem statement, because $|\mathcal{J}| + |\{\mathbf{u}_{(i,j)}, \mathbf{u}_{(j,i)} : (i, j) \in \mathcal{E} \setminus \bar{\mathcal{E}}\}|$ is the exact number of meters the state-preserving attack modifies.

We will prove the following statement for all $n \leq |\mathcal{J}|$, by mathematical induction: for any subset $\bar{\mathcal{J}} \subset \mathcal{J}$ with $|\bar{\mathcal{J}}| = n$, $|\cup_{i \in \bar{\mathcal{J}}} \mathcal{S}_i| \geq n$. For $n = 1, 2, 3$, the statement can be easily verified. Suppose the statement is true for all $n \leq k$ ($k \geq 3$), and $\bar{\mathcal{J}}$ is an arbitrary subset of $\mathcal{J}$ with $|\bar{\mathcal{J}}| = k + 1$. The tree condition guarantees that $\bar{\mathcal{J}}$ can be partitioned into two nonempty sets $\bar{\mathcal{J}}_1$ and $\bar{\mathcal{J}}_2$ such that for any $b_1 \in \bar{\mathcal{J}}_1$ and $b_2 \in \bar{\mathcal{J}}_2$, every path in $\bar{\mathcal{G}}^*$ between b_1 and b_2 contains a node not in $\mathcal{J}$. This implies that $\cup_{b \in \bar{\mathcal{J}}_1} \mathcal{S}_b$ and $\cup_{b \in \bar{\mathcal{J}}_2} \mathcal{S}_b$ are disjoint. By the induction hypothesis, we have $|\cup_{b \in \bar{\mathcal{J}}_1} \mathcal{S}_b| \geq |\bar{\mathcal{J}}_1|$ and $|\cup_{b \in \bar{\mathcal{J}}_2} \mathcal{S}_b| \geq |\bar{\mathcal{J}}_2|$. Thus, $|\cup_{b \in \bar{\mathcal{J}}} \mathcal{S}_b| = |\cup_{b \in \bar{\mathcal{J}}_1} \mathcal{S}_b| + |\cup_{b \in \bar{\mathcal{J}}_2} \mathcal{S}_b| \geq |\bar{\mathcal{J}}_1| + |\bar{\mathcal{J}}_2| = |\bar{\mathcal{J}}|$. Therefore, the induction implies $|\cup_{i \in \mathcal{J}} \mathcal{S}_i| \geq |\mathcal{J}|$, and the theorem statement follows. ■

D. Proof of Theorem 5.1

Suppose meters are protected as described with $\mathcal{T}$ and $\mathcal{B}$. Let $\mathcal{A}$ be the resulting subspace of feasible attack vectors and $\mathcal{U} \triangleq \{\mathbf{u}_1, \ldots, \mathbf{u}_K\}$ denote the basis of $\mathcal{A}$ consisting of unit vectors in $\mathbb{R}^m$. Assume that an undetectable attack can be launched for some target topology $\bar{\mathcal{G}}$ (different from $\mathcal{G}$). We will show that this assumption leads to a contradiction.

Note that $\mathcal{U}$ cannot contain the unit vectors corresponding to the protected measurements. In addition, Theorem 3.2 implies that $\text{Col}(H) \subset \text{Col}(\bar{H}, \mathcal{U})$. These two imply that the lines in $\mathcal{E}_{\mathcal{T}}$ cannot be removed by the attack, because each line has a protected line flow meter.

Let $\widehat{H}$ ($\widehat{\bar{H}}$) denote the submatrix of H ($\bar{H}$) obtained by selecting the rows corresponding to the protected meter measurements. One can easily verify that $\text{Col}(H) \subset \text{Col}(\bar{H}, \mathcal{U})$ if and only if $\text{Col}(\widehat{H}) \subset \text{Col}(\widehat{\bar{H}})$. Hence, we have $\text{Col}(\widehat{H}) \subset \text{Col}(\widehat{\bar{H}})$. This means that for all $\mathbf{x} \in \mathbb{R}^n$, there exists $\mathbf{y} \in \mathbb{R}^n$ such that $\widehat{\bar{H}}\mathbf{y} = \widehat{H}\mathbf{x}$. Let $H_{\mathcal{T}}$ denote the submatrix of $\widehat{H}$ obtained by selecting the rows corresponding to the protected line flow meters on the spanning tree $\mathcal{T}$. Since the lines in $\mathcal{E}_{\mathcal{T}}$ cannot be removed by the attack, the $H_{\mathcal{T}}$ part of H remains the same in $\bar{H}$; hence, $H_{\mathcal{T}}$ is also a submatrix of $\widehat{\bar{H}}$. Thus, $\widehat{\bar{H}}\mathbf{y} = \widehat{H}\mathbf{x}$ implies $H_{\mathcal{T}}\mathbf{y} = H_{\mathcal{T}}\mathbf{x}$. Since $\mathcal{T}$ is a spanning tree and it has one protected line flow meter per line, the protected line meters on $\mathcal{T}$ makes the grid observable [11]. Hence, $H_{\mathcal{T}}$ has full column rank. Consequently, $H_{\mathcal{T}}\mathbf{y} = H_{\mathcal{T}}\mathbf{x}$ implies $\mathbf{y} = \mathbf{x}$, and we have $\widehat{\bar{H}}\mathbf{x} = \widehat{H}\mathbf{x}$. This holds for all $\mathbf{x} \in \mathbb{R}^n$.

Let a be any element in $\mathcal{B}$. We will show that any line in $\mathcal{L}_a$ cannot be a target line. Note that the injection meter at a is protected, so $\widehat{\bar{H}}$ and $\widehat{H}$ have the row corresponding to the injection at a. $\widehat{\bar{H}}\mathbf{x} = \widehat{H}\mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$ implies that the injection at bus a should be the same for $\mathcal{G}$ and $\bar{\mathcal{G}}$ as long as the state is the same for the two cases. When the state is $\mathbf{x}$, the injection at a in $\mathcal{G}$ is $\sum_{k:\{a,k\} \in \tilde{\mathcal{E}}} B_{ak}(x_a - x_k)$, and the injection at a in $\bar{\mathcal{G}}$ is $\sum_{l:\{a,l\} \in \tilde{\bar{\mathcal{E}}}} \bar{B}_{al}(x_a - x_l)$. Thus we have,

$$\sum_{k:\{a,k\} \in \tilde{\mathcal{E}}} B_{ak}(x_a - x_k) = \sum_{l:\{a,l\} \in \tilde{\bar{\mathcal{E}}}} \bar{B}_{al}(x_a - x_l), \ \forall \mathbf{x} \in \mathbb{R}^n,$$

which can be rewritten as follows: for all $\mathbf{x} \in \mathbb{R}^n$,

$$\sum_{k:\{a,k\} \in \tilde{\mathcal{E}} \setminus \tilde{\bar{\mathcal{E}}}} B_{ak}(x_a - x_k) - \sum_{l:\{a,l\} \in \tilde{\bar{\mathcal{E}}} \setminus \tilde{\mathcal{E}}} \bar{B}_{al}(x_a - x_l) = 0.$$

If $\mathcal{L}_a \cap (\tilde{\tilde{\mathcal{E}}} \triangle \tilde{\mathcal{E}})$ is not empty, the above statement is true only when $B_{ak} = 0$ for all $\{a,k\} \in \mathcal{L}_a \cap (\tilde{\tilde{\mathcal{E}}} \triangle \tilde{\mathcal{E}})$. B_{ak} is the susceptance of the line $\{a,k\}$ when it is "connected", and this value is nonzero in practice for every line. Hence, $\mathcal{L}_a \cap (\tilde{\tilde{\mathcal{E}}} \triangle \tilde{\mathcal{E}})$ should be empty; *i.e.*, a line in $\mathcal{L}_a$ cannot be a target line.

It was shown that the lines in $\mathcal{T}$ and $\cup_{a \in \mathcal{B}} \mathcal{L}_a$ cannot be a target line. Thus, the condition (15) implies that no line can be a target line, and this contradicts the assumption that there exists an undetectable topology attack. ■

REFERENCES

[1] "Vulnerability Analysis of Energy Delivery Control Systems," Idaho National Laboratory, September 2011, INL/EXT-10-18381.

[2] C. Paar and J. Pelzl, *Understanding Cryptography: A Textbook for Students and Practitioners*. Springer, 2010.

[3] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," in *Proc. 16th ACM conference on Computer and communications security*, 2009, pp. 21–32.

[4] 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 *First Workshop on Secure Control Systems,CPSWEEK 2010*, Stockholm, Sweeden, Apr 2010.

[5] H. Sandberg, A. Teixeira, and K. H. Johansson, "On security indices for state estimators in power networks," in *First Workshop on Secure Control Systems,CPSWEEK 2010*, Stockholm, Sweeden, Apr 2010.

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

[7] O. Vukovic, K. C. Sou, G. Dan, and H. Sandberg, "Network-layer protection schemes against stealth attacks on state estimators in power systems," in *Smart Grid Communications (SmartGridComm), 2011 IEEE International Conference on*, oct. 2011, pp. 184 –189.

[8] O. Kosut, L. Jia, R. J. Thomas, and L. Tong, "Malicious data attacks on smart grid state estimation: attack strategies and countermeasures," in *Proc. IEEE 2010 SmartGridComm*, Gaithersburg, MD, USA, Oct 2010.

[9] ——, "Malicious data attacks on the smart grid," *IEEE Trans. Smart Grid*, vol. 2, no. 4, pp. 645 –658, dec. 2011.

[10] A. Monticelli and F. F. Wu, "Network observability: Theory," *IEEE Trans. Power Apparatus and Systems*, vol. PAS-104, no. 5, pp. 1042–1048, May 1985.

[11] G. R. Krumpholz, K. A. Clements, and P. W. Davis, "Power system observability: a practical algorithm using network topology," *IEEE Trans. Power App. Syst.*, vol. 99, no. 4, pp. 1534–1542, July 1980.

[12] K. Clements and P. Davis, "Detection and identification of topology errors in electric power systems," *IEEE Trans. Power Syst.*, vol. 3, no. 4, pp. 1748 –1753, nov 1988.

[13] F. F. Wu and W. E. Liu, "Detection of topology errors by state estimation," *IEEE Trans. Power Syst.*, vol. 4, no. 1, pp. 176–183, Feb 1989.

[14] I. Costa and J. Leao, "Identification of topology errors in power system state estimation," *IEEE Trans. Power Syst.*, vol. 8, no. 4, pp. 1531 –1538, nov 1993.

[15] A. Monticelli, "Modeling circuit breakers in weighted least squares state estimation," *IEEE Trans. Power Syst.*, vol. 8, no. 3, pp. 1143 –1149, aug 1993.

[16] A. Abur, H. Kim, and M. Celik, "Identifying the unknown circuit breaker statuses in power networks," *IEEE Trans. Power Syst.*, vol. 10, no. 4, pp. 2029 –2037, nov. 1995.

[17] L. Mili, G. Steeno, F. Dobraca, and D. French, "A robust estimation method for topology error identification," *IEEE Trans. Power Syst.*, vol. 14, no. 4, pp. 1469 –1476, nov 1999.

[18] E. Lourenco, A. Costa, and K. Clements, "Bayesian-based hypothesis testing for topology error identification in generalized state estimation," *IEEE Trans. Power Syst.*, vol. 19, no. 2, pp. 1206 – 1215, may 2004.

[19] A. Jaen, P. Romero, and A. Exposito, "Substation data validation by a local three-phase generalized state estimator," *IEEE Trans. Power Syst.*, vol. 20, no. 1, pp. 264 – 271, feb. 2005.

[20] F. Vosgerau, A. Simoes Costa, K. Clements, and E. Lourenco, "Power system state and topology coestimation," in *Bulk Power System Dynamics and Control (iREP) - VIII (iREP), 2010 iREP Symposium*, aug. 2010, pp. 1 –6.

[21] A. L. Ott, "Experience with pjm market operation, system design, and implementation," *IEEE Trans. Power Syst.*, vol. 18, no. 2, pp. 528–534, May 2003.

[22] L. Xie, Y. Mo, and B. Sinopoli, "False data injection attacks in electricity markets," in *Proc. IEEE 2010 SmartGridComm*, Gaithersburg, MD, USA., Oct 2010.

[23] L. Jia, R. J. Thomas, and L. Tong, "Malicious data attack on real-time electricity market," in *Proc. 2011 IEEE Intl. Conf. Acoust. Speech & Sig. Proc. (ICASSP)*, Prague, Czech Republic, May 2011.

[24] ——, "On the Nonlinearity Effects on Malicious Data Attack on Power System," in *Power and Energy Society General Meeting, 2012 IEEE*, july 2012.

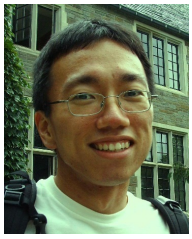
[25] A. Abur and A. G. Expósito, *Power System State Estimation: Theory and Implementation*. CRC, 2000.

[26] J. Kim and L. Tong, "On Data Attacks on a Power Grid: Beyond the DC Model," in *The 47th Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, November 2013, in preparation.

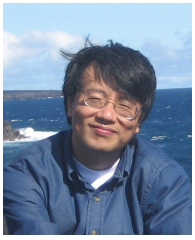
[27] O. Alsac, N. Vempati, B. Stott, and A. Monticelli, "Generalized state estimation," *IEEE Trans. Power Syst.*, vol. 13, no. 3, pp. 1069 –1075, aug 1998.

[28] E. Handschin, F. C. Schweppe, J. Kohlas, and A. Fiechter, "Bad data analysis for power system state estimation," *IEEE Trans. Power App. Syst.*, vol. PAS-94, no. 2, pp. 329–337, Mar/Apr 1975.

[29] R. Christensen, *Plane answers to complex questions: the theory of linear models*. Springer, 2011.



Jinsub Kim received the B.S. degree in electrical engineering from KAIST, Daejeon, Republic of Korea, in 2007, and he is currently pursuing Ph.D. in electrical and computer engineering (with minors in applied mathematics and statistics) at Cornell University, Ithaca, New York. He has conducted research on statistical inference for anomaly detection in communication and power networks. His graduate study was supported by Samsung Scholarship.



Lang Tong (S'87,M'91,SM'01,F'05) is the Irwin and Joan Jacobs Professor in Engineering at Cornell University, Ithaca, New York. He is also the Cornell site director of the Power System Engineering Research Center (PSERC). Lang Tong's current research focuses on inference, optimization, and economic problems in energy and power systems.

He received the B.E. degree in Automation from Tsinghua University, Beijing, China, in 1985, and M.S. and Ph.D. degrees in electrical engineering in 1987 and 1991, respectively, from the University of Notre Dame, Notre Dame, Indiana. He was a Postdoctoral Research Affiliate at the Information Systems Laboratory, Stanford University in 1991. He was the 2001 Cor Wit Visiting Professor at the Delft University of Technology and had held visiting positions at Stanford University and the University of California at Berkeley.

Lang Tong received the 1993 Outstanding Young Author Award from the IEEE Circuits and Systems Society, the 2004 best paper award from IEEE Signal Processing Society, and the 2004 Leonard G. Abraham Prize Paper Award from the IEEE Communications Society. He is also a coauthor of seven student paper awards. He received Young Investigator Award from the Office of Naval Research. He was a Distinguished Lecturer of the IEEE Signal Processing Society.