International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:6, No:6, 2012

# Secure Power Systems Against Malicious Cyber-Physical Data Attacks: Protection and Identification

Morteza Talebi, Jianan Wang, and Zhihua Qu

*Abstract*—The security of power systems against malicious cyber-physical data attacks becomes an important issue. The adversary always attempts to manipulate the information structure of the power system and inject malicious data to deviate state variables while evading the existing detection techniques based on residual test. The solutions proposed in the literature are capable of immunizing the power system against false data injection but they might be too costly and physically not practical in the expansive distribution network. To this end, we define an algebraic condition for trustworthy power system to evade malicious data injection. The proposed protection scheme secures the power system by deterministically reconfiguring the information structure and corresponding residual test. More importantly, it does not require any physical effort in either microgrid or network level. The identification scheme of finding meters being attacked is proposed as well. Eventually, a well-known IEEE 30-bus system is adopted to demonstrate the effectiveness of the proposed schemes.

*Keywords*—Algebraic Criterion, Malicious Cyber-Physical Data Injection, Protection and Identification, Trustworthy Power System.

## I. INTRODUCTION

POWER system is the backbone of a country's economy. The trustworthy issue of the power system is of great crisis towards both human being's and industrial civilization. In order to secure the power system, numerous meters are deployed through power grids, including interconnected generation plants, transmission lines, transformers and loads, to attain updated state information. These information will be provided to the control center or energy management system (i.e. EMS) and analyzed for the prevention from unreliable factors. Most of unreliability accounts for the false date injection, which is usually induced by adversary or hardware failure. The reliability level of system will be tremendously compromised if such injection is not identified and accumulated, especially when it is maliciously initiated by adversary [1][2]. In this respect, research on the power system's protection and identification scheme from malicious data injection is of theoretical and practical interest.

Cyber-physical data attack attempts to deviate the accurate data by introducing erroneous value into certain state variable. Intuitively, such injection is able to be identified by comparing the current state with the outcome of distributed estimation of overall power grid [1][3][4]. The results merely demonstrated that this detection scheme can identify attacks initiated by random phenomena, such as measurement noise, hardware failure or structure error. Recent research [2] indicated a certain type of attack vector, under which the ordinary residual-based scheme is rendered impotent. Apparently, adversary successfully exploits the measurement matrix and manipulates the state variables with malicious data injection composed by a combination of vectors in the null space of $P - I$. In this case, the residual remains unchanged, which fails ordinary bad data detection (BDD). Further, the vulnerability of large-scale power system to malicious data injection can not be omitted due to the significant financial impact of such stealth attack on electricity market [5]. To this end, a greedy algorithm based protection scheme was proposed in [6], which aims at deploying necessary amount secure meters at key buses to ensure a reliable estimation and evade injection. Similar work was introduced in [7] and [9], which illustrates how to secure a state estimator from such injection by encrypting a sufficient/minimum number of meters. The protection strategy of [2] is extended further using a polynomial-time algorithm in [8]. A generalized likehood ratio detection scheme (via convex optimization) is introduced to defense such attack. In addition, several countermeasures to these attacks were also proposed, from additional protected measuring devices [10], to the implementation of improved BDD schemes [2]. Methods to efficiently rank the measurements in terms of their vulnerability and finding sparse attacks requiring the corruption of a low number of measurements were also proposed in [10], [11], and [12]. In [13], a concept of load redistribution (LR) attacks, a special type of false data injection attacks, was introduced and analyzed regarding their damage to power system operation in different time steps with different attacking resource limitations.

From the power system's point of view, the solutions mentioned above are surely functional but they might be too expensive and not be physically practical for expansive distributed network. In this paper, an enhanced protection scheme against malicious false data injection is proposed. An algebraic criterion is derived to ensure a trustworthy power system against malicious cyber-physical data attacks. The proposed protection scheme takes advantage of expansive nature of power grids, reconfigures its subsystem data structure deterministically, and makes it impossible to organize a successful injection. The identification scheme for finding meters being attacked is proposed as well. Then, analysis can be further performed to remove the sources of malicious data injection.

M. Talebi, J. Wang, and Z. Qu are with the Department of Electrical Engineering and Computer Science, University of Central Florida, Orlando, FL, 32816 USA. E-mail: mortezataleby@gmail.com, jianan.wang@ucf.edu, qu@ucf.edu.

The rest of the paper is organized as follows. Section II provides a brief introduction of preliminary results. In section III and IV, we present problem formulation and main results, respectively, to explain how to protect the power system and identify the meters being attacked with our enhanced schemes. Then, an illustrative numerical example of IEEE 30-bus system is elaborated in section V. Section VI concludes the paper and points out the future direction.

## II. Preliminary Results

### A. State Estimation

The state estimation problem in power systems is to determine the power system state variables such as voltage angles and magnitudes at all system buses based on the meter measurements. Given that the general measurement function for the power system is

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

where $x \in \mathbb{R}^n$ is the overall state vector, $z \in \mathbb{R}^m$ is the overall measurement vector and usually $m > n$, $h(x)$ is the nonlinear function derived from the power flow equations of the overall power grid, and $e \in \mathbb{R}^m$ represents the measurement noise whose covariance matrix is $R$. It is assumed that the system (1) is observable that is a very well-established hypothesis for any centralized algorithm of state estimation. The linearized model of the measurement function (1) at time $k$ is

$$z(k) = H(k)x(k) + e(k) \tag{2}$$

with a full-rank observation matrix $H$ as $rank(H) = n$, where $rank(\cdot)$ denotes the rank of matrix. The state estimation problem under the assumption of global observability can be formulated with standard WLS which is given by [14]

$$\hat{x}(k+1) = \hat{x}(k) + K(k)H^T(k)R^{-1}(k)\left[z(k) - H(k)\hat{x}(k)\right], \tag{3}$$

where $\hat{x}$ is the estimate of state and $K(k) = \left[H^T(k)R^{-1}(k)H(k)\right]^{-1}$ is the error covariance.

### B. Bad Data Detection

With the estimated state vector $\hat{x}$ obtained by state estimation algorithm (3), a common approach to verify the integrity of state vector is by computing the $\mathcal{L}-norm$ of measurement residual (i.e. difference between the measurement vector and estimated vector)

$$E \triangleq \|z - H\hat{x}\|. \tag{4}$$

A threshold $C_T$ is pre-defined to control the tolerance of residuals in terms of accuracy of state estimation. If measurement residual is greater than the threshold value, i.e., $E > C_T$, the measurement vector $z$ has a bad data and the state estimation algorithm is not convergent due to either significant measurement/computation errors or gross false data injections. Accordingly, analysis can be performed to position where errors occurs and isolate the suspicious data sources.

### C. Existence of Malicious Data Attacks

In the case that an adversary has access to whole information of $H$, he is able to launch a malicious attack to the system such that the resulting corrupted state can avoid being detected by the residual test in the sense that $E < C_T$ or $E \approx 0$. Following lemma shows how the adversary chooses such a 'stealth' attack which is summarized in [2].

***Lemma 2.1:*** ([2]) Let $z_a \in \mathbb{R}^m$ be amended coordinated attack vector, which will be injected to original measurement vector $z$ in observation equation (2). $z_a = Hc$ where $c$ is the corrupted state induced by the attack vector $z_a$. Let $P = H\left(H^T H\right)^{-1} H^T$, where $P$ is the projection of observation matrix $H \in \mathbb{R}^{m\times n}$ and clearly $PH = H$. All possible choices of coordinated attack vector $z_a \in \mathbb{R}^m$ lie in the null space of matrix $(P - I)$, that is, $(P - I)z_a = \mathbf{0}$.

According to *Lemma 2.1*, the dimension of null space of matrix $(P - I)$ is $n$ regarding the available measurements in power grid. Note that in the power system, it is typical that the number of meters $m$ (both essential and redundant measurements) are greater than number of state variables $n$. The coordinated attack vectors $z_a$ always exists if the adversary can get access to all meters' data, power network topology and line data of subsystem to construct $H$. The attack vectors can be chosen to be a linear combination of the vectors in the null space of $(P-I)$. Secure meters' placement can be considered as one methodology for preventing those coordinated attack vectors and maintaining the subsystem in normal status. However, it could be expensive and physically impractical in expansive distribution network.

## III. Problem Formulation

A power system consists of electric generators, transmission lines, and transformers that form an electrical network. We consider a power system whose electric power grid can be partitioned into a group of $\ell$ subsystems (shown in figure 1). Monitoring the power flow and voltage of each subsystem is important in maintaining system reliability. It is assumed that the subsystem has the capability of reconfiguring its information structure, performing state estimation, and reporting its findings to the upper-level EMS (energy management system).

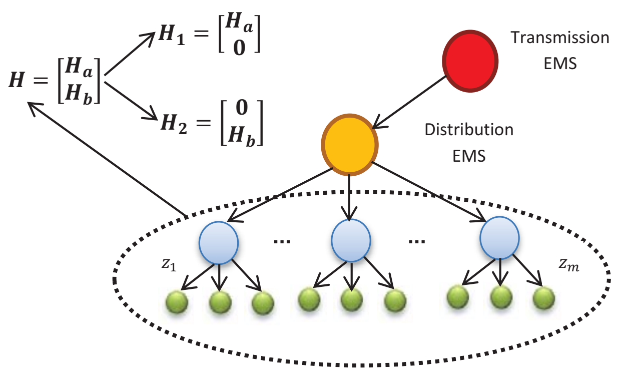


Fig. 1. Rationale of Protection of Power Systems against Malicious Cyber-Physical Data Attack

In the customary case of state estimation, it is common to find the observation matrix $H$ for the subsystem to estimate the state variables. Hence, if the adversary is capable of getting access to information structure, he always can easily fake the eigen-structure of matrix $[P - I]$ and attempt to corrupt the state vector by a stealth false data injection without being detected by the ordinary BDDs according to the *Lemma 2.1*. Obviously, the counter measurement method is secure meters' placement at sufficient number of locations to prevent measurements being manipulated by adversary. Such an approach would work well for certain size of transmission networks but not for expansive distribution networks.

As an alternative solution, an algebraic condition is proposed in the following proposition to secure the power system against malicious data attacks and also depicted in figure 1.

***Proposition 3.1:*** Consider the power system with observation eq. (2), the power system is considered secured from malicious data attacks if the observation matrix $H$ can be reconfigured by $\begin{bmatrix} H_a \\ H_b \end{bmatrix}$ and partitioned by two parts, $H_1 = \begin{bmatrix} H_a \\ 0 \end{bmatrix}$ and $H_2 = \begin{bmatrix} 0 \\ H_b \end{bmatrix}$, such that

$$rank \begin{bmatrix} P_1 - I \\ P_2 - I \end{bmatrix} = m, \quad (5)$$

where $P_1$ and $P_2$ is the projection matrix of $H_1$ and $H_2$, respectively.

***Proof:*** It is straightforward to see that, under condition (5), the only admissible solution of attack vector is $z_a = \mathbf{0}$. In other words, any attack vector rather than $\mathbf{0}$ yields a non-zero residual even if the adversary knows $H$ precisely. ■

It is worthy to note that the proposed method is employing reconfiguration of observation matrix to secure the power system from any attack, and it works as long as the power system has sufficient redundancy to ensure the observability for the sub-areas represented by $H_1$ and $H_2$. The proof for the feasibility of finding sub-matrices $H_1$ and $H_2$ will be provided in the next section.

## IV. Main Results

In this section, we will first present the feasibility of finding sub-matrices $H_1$ and $H_2$ for $H$, and then the protection and identification schemes for power systems against malicious data attacks.

Recall the property of Idempotent Matrix in [15], it is straightforward to see that $P_1$ and $P_2$ are both idempotent. Thus, $I - P_1$ and $I - P_2$ are idempotent as well. Let us define $A = P_1 - I$ and $C = P_2 - I$, then the following facts are obvious:

***Fact 4.1:*** $-A$ and $-C$ are idempotent. Also,

$$A^2 = -A, C^2 = -C, (-A)^\sharp = -A, \quad (6)$$

where $^\sharp$ denotes the generalized inverse of a matrix.

***Fact 4.2:*** $P_1$ is the projection matrix of $H_1$,

$$trace(P_1) = rank(P_1) = rank(H_1) \quad (7)$$

where $trace(\cdot)$ denotes the trace of a matrix.

Based on the above facts, we have the following proposition.

***Proposition 4.3:*** $rank(A) = m - n$ if $rank(H_1) = n$.

***Proof:*** With *Fact 4.2*

$$\begin{aligned} rank(A) &= rank(P_1 - I) = rank(I - P_1) \\ &= trace(I - P_1) = trace(I) - trace(P_1) \\ &= m - rank(H_1) = m - n, \end{aligned} \quad (8)$$

if $H_1$ is observable to the entire system, which means $rank(H_1) = n$. ■

The following lemma will be used for the main result as well.

***Lemma 4.4:*** ([16]) Let $A \in \mathbb{C}^{m\times n}$, $B \in \mathbb{C}^{m\times k}$, $C \in \mathbb{C}^{l\times n}$ and $D \in \mathbb{C}^{l\times k}$. Then,

$$\begin{aligned} rank([A, B]) &= rank(A) + rank(B - AA^\sharp B) \\ &= rank(B) + rank(A - BB^\sharp A) \\ rank(\begin{bmatrix} A \\ C \end{bmatrix}) &= rank(A) + rank(C - CA^\sharp A) \\ &= rank(C) + rank(A - AC^\sharp C) \\ rank(\begin{bmatrix} A & B \\ C & 0 \end{bmatrix}) &= rank(B) + rank(C) \\ &\quad + rank[(I_m - BB^\sharp)A(I_n - C^\sharp C)] \\ rank(\begin{bmatrix} A & B \\ C & D \end{bmatrix}) &= rank(A) \\ &\quad + rank(\begin{bmatrix} 0 & B - AA^\sharp B \\ C - CA^\sharp A & D - CA^\sharp B \end{bmatrix}). \end{aligned} \quad (9)$$

Then, we are ready to present the first main result as follows.

***Theorem 4.5:*** Given $H = \begin{bmatrix} H_a \in \mathbb{R}^{l\times n} \\ H_b \in \mathbb{R}^{m-l\times n} \end{bmatrix}$, $rank \begin{bmatrix} P_1 - I \\ P_2 - I \end{bmatrix} = m$ holds if $H_1 = \begin{bmatrix} H_a \\ 0 \end{bmatrix}$, $H_2 = \begin{bmatrix} 0 \\ H_b \end{bmatrix}$, and $rank(H_1) = rank(H_2) = n$.

***Proof:*** Recall second equation of (9) in *Lemma 4.4*,

$$\begin{aligned} rank(\begin{bmatrix} A \\ C \end{bmatrix}) &= rank(A) + rank(C - CA^\sharp A) \\ &= m - n + rank(C - CA^\sharp A) \end{aligned} \quad (10)$$

due to *Proposition 4.3*, which requires $rank(H_1) = rank(H_a) = n$.

Given $H_1 = \begin{bmatrix} H_a \\ 0 \end{bmatrix}$,

$$P_1 = H_1(H_1^T H_1)^{-1} H_1^T = \begin{bmatrix} H_a(H_a^T H_a)^{-1} H_a^T & 0 \\ 0 & 0 \end{bmatrix}, \quad (11)$$

and $H_2 = \begin{bmatrix} 0 \\ H_b \end{bmatrix}$,

$$P_2 = H_2(H_2^T H_2)^{-1} H_2^T = \begin{bmatrix} 0 & 0 \\ 0 & H_b(H_b^T H_b)^{-1} H_b^T \end{bmatrix}. \quad (12)$$

Then, with *Fact 4.1*,

$$\begin{aligned} C - CA^{\sharp}A &= C(I - A^{\sharp}A) = C(I - A^2) \\ &= C(I + A) = CP_1 \\ &= \begin{bmatrix} -I & 0 \\ 0 & H_b(H_b^T H_b)^{-1}H_b^T - I \end{bmatrix} \\ &\cdot \begin{bmatrix} H_a(H_a^T H_a)^{-1}H_a^T & 0 \\ 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} -H_a(H_a^T H_a)^{-1}H_a^T & 0 \\ 0 & 0 \end{bmatrix} \end{aligned} \tag{13}$$

Thus,

$$rank(C - CA^{\sharp}A) = rank(H_a) = rank(H_1). \tag{14}$$

It finalizes the proof by also noticing that both $H_1$ and $H_2$ are required to be full rank $n$. ■

*Theorem 4.5* provides a mathematical solution to find the sub-matrices $H_1$ and $H_2$ such that eq. (5) holds. Together with *Proposition 3.1*, it also manifests that reconfiguring information structure and corresponding residual test are capable of securing the power system against malicious data attacks.

***Remark 4.6:*** It is worth noting that $rank(\begin{bmatrix} A \\ C \end{bmatrix}) = m - n$ if and only if $H_a = 0_{l\times n}$. It implies that any row elimination of $H$ will contribute the increase of rank. Until eliminating $H_a$ with rank $n$, the full rank will be met. Also note that the full-rank requirement of $H_1$ and $H_2$ leads to $n \le l \le m - n$, which also indicates sufficient measures are required in the sense that $m \ge 2n$.

In what follows, an innovative protection scheme based on *Proposition 3.1* and *Theorem 4.5* is proposed in figure 2 for power system to enhance the security against malicious data attacks. It is a purely mathematical approach and does not require any physical effort either microgrid or network level in comparison with existing work.

Vice verse, the identification scheme is also right on hand based on *Proposition 3.1* and *Theorem 4.5*. The meters that are being attacked by malicious data attack can be identified through the calculation of attack vector $\bar{z}_a$ given the residual vectors $r_1$ and $r_2$ generated by two sub-areas $H_1$ and $H_2$,

$$\bar{z}_a = (\bar{P}^T\bar{P})^{-1}\bar{P}^T \begin{bmatrix} r_1 \\ r_2 \end{bmatrix}, \tag{15}$$

where $\bar{P} = \begin{bmatrix} P_1 - I \\ P_2 - I \end{bmatrix}$. It is true that all the meters corresponding to the non-zero elements in attack vector are being attacked. Further analysis can be performed to remove the sources of malicious data attack. The procedure of identification can be found in figure 3.

The performance of the proposed protection and identification schemes will be illustrated in the next section.

## V. Illustrative Example and Results

In this section, a IEEE modified 30-bus system depicted in figure 4 is adopted to validate the effectiveness of proposed schemes. In terms of the system's setup , bus 1 is the reference bus ($\theta_1 = 0, V_1 = 1$) and the phase angles $\theta_2$ up to $\theta_{30}$ are the state variables due to the simplicity. The voltage magnitude

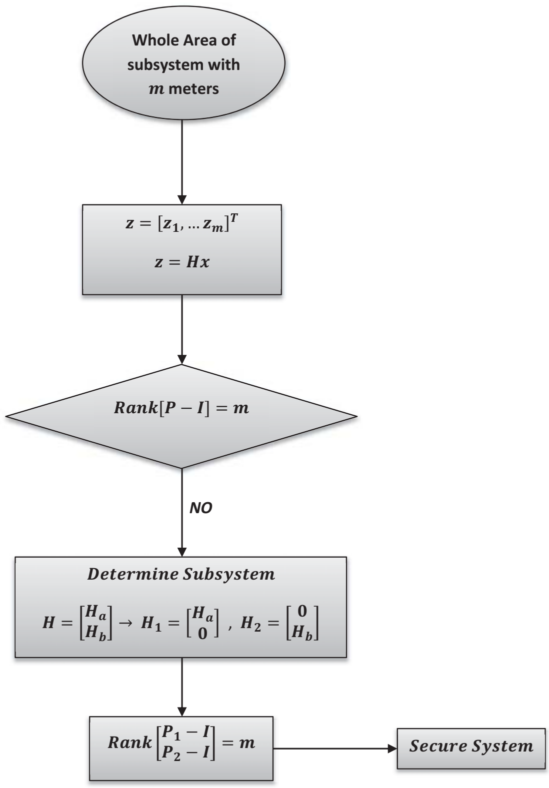


Fig. 2. Protection scheme for Power system against malicious data attack

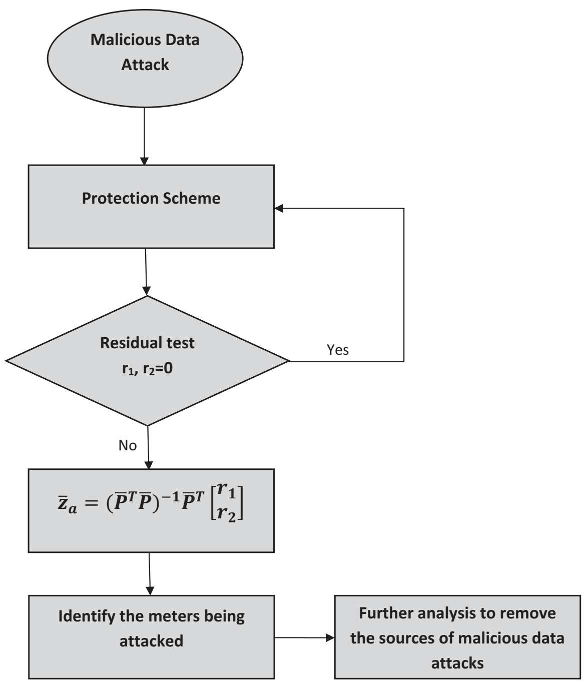


Fig. 3. Identification scheme for Power system against malicious data attack

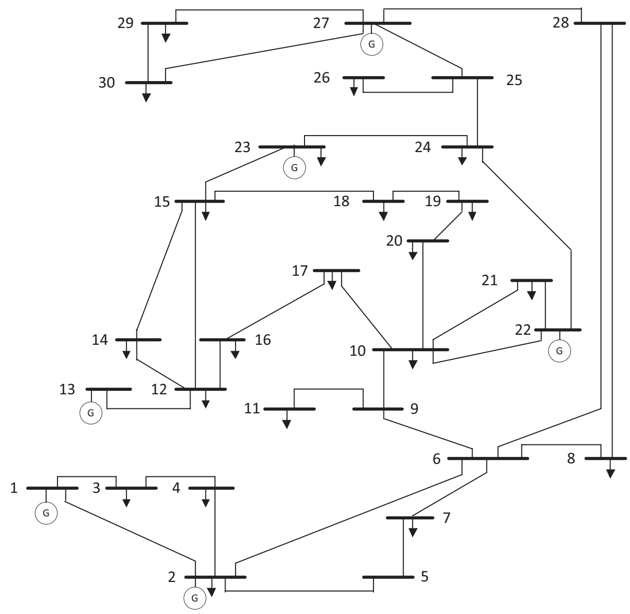


Fig. 4. A single line diagram of modified IEEE 30-bus power system

of each bus is assumed to be known. It is also assumed that the measurement vector $z$ of system is given by a total set of 86 meters which measure 82 active/reactive branch flow and 4 power injection measurements. For more details, line data and operational point of the system are given in appendix A. The observation matrix $H \in \mathbb{R}^{86\times 29}$ are all derived by partial derivative of available measurements with respect to state vector $\theta = \begin{bmatrix}\theta_2 & \cdots & \theta_{30}\end{bmatrix}^T$ as follows. (partial data has been omitted due to the limited space)

$$H = \begin{bmatrix} -15.0358 & 0 & \cdots & 0 & 0 & 0 \\ 0 & -4.8717 & \cdots & 0 & 0 & 0 \\ 5.1686 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 22.6778 & \cdots & 0 & 0 & 0 \\ 4.6507 & 0 & \cdots & 0 & 0 & 0 \\ 4.9159 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \ldots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0.6279 \\ 0 & 0 & \cdots & 0 & -0.8509 & 0.8509 \\ 0 & 0 & \cdots & 1.3204 & 0 & 0 \\ 0 & 0 & \cdots & 4.7335 & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 \end{bmatrix}$$

Note that $m = 86 > 58 = 2n$ guarantees the sufficient redundancy of measurements which is required in *Theorem 4.5*. It can be obtained that $rank(P - I) = 57 < 86$ and hence there are 29 linearly independent choices of coordinated attack vectors. In other words, 29 attack vectors are available to be used for injecting malicious data to corrupt the state estimation. By inspecting the null space of $P - I$, the data attack vectors $z_a$ which correspond to 86 meters are given in table I: (partial data has been omitted due to the limited space)

| $z_a^1$ | $z_a^2$ | $z_a^3$ | $\cdots$ | $z_a^{27}$ | $z_a^{28}$ | $z_a^{29}$ |
|---|---|---|---|---|---|---|
| -0.0013 | 0.0010 | -0.0003 | $\cdots$ | -0.0004 | -0.0003 | -0.0003 |
| 0.0006 | -0.0003 | 0.0002 | $\cdots$ | 0.0000 | 0.0001 | 0.0001 |
| 0.0066 | 0.0231 | -0.0171 | $\cdots$ | -0.0514 | 0.0834 | 0.0488 |
| 0.0398 | 0.0126 | -0.0120 | $\cdots$ | -0.0245 | -0.0911 | 0.0055 |
| 0.0512 | -0.0016 | 0.0137 | $\cdots$ | 0.1520 | -0.0147 | 0.0121 |
| 0.0199 | -0.0303 | 0.0204 | $\cdots$ | 0.0703 | -0.2172 | -0.0807 |
| 0.0524 | 0.0037 | -0.0236 | $\cdots$ | -0.1269 | -0.0262 | -0.0841 |
| $\vdots$ | $\vdots$ | $\vdots$ | $\ldots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| -0.3732 | 0.1800 | -0.0984 | $\cdots$ | -0.0146 | -0.0657 | -0.0755 |
| 0.0027 | -0.0580 | 0.0807 | $\cdots$ | 0.0751 | -0.0631 | -0.0266 |
| -0.0148 | -0.0519 | 0.0385 | $\cdots$ | 0.1156 | -0.1874 | -0.1097 |
| -0.0766 | -0.0242 | 0.0230 | $\cdots$ | 0.0472 | 0.1752 | -0.0105 |
| -0.1201 | 0.0039 | -0.0322 | $\cdots$ | -0.3565 | 0.0345 | -0.0284 |
| -0.0181 | 0.0276 | -0.0185 | $\cdots$ | -0.0639 | 0.1976 | 0.0735 |

TABLE I
CHOICES OF MALICIOUS DATA ATTACK VECTORS

The adversary can choose any linear combination of these 29 non-zero attack vectors to inject malicious data and obviously $(P - I)z_a = 0$ holds. For more clarification, assume that the adversary is injecting $z_a^1$ to real measurement $z$. As we discussed earlier, this type of coordinated attack will not be detected by the residual test since

$$E_1 = \|z + z_a^1 - H\bar{x}\| = 3.1187 \times 10^{-14}$$

which is almost zero and will be surely smaller than the predefined threshold $C_T$.

Next, the proposed schemes will be implemented for the illustration of effectiveness. What is more, the statistical analysis will be adopted to verify the equivalence between standard WLS state estimation and batch state estimation induced by our scheme.

## A. Protection

By noticing the fact that there always exits malicious data attack vectors for the current system, we then follow the protection scheme depicted in Fig. 2 to secure the system. Via row operation, two sub-matrices $H_1$ and $H_2$ can be found by excluding 29 essential meters (independent rows) from the observation matrix $H$ and setting zero for rest of the rows in each of them: (partial data has been omitted due to the limited space)

$$H_1 = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & -0.0062 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ldots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \cdots & -2.4294 & -1.8435 & -1.2754 \\ 0 & 0 & 0 & \cdots & 0 & -1.8435 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1.6560 & -1.6560 \end{bmatrix}$$

with $rank(H_1) = 29$, and another sub-matrix $H_2$ turns out to be: (partial data has been omitted due to the limited space)

$$H_2 = \begin{bmatrix} -15.0458 & -4.8725 & 0 & \cdots & 0 & 0 & 0 \\ 29.6883 & 0 & -5.1688 & \cdots & 0 & 0 & 0 \\ 4.8605 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1.1284 & 0 & \cdots & 0 & 0 & 0 \\ -1.6750 & 0 & 1.6750 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \cdots & -0 & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & -0.8509 & 0.8509 \\ 0 & 0 & 0 & \cdots & 0 & 0.9256 & 0 \end{bmatrix}.$$

with $rank(H_2) = 29$. It can be shown that

$$rank \begin{bmatrix} P_1 - I \\ P_2 - I \end{bmatrix} = 86,$$

which validates the *Theorem 4.5*. Together with *Proposition 3.1*, it reveals that this reconfiguration of power system and corresponding residual test are able to secure the modified IEEE 30-bus system from any malicious data attack.

For more clarification, the following residual test is performed when the same attack vector $z_a^1$ is applied:

$$E_2^1 = \|z + z_a^1 - H_1\bar{x}_1\| = 1.2063,$$

or

$$E_2^2 = \|z + z_a^1 - H_2\bar{x}_2\| = 1.0893,$$

which is obviously easier to be detected with the pre-defined threshold $C_T$ comparing to residual test $E_1$. It can be observed that the malicious attack vectors are no longer 'stealth' within the proposed protection scheme such that the effectiveness of proposed protection scheme is validated.

### B. Identification

In this subsection, the effectiveness of identification scheme will be examined. Given the residual vectors $r_1$ and $r_2$ caused by $z_a^1$ regarding sub-areas $H_1$ and $H_2$,

$$r_1 = \begin{bmatrix} 0.0000 & 0.0000 & \cdots & -0.1201 & -0.0181 \end{bmatrix}^T,$$

and

$$r_2 = \begin{bmatrix} -0.0013 & 0.0006 & \cdots & -0.1211 & -0.0286 \end{bmatrix}^T.$$

Via eq. (15), we can calculate the attack vector $\bar{z}_a^1$ as follows,

$$\bar{z}_a^1 = \begin{bmatrix} -0.0013 & 0.1157 & \cdots & -0.1201 & -0.0181 \end{bmatrix}^T \approx z_a^1$$

Then, we can conclude that all the meters are being attack except meters 2, 25, 40, and 79 since the elements in the attack vector associated with these meters are zero. Furthermore, analysis can be performed to remove the sources of malicious data attack.

### C. Statistical Analysis

For the estimation's purpose, the estimation algorithm under the proposed strategies turns out to be a two-batch estimation algorithm since we partition the whole system by two. Essentially, it is important to see the batch estimation is as good as the standard WLS estimation algorithm from the statistical perspective. Thus, the following covariance analysis from [17] is needed:

$$Cov(\hat{x}, \hat{x}) = \sigma^2 (H^T H)^{-1}$$

where the $\sigma^2$ is a variance of measurement error.

Assume that the $\hat{x}_1$ is the estimation of the state variables using $H_1$ and $\hat{x}_2$ is the estimation of the state variables using $H_2$. It is natural to realize that the estimation of two-batch algorithm $\bar{\hat{x}}$ is the average of these two state estimations. Then, the covariance of two-batch estimation algorithm is calculated as below

$$\begin{aligned} Cov(\bar{\hat{x}}, \bar{\hat{x}}) &= Cov(\frac{\hat{x}_1 + \hat{x}_2}{2}, \frac{\hat{x}_1 + \hat{x}_2}{2}) \\ &= \frac{1}{4}\sigma^2 (H_1^T H_1)^{-1} + \frac{1}{4}\sigma^2 (H_2^T H_2)^{-1}. \end{aligned}$$

For illustrating the equivalence of two algorithms, the well-known Frobenius norm [18] is needed to test the equality of these two covariance matrices

$$d^2 = \frac{1}{n} trace(Cov(\hat{x}, \hat{x}) - Cov(\bar{\hat{x}}, \bar{\hat{x}}))^2$$

where $d$ is the distance between two covariance matrices, $n$ is the number of states. It is clear that if two covariance matrices is exactly the same, i.e., $Cov(\hat{x}, \hat{x}) = Cov(\bar{\hat{x}}, \bar{\hat{x}})$, then $d = 0$. Through the calculation,

$$\begin{gathered} d^2 = \frac{1}{29} trace(Cov(\bar{\hat{x}}, \bar{\hat{x}}) - Cov(\hat{x}, \hat{x}))^2 = 0.2425 \\ \Longrightarrow d = 0.4925 \end{gathered}$$

which indicates the approximate equivalence between two algorithms.

## VI. Conclusion and Future Direction

In this paper, an algebraic criterion to secure the power systems against malicious cyber-physical data attacks is firstly proposed. The feasibility of finding such sub-matrices is proven by reconfiguring the information structure. Then, an enhanced protection and identification schemes for power system against malicious data attacks are proposed as well. It is shown that the proposed scheme makes the power system secure from any malicious cyber-physical data attack with the reconfigured information structure and corresponding residual test, which does not require any physical effort comparing to the solutions in the literature. Furthermore, the identification scheme is capable of identifying the meters being attacked and further analysis can be performed to remove the sources of these attacks. Results applied on the modified IEEE 30-bus systems demonstrate the effectiveness of the proposed schemes.

Future work will mainly focus on extending our work to the security of large-scale power grid systems/microgrids and developing the algebraic criterion from the distributed perspective.

## APPENDIX A

### A. Modified IEEE 30-Bus System

The IEEE 30-Bus system is a well-known and classical example of power system. The configuration of information structure ($H$) of modified IEEE 30-Bus System is derived from the matlab package 'MATPOWER' [19]. The information of the line data, bus data, and steady state operational point of our system is shown in table II and III.

| $Bus$ | $V_{Mag}$ | $V_{Ang}$ | $P_G$ | $Q_G$ | $P_L$ | $Q_L$ |
|---|---|---|---|---|---|---|
| 1 | 1.000 | 0.000 | 30.65 | -2.11 | - | - |
| 2 | 1.000 | -0.532 | 60.97 | 33.04 | 21.70 | 12.70 |
| 3 | 0.983 | -1.705 | - | - | 2.40 | 1.20 |
| 4 | 0.980 | -2.017 | - | - | 7.60 | 1.60 |
| 5 | 0.982 | -2.053 | - | - | - | - |
| 6 | 0.972 | -2.532 | - | - | - | - |
| 7 | 0.967 | -2.887 | - | - | 22.80 | 10.90 |
| 8 | 0.960 | -2.993 | - | - | 30.00 | 30.00 |
| 9 | 0.980 | -3.608 | - | - | - | - |
| 10 | 0.984 | -3.872 | - | - | 5.80 | 2.00 |
| 11 | 0.980 | -4.172 | - | - | 4.5 | 0.00 |
| 12 | 0.985 | -1.892 | - | - | 11.20 | 7.50 |
| 13 | 1.000 | 1.121 | 37.00 | 11.33 | - | - |
| 14 | 0.977 | -2.679 | - | - | 6.20 | 1.60 |
| 15 | 0.980 | -2.697 | - | - | 8.20 | 2.50 |
| 16 | 0.977 | -3.060 | - | - | 3.50 | 1.80 |
| 17 | 0.977 | -3.865 | - | - | 9.00 | 5.80 |
| 18 | 0.968 | -3.903 | - | - | 3.20 | 0.90 |
| 19 | 0.965 | -4.406 | - | - | 9.50 | 3.40 |
| 20 | 0.969 | -4.332 | - | - | 2.20 | 0.70 |
| 21 | 0.993 | -3.980 | - | - | 17.50 | 11.20 |
| 22 | 1.000 | -3.884 | 21.59 | 40.26 | - | - |
| 23 | 1.000 | -1.997 | 19.20 | 7.97 | 3.20 | 1.60 |
| 24 | 0.989 | -3.067 | - | - | 8.70 | 6.70 |
| 25 | 0.990 | -2.061 | - | - | - | - |
| 26 | 0.972 | -2.511 | - | - | 3.50 | 2.30 |
| 27 | 1.000 | -1.160 | 26.91 | 10.57 | - | - |
| 28 | 0.974 | -2.539 | - | - | - | - |
| 29 | 0.980 | -2.461 | - | - | 2.40 | 0.90 |
| 30 | 0.968 | -3.374 | - | - | 10.60 | 1.90 |

TABLE II
OPERATING POINTS AND BUS DATA OF THE MODIFIED IEEE 30-BUS SYSTEM

| $FromBus$ | $ToBus$ | $R$ | $X$ | $B/2$ |
|---|---|---|---|---|
| 1 | 2 | 0.02 | 0.06 | 0.015 |
| 1 | 3 | 0.05 | 0.19 | 0.01 |
| 2 | 4 | 0.06 | 0.17 | 0.01 |
| 3 | 4 | 0.01 | 0.04 | 0 |
| 2 | 5 | 0.05 | 0.2 | 0.01 |
| 2 | 6 | 0.06 | 0.18 | 0.01 |
| 4 | 6 | 0.01 | 0.04 | 0 |
| 5 | 7 | 0.05 | 0.12 | 0.01 |
| 6 | 7 | 0.03 | 0.08 | 0.01 |
| 6 | 8 | 0.01 | 0.04 | 0 |
| 6 | 9 | 0 | 0.21 | 0 |
| 6 | 10 | 0 | 0.56 | 0 |
| 9 | 11 | 0 | 0.21 | 0 |
| 9 | 10 | 0.0 | 0.1100 | 0.0 |
| 4 | 12 | 0.0 | 0.260 | 0.0 |
| 12 | 13 | 0.0 | 0.1400 | 0.0 |
| 12 | 14 | 0.12 | 0.2600 | 0.0 |
| 12 | 15 | 0.07 | 0.13 | 0.0 |
| 12 | 16 | 0.09 | 0.2 | 0.0 |
| 14 | 15 | 0.22 | 0.2 | 0.0 |
| 16 | 17 | 0.08 | 0.19 | 0.0 |
| 15 | 18 | 0.11 | 0.22 | 0.0 |
| 18 | 19 | 0.06 | 0.13 | 0.0 |
| 19 | 20 | 0.03 | 0.07 | 0.0 |
| 10 | 20 | 0.09 | 0.21 | 0.0 |
| 10 | 17 | 0.03 | 0.08 | 0.0 |
| 10 | 21 | 0.03 | 0.07 | 0.0 |
| 10 | 22 | 0.07 | 0.15 | 0.0 |
| 21 | 23 | 0.01 | 0.02 | 0.0 |
| 15 | 23 | 0.1000 | 0.20 | 0.0 |
| 22 | 24 | 0.12 | 0.180 | 0.0 |
| 23 | 24 | 0.1300 | 0.2700 | 0.0 |
| 24 | 25 | 0.19 | 0.3300 | 0.0 |
| 25 | 26 | 0.25 | 0.3800 | 0.0 |
| 25 | 27 | 0.1100 | 0.2100 | 0.0 |
| 28 | 27 | 0.0 | 0.400 | 0.0 |
| 27 | 29 | 0.2200 | 0.4200 | 0.0 |
| 27 | 30 | 0.3200 | 0.60 | 0.0 |
| 29 | 30 | 0.2400 | 0.4500 | 0.0 |
| 8 | 28 | 0.0600 | 0.2000 | 0.0214 |
| 6 | 28 | 0.02 | 0.0600 | 0.065 |

TABLE III
LINE DATA OF THE MODIFIED IEEE 30-BUS SYSTEM

## REFERENCES

[1] E. Handschin, F. Schweppe, J. Kohlas, and A. Fiechter, Bad data analysis for power system state estimation, IEEE Transactions on Power Apparatus and Systems, vol. 94, no. 2, pp. 329-337, 1975.
[2] Y. Liu, P. Ning, and M. K. Reiter, False data injection attacks against state estimation in electric power grids, in Proc. ACM Conf. Comput. Commun. Security, Chicago, IL, Nov. 2009.
[3] K. Clements, G. Krumpholz, and P. Davis, Power system state estimation residual analysis: an algorithm using network topology, IEEE Transactions on Power Apparatus and Systems, no. 4, pp. 1779-1787, 1981.
[4] A. Monticelli, Electric power system state estimation, Proceedings of the IEEE, vol. 88, no. 2, pp. 262282, 2000.
[5] L. Xie, Y. Mo, and B. Sinopoli, False data injection attacks in electricity markets, in 2010 First IEEE International Conference on Smart Grid Communications (SmartGridComm)., pp. 226-231,IEEE 2010.
[6] T. T. Kim and H. V. Poor, Strategic protection against data injection attacks on power grids, IEEE Transactions on Smart Grid, vol. 2, pp.326-333, 2011.
[7] R. Bobba, K. Rogers, Q. Wang, H. Khurana, K. Nahrstedt, and T. Overbye, Detecting false data injection attacks on dc state estimation, in CPSWEEK 2010, the First Workshop on Secure Control Systems, 2010.
[8] O. Kosut, L. Jia, R. Thomas, and L. Tong, Malicious data attacks on the smart grid, IEEE Transactions on Smart Grid, no. 4, pp. 645-658,2011.
[9] G. Dan and H. Sandberg, Stealth attacks and protection schemes for state estimators in power systems, in First IEEE International Conference on Smart Grid Communications (SmartGridComm), pp. 214-219, 2010.
[10] A. Giani, E. Bitar, M. McQueen, P. Khargonekar, K. Poolla, and M. Garcia. Smart grid data integrity attacks: Characterizations and countermeasures. In Proceedings of the IEEE SmartGridComm, October 2011.
[11] H. Sandberg, A. Teixeira, and K. H. Johansson. On security indices for state estimators in power networks. In Preprints of the FirstWorkshop on Secure Control Systems, CPSWEEK 2010, Stockholm, Sweden, April 2010.
[12] K. C. Sou, H. Sandberg, and K. H. Johansson. Electric power net- work security analysis via minimum cut relaxation. In Proceedings of the 50th IEEE Conference on Decision and Control, December 2011.
[13] Y. Yuan, Z. Li, and K. Ren, Modeling load redistribution attacks in power systems, IEEE Transactions on Smart Grid, vol. 2. no. 2, pp. 382-390, Jun. 2011.
[14] A. Gomez-Exposito, A. Abur, A. de la Villa Jaen, and C. Gomez-Quiles, Amultilevel state estimation paradigm for smart grids, Proceedings of the IEEE, vol. 99, pp. 952-976, 2011.
[15] Ben-Israel Adi and Greville Thomas N.E., Generalized Inverses, 2nd Edition, Wiley-Interscience, Chpt. 2, Sec.4, 2003.
[16] G. Marsaglia and G. P. H. Styan. Equalities and inequalities for ranks of matrices. Linear and Multi-linear Algebra, 269-292, 2 1974.
[17] R. W. Farebrother, Linear least squares computations, Marcel Dekker INC, pp. 160, 1988.
[18] Muni S. Srivastava and Hirokazu Yanagihara, Testing the equality of several covariance matrices with fewer observations than the dimension,

Elsevier, Journal of Multivariate Analysis 101,pp.1323, 2010.
[19] R.D. Zimmerman and C.E. Murillo-Sanchez. MATPOWER, A MATLAB Power System Simulation Package. http://www.pserc.cornell.edu/matpower/manual.pdf, September 2007.

**Morteza Talebi** is currently a PhD student at the the Department of Electrical Engineering and Computer Science at University of Central Florida, USA. He received his B.S. in electrical engineering from Gilan university, Iran in 2006. He received his M.S. from the North Carolina Agricultural and Technical State University, USA in power and control in 2010. His research interest is application of the control to power system.

**Jianan Wang** (SM'09-M'12) is currently a Postdoctoral Associate in the Department of Electrical Engineering and Computer Science at University of Central Florida, USA. He received his B. S. and M.S. in Control Science and Engineering from the Beijing Jiaotong University and Beijing Institute of Technology, Beijing, China, in 2004 and 2007, respectively. He received his Ph. D in Aerospace Engineering at Mississippi State University in 2011. His research interests include cooperative control of multiple dynamic systems, UAV formation control, obstacle/collision avoidance, trustworthy networked system, and estimation of sensor networks. Dr. Wang is also a senior member of American Institute of Aeronautics and Astronautics (AIAA).

**Zhihua Qu** (M'90-S'93-F'09) received the Ph.D. degree in electrical engineering from the Georgia Institute of Technology, Atlanta, in June 1990.

Since then, he has been with the University of Central Florida (UCF), Orlando, currently a Professor and Chair of Electrical and Computer Engineering and the SAIC Endowed Professor of UCF. His areas of expertise are nonlinear systems and control, energy and power systems, autonomous vehicles, and robotics. In energy systems, his research covers such subjects as low-speed power generation, dynamic stability of distributed power systems, anti-islanding control and protection, distributed generation and load sharing control, distributed VAR compensation, distributed optimization, and cooperative control. Dr. Qu is the author of three books: Robust Tracking Control of Robot Manipulators (Piscataway, NJ: IEEE Press, 1996), Robust Control of Nonlinear Uncertain Systems(New York: Wiley, 1998), and Cooperative Control of Dynamical Systems with Applications to Autonomous Vehicles (New York: Springer-Verlag, 2009).

Dr. Qu is currently serving on the Board of Governors of the IEEE Control Systems Society and as an Associate Editor for Automatica, IEEE Transactions on Automatic Control, and the International Journal of Robotics and Automation.