processes

MDPI

Article

Random-based Hidden Moving Target Defense against Alert False Data Injection Attackers

Bo Liu[1,*], Hongyu Wu[1], Qihui Yang[1] and Hang Zhang[1]

[1] Mike Wiegers Department of Electrical and Computer Engineering, Kansas State University, Manhattan 66506, KS, USA

* Correspondence: liubo1793@ksu.edu

Abstract: Hidden moving target defense (HMTD) is a proactive defense strategy stealthy to attackers by changing the reactance of transmission lines to thwart false data injection (FDI) attacks. However, alert attackers with strong capabilities pose additional risks to the HMTD and thus, it is a much-needed effort to evaluate the hiddenness of the HMTD. This paper first summarizes two existing alert attacker models, i.e., bad-data-detection-based alert attackers and data-driven alert attackers. Further, this paper proposes a novel model-based alert attacker model that uses the MTD operation models to estimate the dispatched line reactance. The proposed attacker model can construct stealthy FDI attacks against HMTD methods that lack randomness by using the estimated line reactance. We propose a novel random-based HMTD (RHMTD) operation method, which utilizes random weights to introduce randomness and uses the derived hiddenness operation conditions as constraints. RHMTD is theoretically proven to be stealthy to three alert attacker models. In addition, we analyze the detection effectiveness of the RHMTD against three alert attacker models. Simulation results on the IEEE 14-bus systems show that traditional HMTD methods fail to detect attacks by the model-based alert attacker, and RHMTD is stealthy to three alert attackers and effective in detecting attacks by three alert attackers.

Keywords: False data injection attack; hidden moving target defense; alert attacker model; state estimation; D-FACTS device; unsupervised learning

1. Introduction

Modern power systems suffer from significant threats from cyber-physical attacks due to the vulnerabilities of widely used information and communication technology (ICT) enabled devices and Internet of things (IoT) technologies. In addition, energy sources such as wind and solar energy have inherent instability that might compromise the stability of the system [1]. According to the U.S. Department of Energy, 362 power interruptions related to cyber-physical attacks were reported between 2011 and 2014 [2]. False data injection (FDI) attacks are one of the most destructive cyber-physical attacks against smart grids. FDI attacks compromise measurements in the supervisory control and data acquisition (SCADA) system, which aim to manipulate the voltage estimated by the state estimation in the energy management system of the power system. The FDI attacks can cause severe consequences, including line overloading, load shedding, unstable system states and even voltage collapse [3].

Moving target defense (MTD) is introduced into the physical layer of power systems to detect FDI attacks. MTD actively perturbs the branch impedance using distributed flexible AC transmission system (D-FACTS) devices, such that the time-variant system configuration invalidates attackers' knowledge about the actual power system configurations. The first MTD work against FDI attacks [4] proposed a random MTD (RMTD)

Citation: To be added by editorial staff during production.

Academic Editor: Firstname Lastname

Received: date
Revised: date
Accepted: date
Published: date

method in which the reactance of an arbitrary subset of D-FACTS-equipped lines is randomly changed. It was proved that MTD methods could effectively detect FDI attacks [5]–[9], cyber-physical attacks [10], and Stuxnet attacks [11].

1.1. Related Work

MTD planning and MTD operation are two essential steps in implementing the MTD method. MTD planning refers to installing D-FACTS devices on an identified subset of transmission lines, and MTD operation refers to adjusting the D-FACTS setpoints under different load conditions. Reference [12] proved that MTD planning determines the detection effectiveness of MTDs. Max-rank placement [5], [12] can achieve the maximum rank of the composite matrix, which is the widely-used metric of MTD detection effectiveness. Arbitrary placement and full placement are the two simplest D-FACTS placement strategies without considering the detection effectiveness of MTDs. Arbitrary placement installed D-FACTS devices on randomly selected lines [4], while full placement installed D-FACTS devices on every transmission line [13]. The placement of D-FACTS devices was optimized in [14], which aims to reduce the number of measurements that can be manipulated by the attacker. It also proved that the coordinated design of consecutive perturbation schemes within an MTD cycle could improve the MTD's performance in detecting FDI attacks.

MTD operation methods mainly determine the function of D-FACTS devices in MTD. The arbitrary operation method, the simplest MTD operation method, randomly perturbed the D-FACTS setpoints [4] without considering the economic benefits and detection effectiveness. Optimal-power-flow (OPF)-based operation methods utilized the D-FACTS devices and OPF model to control the power flow and minimize the system losses or generation costs [12], [15], [16]. Optimization-based operation methods maximized the metric of detection effectiveness or the economic cost to optimally dispatch the D-FACTS setpoints [13], [17]. Recently, a double-benefit moving target defense was proposed to protect the SG from cyber–physical attacks (CPAs) and also gain generation-cost benefits in DC power system model [9]. Reference [18] studied the effectiveness and hiddenness of MTD using measurement residuals in three-phase AC distribution system state estimation and further formulated the optimization problem for MTD to jointly optimize the effectiveness and hiddenness considering voltage stability. Reference [19] developed two strategies to make the increasing operation cost zero for activating the MTD. In addition, it studied the impact of MTD on the system dynamics using small signal stability.

A strong and alert adversary can detect the existence of MTD in place, which can drive the attacker to postpone the attack using the incorrect line impedance. Consequently, the attacker can invest more resources to obtain the current power system configuration, and potentially launch stealthy attacks with a higher-level threat. The concept of hidden MTD is proposed in transmission systems [6] and distribution systems [20], in which the defender delicately modifies the line impedance to maintain MTD hidden to the attacker.

There are three types of hidden MTD methods in the literature. In the first type, referred to as watermarking HMTD [21], the defender slightly changes the line impedance such that the status of the power system will not significantly change, and the attacker will not realize the existence of MTD. However, small line impedance changes cause the Chi-square bad data detector (BDD) in state estimation fails to detect the FDI attacks. The defender had to utilize the CUSUM detector to detect FDI attacks. Due to the characteristic of CUSUM [22], the CUSUM detector cannot immediately detect the FDI attacks, resulting in the system that will suffer from FDI attacks for multiple time instants. In the second type, referred to as secure-meter-based HMTD [23], multiple protected meters were utilized in each loop of the power system topology to cover the status change of the power system and the power flow changes caused by the line impedance changes. It is assumed that attackers have no read access to the protected meters such that alert attackers (AA) cannot detect the existence of MTD using the remaining measurements through the state

estimation. However, this method is expensive for the defender, as the expensive protected meters are only used for ensuring the hiddenness of MTD, rather than improving the detection effectiveness against attacks. In the third type, referred to as model-based HMTD [5], [20], the defender delicately changes the line impedance such that the power flow of each transmission line is the same before and after the MTD. However, the model-based HMTD methods utilize optimization models without any uncertainties, which are not consistent with the dynamic defense nature of MTD. Randomness and diversity are two essential components in the dynamic defense strategy [24]. Without randomness in MTD, the attacker can apply the same HMTD method to estimate the exact line impedance dispatched by the system operator, if the attacker knows which model-based HMTD is used. Therefore, it is necessary to model possible alert attackers, and further improve the hiddenness and detection effectiveness of MTD methods against the different types of smart and alert attackers.

1.2. Research Gap

The research gap is that existing alert attackers need to be summarized and modeled, and novel alert adversaries with strong and advanced capabilities are necessary to be modeled. With clearly-defined alert attacker models, these alert attacker models can be used as a metric to comprehensively evaluate the hiddenness and detection effectiveness of any novel MTD methods. In this paper, two existing alert attackers against MTD are modeled, i.e., BDD-based alert attacker (BDD-AA), and data-driven alert attacker (DD-AA). In addition, this paper proposes a novel model-based alert attacker (M-AA). These three alert attacker models can be used to analyze the drawbacks of existing HMTD methods.

This paper further proposes a novel HMTD method that is hidden to three alert attacker models. We compare the proposed HMTD method with the existing methods regarding the hiddenness and detection effectiveness against three alert attacker models in Table 1. Table 1 presents the drawbacks of existing HMTD methods, highlights the necessity of the proposed model-based alert attacker, and demonstrates the novelties of the proposed HMTD method. Note that the first Yes (Y) or No (N) indicates whether the HMTD method is hidden to a given alert attacker, and the second Y or N indicates whether the HMTD method is able to detect the attacks by the attacker.

Table 1. Comparison of the proposed and existing HMTD methods regarding the hiddenness and detection effectiveness.

Method	BDD-AA	DD-AA	M-AA	Characteristics
Watermarking HMTD [21]	Y/Y	Y/Y	Y/Y	Detection delay of FDI attack
Secure-meter-based HMTD [23]	Y/Y	N/Y	Y/Y	Extra expensive protected meters
Model-based HMTD [5], [20]	Y/Y	Y/Y	Y/N	Lack of randomness
This paper	Y/Y	Y/Y	Y/Y	No detection delay and no protected meters with randomness

1.3. Contribution

To fill the research gap, this paper summarizes two alert attacker models and further proposes a novel alert attacker model. These three alert attacker models formulate a metric to fully evaluate the hiddenness and detection effectiveness of any HMTD method. Then, this paper proposes a novel random-based HMTD (RHMTD) operation model which is stealthy to the three alert attackers. The contribution of this paper is summarized as follows:

- We summarize two alert attacker models against MTD in the literature: *i)* a *BDD-based alert attacker* who uses Chi-square BDD to detect the existence of MTD; and

ii) a data-driven alert attacker who uses dimension reduction and unsupervised learning methods to detect the existence of MTD.

- We propose a novel alert attacker model, i.e., a *model-based alert attacker*, who uses the MTD operation model to calculate the dispatched line reactance and then uses Chi-square BDD to verify the correctness of the estimated reactance. This attacker model can construct stealthy FDI attacks against HMTD methods that lack randomness by using the estimated line parameters.
- We propose a novel random-based HMTD (RHMTD) operation model in the DC power system model, which maximizes the weighted line reactance changes and integrates the derived MTD hiddenness operation condition as constraints. The weights of the line reactance in the objective function follow the uniform distribution for introducing the randomness.
- We theoretically prove that the hiddenness of the proposed RHMTD method against three alert attacker models. We further analyze the attack detection effectiveness of the proposed method against three alert attacker models.

The rest of this paper is organized as follows. In Section II, we define three alert attacker models. In Section III, we derive a novel RHMTD operation model, prove the hiddenness of RHMTD to three alert attackers, and evaluate the attack detection effectiveness of RHMTD against three alert attackers. The case studies in the IEEE 14-bus system are conducted in Section IV. The paper is concluded in Section V.

2. Alert Attacker Models

In this section, we first define variables used in this paper and then define three alert attacker models.

2.1. Notation

Variables used throughout the paper are summarized in Table 2. "D-FACTS lines" and "non-D-FACTS lines" stand for the set of lines equipped with and without D-FACTS devices, respectively.

Table 2. Nomenclature

Symbol	Definition
$\boldsymbol{\theta}$	Voltage angle of buses excluding reference bus
$\mathbf{z}$	Measurement vector
$\mathbf{a}$	FDI attack vector
$\mathbf{H_0}$	DC measurement matrix in SE before MTD
$\mathbf{H}$	DC measurement matrix in SE after MTD
$\mathbf{A}$	Incident matrix of power system graph
$\mathbf{X}$	Diagonal line reactance matrix
x_{ij}	The reactance of line i–j (between bus i and j)
n	Total number of system buses
m	Total number of measurements
p	Total number of lines

2.2. BDD-based Alert Attacker Model

The first BDD-based alert attacker model is proposed in [6]. Here, we refine the BDD-based alert attacker with the capability of topology learning capability.

Attack goal. The BDD-based alert attacker aims to launch traditional stealthy FDI attacks using correct line impedance under the MTD. **Assumption.** We assume that the attacker knows the original configuration of the system without MTD, including the system topology and the line impedance b_0, but doesn't know the actual line impedance dispatched by MTD on the current time instant. **Attacker's capability.** The attacker has read access to all SCADA measurements in the power system to detect MTD, and write access to all measurements to inject FDI attacks. The attacker can perform SE and BDD to detect MTD, and can launch the topology learning (TL) methods [25] to learn the current line impedance $\hat{b}_a$. **Attack logic.** The flowchart of the BDD-based alert

attacker is shown in Fig. 1. The attacker conducts SE using the original line impedance before the MTD, and then performs BDD to calculate the estimation residual by (1).

$$r_a = \left\| \mathbf{z}_1 - \mathbf{H}_0 (\mathbf{H}_0^T \mathbf{H}_0)^{-1} \mathbf{H}_0^T \mathbf{z}_1 \right\|_2 = \left\| \mathbf{H}_1 \mathbf{x}_1 - \mathbf{H}_0 (\mathbf{H}_0^T \mathbf{H}_0)^{-1} \mathbf{H}_0^T \mathbf{z}_1 \right\|_2 \quad (1)$$

where the measurement matrix under the MTD is $\mathbf{H}_1$ and the attacker's original measurement matrix before MTD is $\mathbf{H}_0$.

If the attacker's estimation residual is less than the threshold, i.e., $r_a < r_{th}$, it indicates the attacker's knowledge of the line impedance is correct and no MTD is applied in the field. Then, the attacker can launch stealthy FDI attacks using the original system configuration $\mathbf{H}_0$. If $r_a > r_{th}$, the alert attacker suspects the accuracy of the line impedance due to the MTD and postpones launching attacks until bypassing the BDD check by estimating the actual line impedance with the topology learning methods.

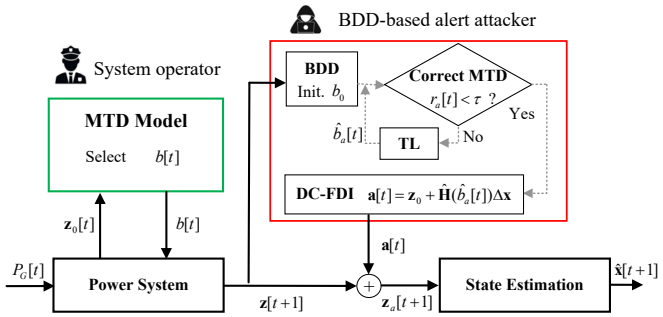


Fig. 1. The attack logic flowchart of the BDD-based alert attacker.

2.3. Data-driven Alert Attacker Model

The first data-driven alert attacker model against MTD is proposed in [21]. Here, we generalize the data-driven alert attacker model, and enable the attacker with stronger attack capability. Currently, only watermarking HMTD has been evaluated to remain hidden to the data-driven attacker through simulation. However, the hiddenness of the secure-meter-based HMTD and model-based HMTD has not been evaluated against the data-driven attacker.

Attack goal. The attacker aims to launch data-driven FDI attacks under the MTD. **Assumption.** It is assumed the attacker doesn't know the configuration of the system before and after the MTD, including the system topology and the line impedance, but he knows MTD may be applied in the system. **Attacker's capability.** The attacker has read access and write access to all SCADA measurements. The attacker can collect historical measurement data over time and the attacker can use unsupervised machine learning methods to analyze the data.

Attack logic. The attack logic of the data-driven alert attacker is shown in Fig. 2. First, the attacker adds all eavesdropped measurements in $\mathbf{Z}$ matrix. Then, the attacker applies the dimension reduction method (e.g., PCA) on the collected historical measurement $\mathbf{Z}$ to 2D for visualization. If the low-dimensional historical measurements form more than one cluster, it reflects the pattern of the power flow measurements significantly changes, indicating MTD could exist in the field. Then, the attacker can apply clustering algorithms (e.g., K-means and DBSCAN) to identify all historical measurements responding to the current MTD, and construct data-driven FDI attacks using the identified historical measurements. However, the number of the identified historical measurements depends on the frequency of MTD. Therefore, under MTD, the number of historical measurements that can be used for constructing data-driven FDI attacks is significantly reduced. Since the performance of data-driven FDI attacks heavily relies on the number of measurements, advanced data-driven FDI attack methods need to be applied, such as matrix-reconstruction FDI [26]. If the low-dimensional historical measurements only form one cluster, it indicates MTD is not applied in the field. Therefore, all collected historical measurements can be used to construct data-driven FDI attacks. With sufficient historical measurements, the attack has more data-driven FDI attack methods to choose from for constructing malicious injection vectors, such as PCA-FDI [27] and subspace FDI [28].

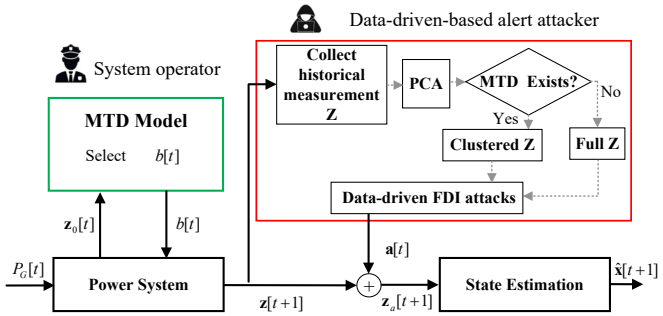


Fig. 2. The attack logic flowchart of the data-driven alert attacker.

2.4. Model-based Alert Attacker Model

For the first time, this paper proposes a model-based alert attacker model. This alert attacker model is designed for the existing MTD or HMTD methods [5], [12], [15], [19], [20], which are based on the optimization problem without considering any uncertainties. If the attacker applies the same MTD model, it is easy to obtain the actual line impedance dispatched in the field.

Attack goal. The attacker aims to launch traditional FDI attacks using the correct line impedance under the MTD. **Assumptions.** We assume the attacker knows the configuration of the system, including the system topology and the line impedance before the MTD. **Attacker's capability.** The attacker has read access and write access to all SCADA measurements. In addition, the attacker is assumed to know the multiple MTD operation models, including the method used by the system operator. **Attack logic.** As shown in Fig. 3, the model-based alert attacker utilizes the MTD operation model to calculate the dispatched line impedance $\hat{b}_a$ and measurement matrix $\hat{\mathbf{H}}$. Then, the attacker can further evaluate the correctness of the solved line impedance by $r_a = \left\| \mathbf{z}_1 - \hat{\mathbf{H}}\left(\hat{\mathbf{H}}^T\hat{\mathbf{H}}\right)^{-1}\hat{\mathbf{H}}^T\mathbf{z}_1 \right\|_2$. If $r_a < r_{th}$, it indicates the attacker obtains the actual line impedance under the MTD. If the estimated residual is larger than the threshold, i.e., $r_a > r_{th}$, the alert attacker needs to change an MTD operation method until the system operator's current MTD operation model is found and correct line impedance is obtained. Then, the attacker can launch the traditional FDI attacks using $\hat{\mathbf{H}}$.

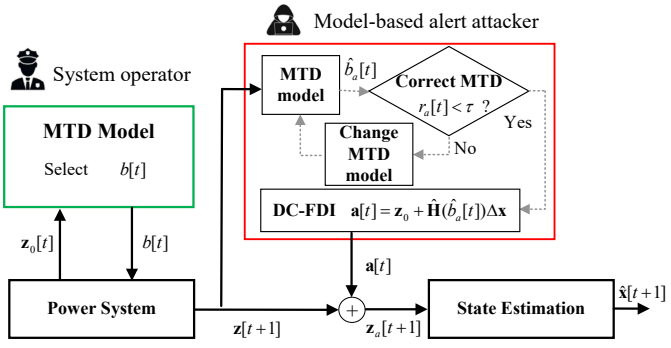


Fig. 3. The attack logic flowchart of the model-based attacker.

3. Random-based HMTD

In this section, we first derive a novel hiddenness operation condition of HMTD, and then propose a novel RHMTD operation model. Finally, we prove that the proposed RHMTD is hidden to three alert attackers, and analyze the attack detection effectiveness of RHMTD.

3.1. Hiddenness Operation Condition

Assume MTD changes the line impedance of the transmission lines. Accordingly, the measurement matrix is changed from $\mathbf{H}_0$ to $\mathbf{H}_1$, and system states are changed from $\boldsymbol{\theta}_0$ to

$\boldsymbol{\theta}_1$. SCADA measurements are changed from $\mathbf{z}_0 = \mathbf{H}_0\boldsymbol{\theta}_0$ (the measurements before MTD) to $\mathbf{z}_1 = \mathbf{H}_1\boldsymbol{\theta}_1$ (the measurements after MTD).

We will use the decomposition of $\mathbf{H}$ matrix to demonstrate the impact of D-FACTS devices on $\mathbf{H}$ and the relationship between $\mathbf{H}_0$ and $\mathbf{H}_1$. First, we separate matrix $\mathbf{H}_0$ into two submatrices, i.e., $\mathbf{H}_0^1$ and $\mathbf{H}_0^2$, which correspond to the measurements related to the lines with and without D-FACTS devices, respectively. Then, we apply the matrix decomposition [12] on $\mathbf{H}_0^1$ and $\mathbf{H}_0^2$, respectively:

$$\mathbf{H}_0 = \begin{bmatrix} \mathbf{H}_0^1 \\ \mathbf{H}_0^2 \end{bmatrix} = \begin{bmatrix} \mathbf{D}_1 \cdot \mathbf{X}_1 \cdot \mathbf{A}_1 \\ \mathbf{D}_2 \cdot \mathbf{X}_2 \cdot \mathbf{A}_2 \end{bmatrix} \tag{2}$$

where $\mathbf{X}_1 \in \mathbb{R}^{p1 \times p1}$ and $\mathbf{X}_2 \in \mathbb{R}^{p2 \times p2}$ are the diagonal reactance matrix of p_1 D-FACTS lines and p_2 non-D-FACTS lines, respectively; $\mathbf{A}_1 \in \mathbb{R}^{n-1 \times p1}$ and $\mathbf{A}_2 \in \mathbb{R}^{n-1 \times p2}$ are the reduced bus-branch incidence matrix of the graphs composed of the D-FACTS lines and non-D-FACTS lines, respectively; $\mathbf{D}_1$ and $\mathbf{D}_2$ is the meter deployment matrix of graph $\mathbf{A}_1$ and $\mathbf{A}_2$, respectively. Here, D-FACTS lines refer to the transmission lines equipped with D-FACTS devices, and non-D-FACTS lines refer to the remaining transmission lines in the power system. Similarly, $\mathbf{H}_1$ can be expressed by (3):

$$\mathbf{H}_1 = \begin{bmatrix} \mathbf{H}_1^1 \\ \mathbf{H}_1^2 \end{bmatrix} = \begin{bmatrix} \mathbf{D}_1 \cdot \mathbf{X}_1' \cdot \mathbf{A}_1 \\ \mathbf{D}_2 \cdot \mathbf{X}_2 \cdot \mathbf{A}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{D}_1 \cdot (\mathbf{X}_1 + \Delta\mathbf{X}) \cdot \mathbf{A}_1 \\ \mathbf{D}_2 \cdot \mathbf{X}_2 \cdot \mathbf{A}_2 \end{bmatrix} \tag{3}$$

where $\mathbf{X}'$ is the diagonal reactance matrix of D-FACTS lines after D-FACTS devices modify the line reactance; and $\Delta\mathbf{X}$ is the incremental line reactance matrix, i.e., $\Delta\mathbf{X} = \mathbf{X}_1' - \mathbf{X}_1$. (2) and (3) intuitively demonstrate the impact of MTD on the measurement matrix. MTD only modified the submatrix of the measurement matrix related to the D-FACTS devices.

According to [6], HMTD remains hidden to BDD-based attackers by remaining all measurements unchanged after the setpoint changes of D-FACTS devices, i.e., $\mathbf{z}_0 = \mathbf{z}_1$. In the noiseless condition, the unchanged measurement condition can be reformulated:

$$\begin{bmatrix} \mathbf{H}_0^1 \\ \mathbf{H}_0^2 \end{bmatrix} \boldsymbol{\theta}_0 = \begin{bmatrix} \mathbf{H}_1^1 \\ \mathbf{H}_1^2 \end{bmatrix} (\boldsymbol{\theta}_0 + \Delta\boldsymbol{\theta}) \tag{4}$$

where $\Delta\boldsymbol{\theta}$ is the incremental state by MTD, i.e., $\Delta\boldsymbol{\theta} = \boldsymbol{\theta}_1 - \boldsymbol{\theta}_0$. When we substitute (2) and (3) into (4), we can obtain:

$$\begin{cases} \mathbf{D}_1 \cdot \mathbf{X}_1 \cdot \mathbf{A}_1 \cdot \boldsymbol{\theta}_0 = \mathbf{D}_1 \cdot (\mathbf{X}_1 + \Delta\mathbf{X}) \cdot \mathbf{A}_1 \cdot (\boldsymbol{\theta}_0 + \Delta\boldsymbol{\theta}) \\ \mathbf{D}_2 \cdot \mathbf{X}_2 \cdot \mathbf{A}_2 \cdot \Delta\boldsymbol{\theta} = 0 \end{cases} \tag{5}$$

Since $\mathbf{D}_2 \cdot \mathbf{X}_2 \cdot \mathbf{A}_2$ is a fixed matrix, $\Delta\boldsymbol{\theta}$ determined by HMTD should belong to the null space of $\mathbf{D}_2 \cdot \mathbf{X}_2 \cdot \mathbf{A}_2$, i.e., $\Delta\boldsymbol{\theta} \in Null(\mathbf{D}_2 \cdot \mathbf{X}_2 \cdot \mathbf{A}_2)$. Thus, $\Delta\boldsymbol{\theta}$ can be represented by the kernel bases of $\mathbf{D}_2 \cdot \mathbf{X}_2 \cdot \mathbf{A}_2$. Therefore, the hiddenness condition of the HMTD can be summarized as follows:

$$\mathbf{D}_1 \cdot \mathbf{X}_1 \cdot \mathbf{A}_1 \cdot \boldsymbol{\theta}_0 = \mathbf{D}_1 \cdot (\mathbf{X}_1 + \Delta\mathbf{X}) \cdot \mathbf{A}_1 \cdot (\boldsymbol{\theta}_0 + \mathbf{KW}) \tag{6}$$

where $\mathbf{K} = [k_1, k_2, ..., k_s] \in \mathbb{R}^{p1 \times s}$ is the matrix of kernel bases of $\mathbf{D}_2 \cdot \mathbf{X}_2 \cdot \mathbf{A}_2$; $\mathbf{W} = [w_1, w_2, ..., w_s]^T \in \mathbb{R}^s$ is the weight determined by the system operator; and s is the dimension of kernel bases.

3.2. The Random-based HMTD model

In order to remain stealthy to three alert attackers and ensure the attack detection effectiveness, an HMTD operation model should simultaneously meet the following four requirements. First, for the BDD-based alert attacker, the measurements need to remain unchanged before and after the implementation of MTD. Essentially, the setpoints of D-

FACTS devices in the HMTD operation model should satisfy the derived hiddenness condition (6). Secondly, for the data-driven alert attacker, MTD ought to avoid introducing distinct changes in measurements. Note that this requirement is less restrictive than that of the BDD-based alert attacker. Thirdly, for the model-based alert attacker, it is necessary to introduce unpredicted randomness into the HMTD operation model. In this case, even though the model-based alert attacker applies the same HMTD operation algorithm used by the system operator, the attacker still fails to obtain the actual line reactance dispatched by the system operator. Finally, sufficient line reactance changes are needed to guarantee a fast and effective attack detection capability [12].

We propose a non-convex, nonlinear, optimization-based RHMTD operation model in (7), which aims to remain stealthy to three alert attacker models and ensure the attack detection effectiveness. The proposed RHMTD model maximizes the weighted square of the line reactance changes using uniformly distributed random weights. The maximized line reactance changes ensure the attack detection effectiveness, while the random weights contribute to providing uncertainties to the model-based alert attacker. Constraint (7.1) is the derived hiddenness condition, which ensures the RHMTD to be hidden to the BDD-based and data-driven alert attackers. Constraint (7.2) defines the kernel bases of $\mathbf{D}_2 \cdot \mathbf{X}_2 \cdot \mathbf{A}_2$ for the hiddenness condition. Constraint (7.3) is the physical constraint of D-FACTS devices working setpoints. Generally, the MTD magntiude μ is 0.2 [6].

$$\max_{\Delta\mathbf{X},\mathbf{W}} \quad \mathrm{diag}(\Delta\mathbf{X})^T \boldsymbol{\lambda} \mathrm{diag}(\Delta\mathbf{X}) \tag{7}$$

$$s.t. \quad \mathbf{D}_1 \cdot \mathbf{X}_1 \cdot \mathbf{A}_1 \cdot \boldsymbol{\theta}_0 = \mathbf{D}_1 \cdot (\mathbf{X}_1 + \Delta\mathbf{X}) \cdot \mathbf{A}_1 \cdot (\boldsymbol{\theta}_0 + \mathbf{KW}) \tag{7.1}$$

$$\mathbf{K} = Null(\mathbf{D}_2 \cdot \mathbf{X}_2 \cdot \mathbf{A}_2) \tag{7.2}$$

$$-\mu \mathrm{diag}(\mathbf{X}_1) \le \mathrm{diag}(\Delta\mathbf{X}) \le \mu \mathrm{diag}(\mathbf{X}_1) \tag{7.3}$$

where the weight parameter $\boldsymbol{\lambda}$ is random variables following the uniform distribution between 0 and 1, i.e., $\boldsymbol{\lambda}_i \in U(0,1)$, i = 1,2,..., $\|\mathbf{X}_1\|_0$.

The RHMTD operation model can be seamlessly integrated into the existing energy management system of the power system. The defender, i.e., the system operator, can assign the weight and then calculate the setpoints of the D-FACTS devices by solving model (7) after the optimal power flow (OPF) function determines the optimal generation. Then, the D-FACTS setpoints are sent to the field devices for implementation through encrypted communication.

3.3. Hiddenness of the RHMTD against alert attackers

In this section, we prove the hiddenness of the proposed method to three alert attackers. Assume the measurements before the MTD are $\mathbf{z}_0 = \mathbf{H}_0 \mathbf{x}_0$, and the measurements after the RHMTD is $\mathbf{z}_1 = \mathbf{H}_1 \mathbf{x}_1$, where $\mathbf{H}_1$ is determined by (7). Note that $\mathbf{z}_0 = \mathbf{z}_1$ holds in the noiseless condition due to the hiddenness operation constraints.

Theorem 1. *The RHMTD model is hidden to the BDD-based alert attacker.*

Proof. The BDD-based alert attacker uses the system configuration $\mathbf{H}_0$ to calculate the estimation residual, and the estimation residual of the proposed RHMTD is zero in the noiseless condition, as follows. Thus, RHMTD is hidden to the BDD-based alert attacker.

$$\begin{aligned} r_a &= \left\| \mathbf{z}_1 - \mathbf{H}_0 (\mathbf{H}_0^T \mathbf{H}_0)^{-1} \mathbf{H}_0^T \mathbf{z}_1 \right\|_2 = \left\| \mathbf{z}_0 - \mathbf{H}_0 (\mathbf{H}_0^T \mathbf{H}_0)^{-1} \mathbf{H}_0^T \mathbf{z}_0 \right\|_2 \\ &= \left\| \mathbf{z}_0 - \mathbf{H}_0 (\mathbf{H}_0^T \mathbf{H}_0)^{-1} \mathbf{H}_0^T \mathbf{H}_0 \mathbf{x}_0 \right\|_2 = \left\| \mathbf{z}_0 - \mathbf{H}_0 \mathbf{x}_0 \right\|_2 = 0 \end{aligned} \tag{8}$$

□

Theorem 2. *The RHMTD is hidden to the data-driven alert attacker.*

Proof. The data-driven alert attacker collects a set of historical measurements to conduct the UL detection. It is assumed that the attacker arranges all eavesdropped measurement vectors of T time instants into a historical measurement matrix $\mathbf{Z}^{Hist} = [\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_T]$, where $\mathbf{Z}^{Hist} \in \mathbb{R}^{m \times T}$. Let us separate T time instants into two parts, i.e., $T = T_1 + T_2$, and accordingly, let $\mathbf{Z}_1$ and $\mathbf{Z}_2$ be the historical measurement matrix of T_1 and T_2 time instants, respectively. When there are no MTDs applied in the system over T times instants, the historical measurement matrix is denoted by $\mathbf{Z}_0^{Hist} = \begin{bmatrix} \mathbf{Z}_{1,0} & \mathbf{Z}_{2,0} \end{bmatrix}$. The data-driven attacker first applies the PCA on $\mathbf{Z}_0^{Hist}$ to reduce the dimension, and cluster the the low-dimension data $\mathbf{Z}_0^{PCA}$ as follows:

$$\mathbf{Z}_0^{PCA} = PCA(\mathbf{Z}_0^{Hist}) \tag{9}$$

$$y_i = Cluster(\mathbf{Z}_0^{PCA}) \tag{10}$$

where y_i is the cluster index of i-th dimension-reduced measurement vector.

Assume the RHMTD model is applied since T_2–th time instants, and the historical measurement matrix collected by the attacker becomes $\mathbf{Z}_{RH}^{Hist} = \begin{bmatrix} \mathbf{Z}_{1,0} & \mathbf{Z}_{2,H} \end{bmatrix}$. Due to the hiddenness operation condition, the measurement vector in the T_2 time instants remain unchanged with and without RHMTD, i.e., $\mathbf{Z}_{2,0} = \mathbf{Z}_{2,H}$. Thus, $\mathbf{Z}_0^{Hist} = \mathbf{Z}_{RH}^{Hist}$ holds. Then, the dimension-reduced vectors of $\mathbf{Z}_{RH}^{Hist}$ are same as that of $\mathbf{Z}_0^{Hist}$, i.e., $PCA(\mathbf{Z}_{RH}^{Hist}) = \mathbf{Z}_0^{PCA}$. Since the input of clustering algorithm remain unchanged, the RHMTD will not change the clustering results. Therefore, the proposed RHMTD is hidden to the data-driven alert attacker. □

Theorem 3. *The RHMTD is hidden to the model-based alert attacker.*

Proof. It is assumed the model-based alert attacker applies the RHMTD model (7) using the eavesdropped measurements $\mathbf{z}_1$, and obtains the system configuration $\hat{\mathbf{H}}$. Even though the input measurement of RHMTD model conducted by the system operator and the attacker are the same ($\mathbf{z}_0 = \mathbf{z}_1$), different weights result in different D-FACTS setpoints, i.e., $\mathbf{H}_{RH} \neq \hat{\mathbf{H}}$. Due to the hiddenness condition, $\mathbf{z}_1 = \hat{\mathbf{H}}\mathbf{x}_2$ holds. The estimation residual computed by the model-based alert attacker using $\hat{\mathbf{H}}$ is zero as follows.

$$\begin{aligned} r_a &= \left\| \mathbf{z}_1 - \hat{\mathbf{H}}\left(\hat{\mathbf{H}}^T\hat{\mathbf{H}}\right)^{-1}\hat{\mathbf{H}}^T\mathbf{z}_1 \right\|_2 = \left\| \mathbf{z}_1 - \hat{\mathbf{H}}\left(\hat{\mathbf{H}}^T\hat{\mathbf{H}}\right)^{-1}\hat{\mathbf{H}}^T\hat{\mathbf{H}}\mathbf{x}_2 \right\|_2 \\ &= \left\| \mathbf{z}_1 - \hat{\mathbf{H}}\mathbf{x}_2 \right\|_2 = 0 \end{aligned} \tag{11}$$

Note that if the attacker happens to use the same weight as that used by the system operator, $\mathbf{H}_{RH} = \hat{\mathbf{H}}$ holds. However, it doesn't impact the hiddenness of the RHMTD in Theorem 3. It only degrades the attack detection effectiveness of the proposed RHMTD, but it happens with very low probability. □

3.4. Detection Effectiveness of the RHMTD against alert attackers

In this section, we analyze the attack detection effectiveness of RHMTD against the attacks by the BDD-based and model-based alert attackers due to the straightforward analysis, and then prove that the RHMTD has the maximum detection effectiveness against the PCA-FDI attacks by the data-driven attackers.

For the BDD-based alert attacker, the stealthiness of the RHMTD misleads the attacker to adopt the traditional FDI attacks without the aid of topology learning. It is proved that the placement of D-FACTS determines the attack detection effectiveness of MTD against the traditional FDI attacks [12]. The max-rank HMTD placement [5] adopted in this paper guarantees the maximum attack detection effectiveness under the assumption that the reactance of all D-FACTS lines is changed by the D-FACTS devices. This assumption is satisfied by the RHMTD operation by maximizing the line reactance changes introduced

by D-FACTS devices. Therefore, the RHMTD under the max-rank HMTD placement has the maximum detection effectiveness of the attacks by the BDD-based attackers.

The model-based alert attacker constructs FDI attacks using $\hat{\mathbf{H}}$ under the HRMTD. According to the MTD detection effectiveness metric [6], [12], [13], the detection effectiveness of RHMTD with $\mathbf{H}_{RH}$ against the model-based alert attacker depends on the rank of the composite matrix, i.e., $rank([\hat{\mathbf{H}} \quad \mathbf{H}_{RH}])$. We can apply the graph-theory analysis on deriving the value of $rank([\hat{\mathbf{H}} \quad \mathbf{H}_{RH}])$. Note that in the D-FACTS placement problem, it is the difference between the original line reactance (attacker's knowledge) and defender's dispatched line reactance that determines the detection effectiveness. However, it is the difference between the attacker's estimated line reactance $\hat{b}_a$ (attacker's knowledge) and defender's dispatched line reactance b that plays an important role in the detection effectiveness of the model-based attacker's attacks. Thus, we treat the difference between the attacker's estimated line reactance and the defender's dispatched line reactance, i.e., $\hat{b}_a - b$, as the contribution of D-FACTS devices. If $\hat{b}_a(i) = b(i)$ holds for the i-th D-FACTS line, it indicates the D-FACTS device on this line does not exist from the perspective of the alert attacker, referred to as the equivalently removed D-FACTS line hereafter; if $\hat{b}_a(i) \neq b(i)$, the D-FACTS device on this line works. In this case, if $\hat{b}_a \neq b$ holds for all D-FACTS lines, the adopted max-rank HMTD placement ensures the maximum attack detection effectiveness i.e., $\max\left(rank([\hat{\mathbf{H}} \quad \mathbf{H}_{RH}])\right) = p$ based on the graph-theory analysis of MTD [12]. If the attacker accurately estimates the defender's dispatched reactance of some D-FACTS lines, the rank of the composite matrix in MTDs is determined by the number of loops in G_a as follows:

$$rank([\hat{\mathbf{H}} \quad \mathbf{H}_{RH}]) = p - lp_{\overline{DF}} \tag{12}$$

where $lp_{\overline{DF}}$ is the number of loops in G_a and G_a is a graph constructed from the view of the attackers, consisting of all buses, non-D-FACTS lines and equivalently removed D-FACTS lines.

For the data-driven alert attacker, the hiddenness of RHMTD misleads the alert attacker to estimate the principle components of $\mathbf{H}_0$ before the MTD, rather than that of the actual $\mathbf{H}_{RH}$. Consequently, the stealthiness of the PCA-FDI attack greatly degrades. Specifically, the stealthiness of the PCA-FDI attacks depends on the difference between column space of $\mathbf{H}_0$ and that of $\mathbf{H}_{RH}$. In Theorem 4, we prove that the proposed RHTMD under the max-rank HMTD placement maximizes the difference between the column space of $\mathbf{H}_0$ and that of $\mathbf{H}_{RH}$. such that the stealthy attack space is minimized.

Theorem 4. *The RHMTD model has the maximized attack detection probability to the PCA-FDI attack by the data-driven alert attacker.*

Proof. The alert attacker collects historical measurements under RHMTD over T times, and the historical measurement matrix is denoted by $\mathbf{Z}_{RH}^{Hist}$. Similar to the proof of Theorem 3, $\mathbf{Z}_0^{Hist} = \mathbf{Z}_{RH}^{Hist}$ holds in the noiseless condition due to the hiddenness of the RHMTD. Therefore, the estimated $\mathbf{H}$ matrix under the RHMTD $H_{RH}^{PCA} = PCA(\mathbf{Z}_{RH}^{Hist})$ is the same as that without MTD $H_0^{PCA} = PCA(\mathbf{Z}_0^{Hist})$ in the noiseless condition, i.e., $H_{RH}^{PCA} = H_0^{PCA}$. Then, PCA-FDI attacks are constructed by $a = H_{RH}^{PCA} c = H_0^{PCA} c$.

According to the principle of FDI attack [6], if the attack vector a belongs to the column space of $\mathbf{H}_{RH}$, i.e., $a \in col(\mathbf{H}_{RH})$, the constructed PCA-FDI attack is stealthy to RHMTD. Specifically, a PCA-FDI attack is stealthy to RHMTD, if $a \in col(\mathbf{H}_{RH}) \bigcap col(\mathbf{H}_0^{PCA})$. Then, the dimension of the stealthy attack space can be expressed as:

$$\begin{aligned}&\left|col(\mathbf{H}_{RH})\bigcap col(\mathbf{H}_0^{PCA})\right|\\&=r(\mathbf{H}_{RH})+r(\mathbf{H}_0^{PCA})-r([\mathbf{H}_0^{PCA}\quad \mathbf{H}_{RH}])\\&=2\times(n-1)-r([\mathbf{H}_0^{PCA}\quad \mathbf{H}_{RH}])\end{aligned}\tag{13}$$

Since attacker's $\mathbf{H}_0^{PCA}$ is unknown to the system operator, it is assumed that the attacker can accurately approximates the column space of $\mathbf{H}_0$, i.e., $col(\mathbf{H}_0^{PCA})=col(\mathbf{H}_0)$. Then, the dimension of stealthy attack space becomes:

$$\left|col(\mathbf{H}_{RH})\bigcap col(\mathbf{H}_0^{PCA})\right|=2\times(n-1)-r([\mathbf{H}_0\quad \mathbf{H}_{RH}])$$

The adopted max-rank HMTD placement guarantees the maximum value of $r([\mathbf{H}_0\quad \mathbf{H}_{RH}])$. Therefore, the dimension of stealthy attack space is minimized under the RHMTD. Therefore, RHMTD has the maximized attack detection effectiveness to PCA-FDI attacks by the data-driven alert attacker. □

4. Numerical Results

4.1. Test Systems

We evaluate the HMTD operation model in the IEEE 14-bus system [29]. We solve the HMTD operation model using fmincon function of MATLAB. We use MATLAB to simulate the BDD-based and model-based alert attackers and use Python to simulate the data-driven alert attacker. The measurement noise is assumed to be Gaussian distributed with zero mean and the standard deviation as 1% of the actual measurement. The threshold of the Chi-square detector in the BDD used by attackers and defenders is set to have a 0.1% false-positive rate.

4.2. Uncertainties of RHMTD

First, we demonstrate the effectiveness of random weights in providing uncertainties to the line reactance in the RHMTD. Under the same load condition, we conduct the RHMTD operation model for 20 times using different weights. Fig. 4 shows the dispatched line reactance of each D-FACTS line in the 20 RHMTDs. It is seen the reactance of each D-FACTS lines in the 20 RHMTD are different. The uncertainties can contribute to the hiddenness and detection effectiveness of RHMTD to the model-based attacker. However, for some RHMTDs, they have similar reactance of 7-th D-FACTS lines, which will negatively impact the detection effectiveness of RHMTD against the attacks by the model-based alert attacker. This impact will be evaluated in Section 4.4.

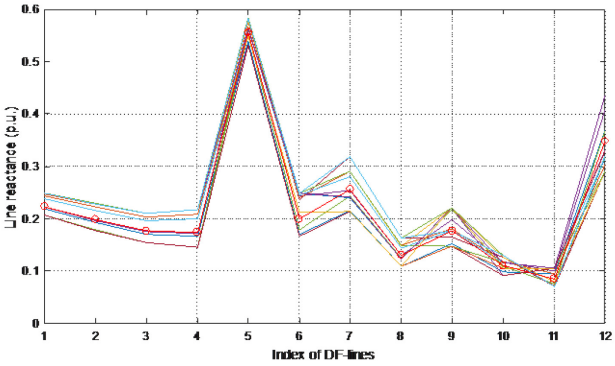


Fig. 4. The reactance of D-FACTS-lines in 20 RHMTDs under a given load.

We utilize the L1-norm distance between the line reactance generated by the system operator and that by the model-based attacker to measure the uncertainties in the RHTMD. Based on the distance, we demonstrate the impact of MTD magnitude on the uncertainties. Under each MTD magnitude, we generate one RHMTD operation point for the system

operator as the reference, and then generate 50 RHMTD operation points as the model-based alert attacker's estimation by running the RHMTD model. Figure 5 shows the boxplot of the L1-norm distance under different MTD magnitudes. It is seen that a larger MTD magnitude generally results in a large L1-norm distance. The median of the L1-norm distance under 0.2 MTD magnitude is lower than that under 0.18. It indicates that a larger MTD magnitude doesn't guarantee a larger L1-norm distance or a better attack detection effectiveness against the model-based alert attacker due to the random uncertainties.

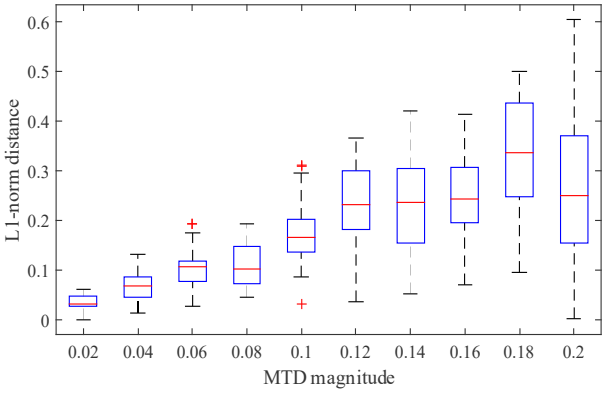


Fig. 5. L1-norm distance under MTD magnitudes.

4.3. Hiddenness of RHMTD against Three Alert Attackers

In this section, we evaluate the hiddenness of RHMTD to three alert attackers. First, we evaluate the hiddenness of RHMTD to the BDD-based alert attacker by comparing the defense stealthiness probability (DSP) of RHMTD and RMTD under different MTD magnitudes. The DSP is a widely used metric to measure the MTD hiddenness from the perspective of attackers, which is defined as the ratio of the number of MTDs hidden to attackers to the total number of launched MTDs.

To study the impact of MTD magnitude on the MTD hiddenness, we increase the MTD magnitude from 0.02 to 0.2 with an incremental of 0.02. For each MTD magnitude, we generate 100 RMTDs and 100 RHMTDs under different load conditions, respectively. In addition, we repeat this MTD generation process under two different noise conditions to evaluate the impact of noise on the MTD hiddenness. The DSP of RMTD and RHMTD against the BDD-based alert attacker is shown in Fig. 6. As seen, when the MTD magnitude is small (less than 0.04), it is likely that RMTD remains hidden to the attacker. This is because the tiny line reactance mismatch has limited capability to increase the estimation residual in the attacker's BDD. With the increase of MTD magnitude, the DSP drops to zero, indicating that RMTD is no longer hidden to the attacker. For the RHTMD, its DSP is larger than 0.95 regardless of MTD magnitudes and noise standard deviation, indicating the hiddenness to the BDD-based attacker.

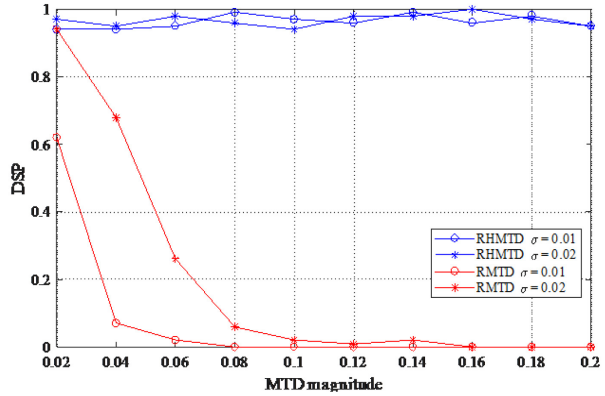


Fig. 6. The hiddenness of RMTD and RHMTD against the BDD-based alert attacker under different noise conditions.

Then, we evaluate the hiddenness of RHMTD to the model-based alert attacker under different noise conditions. We generate 100 RHMTD operation setpoints under different load conditions under MTD magnitudes from 0.02 to 0.2 with an incremental of 0.02. The measurements of RHMTD are sampled in the noiseless condition and noisy conditions with standard deviation $\sigma = 1\%$, $\sigma = 2\%$, and $\sigma = 3\%$, respectively. It is assumed that the model-based attacker applies the RHMTD model (7) to estimate the line reactance dispatched in the field, and then applies SE to calculate the estimation residual to detect the existence of MTD. The DSP of RHMTD against the model-based alert attacker is shown in Fig. 7. In the noiseless condition, the DSP of RHMTD is always 1.0 regardless of MTD magnitudes. In noisy conditions, the DSP of RHMTD is more than 95%. It is seen that MTD magnitude and noise magnitude don't impact the hiddenness of RHMTD.

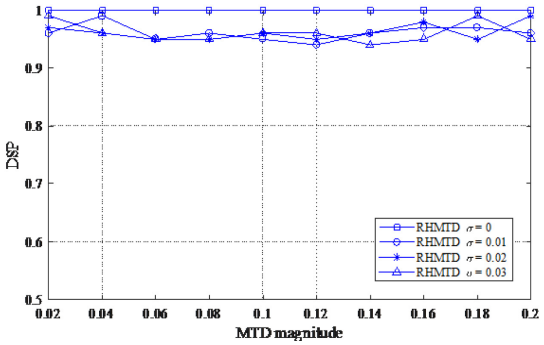
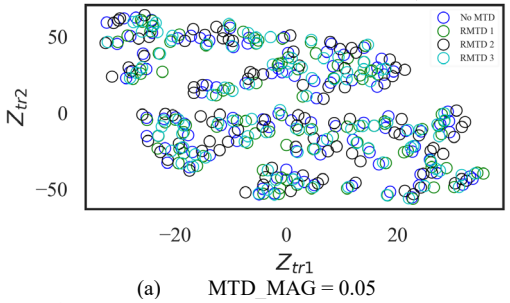
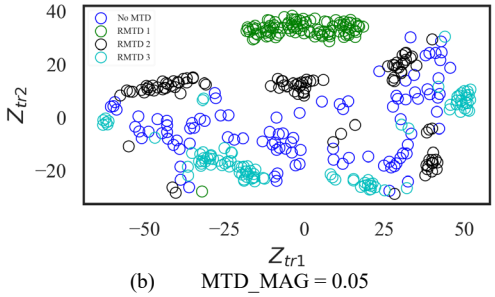


Fig. 7. The hiddenness of RHMTD against the model-based alert attacker under noiseless and noise conditions.

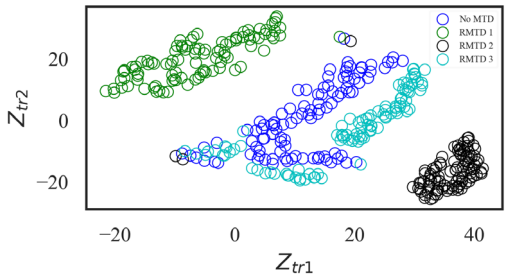
Finally, we demonstrate the drawbacks of RMTD against the data-driven alert attacker, and further evaluate the hiddenness of RHMTD to the data-driven alert attacker. To simulate historical measurements free from MTD collected by the data-driven alert attacker, the power flow problem is solved for multiple time instants. In this paper, we use 100 load conditions to generate historical measurements of 100 time instants. First, we generate 3 RMTD groups under 100 different load conditions under 0.05, 0.10, and 0.15 MTD magnitudes, respectively. Specifically, let RMTD 1, RMTD 2, and RMTD 3 refer to these generated RMTD groups, and RMTD i (i = 1, 2, 3) has 100 different operation setpoints for each MTD magnitude. After the SCADA measurements are collected by the attacker, a dimension reduction algorithm, i.e., PCA, is applied on the 100 measurement vectors under RMTDs and 100 measurement vectors free of MTD to visualize the difference between the normal data (no MTD measurements) and MTD measurements.



(a) MTD_MAG = 0.05



(b) MTD_MAG = 0.05



(c) MTD_MAG = 0.15

Fig. 8. The projection of RMTD and no MTD measurements in the $\mathbb{R}^2$ space by PCA under different MTD magnitudes.

The projection of RMTD and no MTD measurement data in the $\mathbb{R}^2$ space under different MTD magnitudes are shown in Fig. 8. When the MTD magnitude is 0.05, the RMTD data points and no MTD data points are overlapped, indicating data-driven alert attacker cannot detect the existence of RMTD. When the MTD magnitude becomes 0.10, RMTD 1 is projected into a new cluster, while RMTDs 2 and 3 are still overlapped with no MTD data. When the MTD magnitude increases to 0.15, RMTDs 1 and 2 forms two new clusters, and data points of RMTD 3 also remain separated from no MTD data. For the data-driven attacker, a new cluster indicates the detection of MTD. Thus, the hiddenness of RMTD degrades with the increase of MTD magnitude, which is consistent with the performance of RMTD against the BDD-based alert attacker.

To evaluate the hiddenness of RHMTD, we apply RHMTD algorithm under 100 load conditions with 0.20 MTD magnitude. For comparison, we also generate 10 RMTD groups with 0.20 MTD magnitude. The projection of RHMTD, 10 RMTD and no MTD measurements in the $\mathbb{R}^2$ space is shown in Fig. 9. As seen, under 0.20 MTD magnitude, all RMTD groups form new clusters that locate far from the cluster of no MTD measurements. All data points of RHMTD remain inside of the cluster of the no MTD, as shown in Fig. 10. Therefore, these RHMTD are stealthy to the data-driven attacker. Since RHMTDs with 0.20 MTD magnitude could remain stealthy, it infers that the RHMTD with a smaller MTD magnitude could also remain stealthy, according to the impact of the MTD magnitude on the MTD stealthiness.

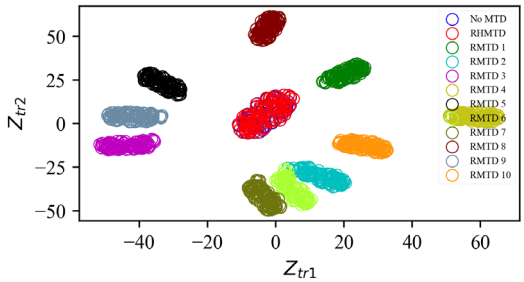


Fig. 9. The projection of RHMTD, 10 RMTD and no MTD measurements in the $\mathbb{R}^2$ space.

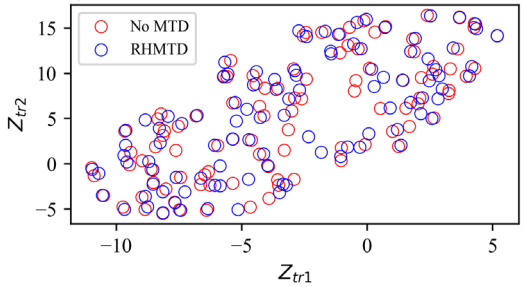


Fig. 10. The projection of RHMTD and no MTD measurements in the $\mathbb{R}^2$ space.

We compare the hiddenness of RHMTD with two existing HMTD methods, i.e., watermarking HMTD [21] and model-based HMTD [4] against three alert attacker models. We can see that the proposed RHMTD is hidden to three alert attackers, consistent with

the hiddenness theorems in Section 3.3. It is seen that these three HMTD methods are all hidden to BDD-based alert attackers. This is because, as the first proposed alert attacker in the literature, these HMTD methods consider the estimation residual changes in the alert attacker's BDD. All three HMTD methods are hidden to data-driven alert attackers since these HMTD methods avoid significant measurement changes before and after MTD. For the proposed model-based alert attacker, the DSP of watermarking HMTD is lower than its DSP against BDD-based alert attacker. It is because the randomness in the watermarking HMTD makes the attacker's estimated line parameters different from the actual dispatched parameters. The difference results in an increase in the attacker's estimation residual. Even though the model-based HMTD is hidden to the model-based alert attacker, the attacker can accurately estimate actual dispatched line parameters due to the lack of randomness in model-based HMTD. As a consequence, the model-based HMTD cannot detect the attacks by the model-based alert attacker, which is shown in Fig. 12 and Table 4.

Table 3. DSP of existing HMTD methods and the RHMTD against three alert attackers.

Method	BDD-AA	DD-AA	M-AA
Watermarking HMTD	94%	100%	83%
Model-based HMTD	93%	100%	96%
RHMTD	95%	100%	96%

4.4. Attack Detection Effectiveness of the RHMTD against Three Alert Attackers

In this subsection, we evaluate the attack detection effectiveness of the RHMTD against three alert attackers. First, we prepare the defense pool of RHMTD. We increase the MTD magnitude from 0.02 to 0.2 with an incremental of 0.02, and then generate 100 RHMTD operation setpoints for each MTD magnitude. In total, there are 1000 RHMTD operation setpoints as the defense pool. In the simulation, the widely-used attack detection probability is applied to measure the attack detection effectiveness of an MTD, which is defined as the ratio of number of FDIs detected by the MTD to the total number of FDI attacks.

For the BDD-based attacker, RHMTD misleads the attacker to construct traditional FDI attacks without the aid of topology learning. Therefore, the BDD-based alert attacker constructs 100 single-bus FDI attacks using $\mathbf{H}_0$ for each RHMTD in the defense pool. The ADP of RHMTD against the BDD-based alert attacker under different MTD magnitudes is shown in Fig. 11. The ADP increases with the MTD magnitude. This is because the tiny line changes cannot cause sufficient residual incremental in the defender's BDD, and thus the ADP under the low MTD magnitudes is low. When the MTD magnitude is larger than 0.08, the ADP becomes 93.3%. This is because MTD can not detect the single-bus FDI attack on Bus 8, which only has one transmission line. This is the drawback of MTD identified by our previous work [30].

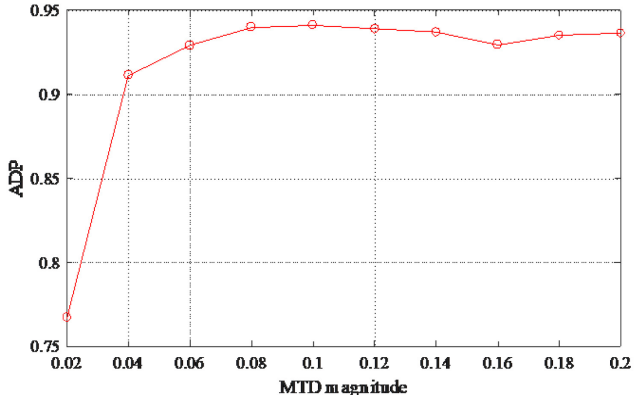


Fig. 11. The ADP of RHMTD against the BDD-based alert attacker under different MTD magnitudes.

The model-based attacker utilizes the same RHMTD model to estimate the reactance of D-FACTS lines and uses $\hat{\mathbf{H}}$ to construct FDI attacks. We compare the detection effectiveness of the RHTMD and HMTD against the model-based attacker. We generate 100 HMTD under each MTD magnitude. For each HMTD and RHMTD, the model-based attacker launches 100 FDI attacks. The ADP of RHMTD and HMTD against the model-based alert attacker under different MTD magnitudes is shown in Fig. 12. It is seen that HMTD cannot detect the attacks by the model-based attacker. The lack of uncertainties causes the model-based attacker can accurately estimate the reactance of D-FACTS lines, and the attacks can bypass the defender's BDD. Compared with the low ADP of HMTD, the ADP of the RHMTD can reach 80%. We can see that the ADP of RHMTD against the model-based attacker is lower than the ADP of RHMTD against the BDD-based attacker. This is because the reactance of some D-FACTS lines estimated by the attacker is very close to the actual reactance dispatched by the defender. For a single-bus FDI attack by the model-based attacker, if line parameters of all connected lines associated with the target bus are accurately or approximately estimated, the FDI attack is very likely to remain stealthy to the RHMTD. The detection effectiveness of RHMTD against the model-based alert attacker is analyzed in Section 3.4.

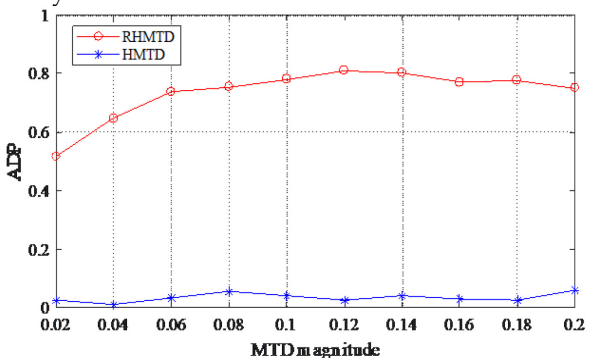


Fig. 12. The ADP of RHMTD and HMTD against the model-based alert attacker under different MTD magnitudes.

The data-driven attacker constructs PCA-FDI attacks under RHMTD with 0.2 MTD magnitude. It is assumed that the attacker collects the historical measurements of 5000-time instants. The RHMTD is conducted under each time instant. In the PCA-FDI attacks, the number of the attacked buses are 1, 3, and 5, respectively. Here, the incremental voltage of the PCA-FDI attack is defined as $\mathbf{c} = k \times \boldsymbol{\theta}_0$, where $\boldsymbol{\theta}_0$ is the actual voltage angle of the power system at the attacked time instant, and k is the FDI magnitude varying from 0.05 to 0.4. The ADP of RHMTD against the PCA-FDI attacks by the data-driven alert attacker is shown in Fig. 13. It is seen that the ADP increases with the FDI attack magnitude and the number of attacked buses.

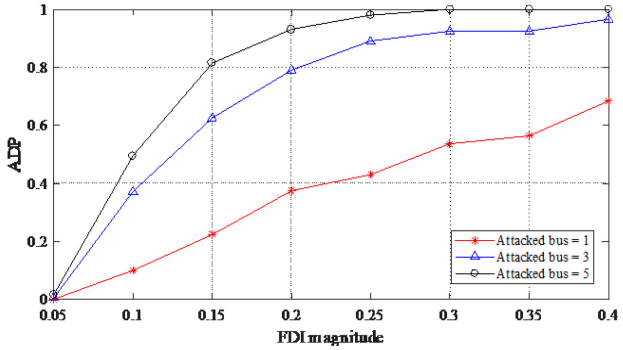


Fig. 13. The ADP of RHMTD against the data-driven alert attacker.

We compare the attack detection effectiveness of RHMTD with two existing HMTD methods against three alert attacker models. We use the Chi-2 detector for three HMTD methods to detect FDI attacks. Due to small line parameter changes, the watermarking HMTD has very low ADP against three attackers. Model-based HMTD has the same ADP as RHMTD against the BDD-based alert attacker. However, its ADP against the model-based alert attacker is close to zero. This is because the attacks constructed by the model-based alert attacker are based on accurately estimated line parameters. RHMTD has higher ADP than other HMTD methods due to its randomness and sufficient line impedance changes.

Table 4. ADP of existing HMTD methods and RHMTD against three alert attackers.

Method	BDD-AA	M-AA	DD-AA
Watermarking HMTD	37.8%	47.3%	45%
Model-based HMTD	93.9%	6.0%	59%
RHMTD	93.6%	75.1%	68%

5. Conclusions

This paper points out the drawbacks of existing HMTD operation methods, including the delay of attack detection, extra costs on secure meters, and the lack of randomness. To fully evaluate the hiddenness of HMTD methods, this paper first summarizes the BDD-based alert attacker model and the data-driven alert attacker model, and then proposes a novel model-based alert attacker model. By analyzing the three alert attackers, this paper proposes a novel random-based HMTD, which maximizes the weighted square of line reactance changes, and introduces random variables into the weights of the objective function. In addition, the proposed model utilizes the novel derived hiddenness operation conditions as constraints to ensure the measurements before and after MTD remain unchanged. We theoretically prove the hiddenness of the proposed RHMTD to three alert attacker models, and analyze the effectiveness of RHMTD in detecting FDI attacks constructed by three alert attackers.

The simulation results show that the random weights in RHMTD successfully introduce the randomness into the setpoints of D-FACTS devices. The randomness makes the model-based alert attacker difficult to accurately estimate the actually dispatched setpoints of D-FACTS devices by the defender. The RHMTD method is hidden to both the BDD-based and model-based alert attackers with more than 95% DSP. The RHMTD method is also hidden to the data-driven alert attacker, since the projection of RHMTD and no MTD measurements overlaps after the dimension reduction. Simulation results also evaluate the detection effectiveness of RHMTD against three alert attackers. The traditional HMTD fails to detect FDI attacks by the model-based alert attacker, while RHTMD can detect these attacks with 80% ADP. RHTMD is effective in detecting FDI attacks by the BDD-based and data-driven alert attackers with more than 90% ADP.

In the future, we will extend the proposed HMTD operation method in the DC power system model to the AC power system model. In addition, we will define more alert adversary models using advanced machine-learning techniques and limited data resources.

Author Contributions: Conceptualization, B.L. and H.W.; methodology, B.L., H.W. Q.Y. and H.Z.; software, B.L. and H.W.; validation, B.L., H.W. Q.Y. and H.Z.; formal analysis, B.L., H.W. Q.Y. and H.Z.; investigation, B.L., H.W. Q.Y. and H.Z.; resources, B.L., H.W. Q.Y. and H.Z.; data curation, B.L., H.W. Q.Y. and H.Z.; writing—original draft preparation, B.L. and H.W.; writing—review and editing, B.L., H.W. Q.Y. and H.Z.; visualization, B.L., H.W. Q.Y. and H.Z.; supervision, B.L., H.W. Q.Y. and H.Z.; project administration, B.L.; funding acquisition, H.W.. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by U.S. National Science Foundation, grant number No. 1929147 and No. 2146156.

Data Availability Statement: Not applicable.

Abbreviations: MTD, Moving Target Defense; FDI, False Data Injection; SE, State Estimation; PCA, Principal component analysis; D-FACTS, distributed flexible AC transmission system; BDD, Bad Data Detection.

References

[1] S. Balouch *et al.*, Optimal Scheduling of Demand Side Load Management of Smart Grid Considering Energy Efficiency, *Frontiers in Energy Research*, vol. 10, May 2022, doi: 10.3389/fenrg.2022.861571.

[2] A. S. Musleh, G. Chen, and Z. Y. Dong, A Survey on the Detection Algorithms for False Data Injection Attacks in Smart Grids, *IEEE Transactions on Smart Grid*, vol. 11, no. 3, pp. 2218–2234, May 2020, doi: 10.1109/TSG.2019.2949998.

[3] H. Zhang, B. Liu, and H. Wu, Smart Grid Cyber-Physical Attack and Defense: A Review, *IEEE Access*, vol. 9, pp. 29641–29659, 2021, doi: 10.1109/ACCESS.2021.3058628.

[4] M. A. Rahman, E. Al-Shaer, and R. B. Bobba, Moving Target Defense for Hardening the Security of the Power System State Estimation, in *Proceedings of the First ACM Workshop on Moving Target Defense*, New York, NY, USA, 2014, pp. 59–68. doi: 10.1145/2663474.2663482.

[5] B. Liu and H. Wu, Optimal Planning and Operation of Hidden Moving Target Defense for Maximal Detection Effectiveness, *IEEE Transactions on Smart Grid*, pp. 1–1, 2021, doi: 10.1109/TSG.2021.3076824.

[6] J. Tian, R. Tan, X. Guan, and T. Liu, Enhanced Hidden Moving Target Defense in Smart Grids, *IEEE Transactions on Smart Grid*, vol. 10, no. 2, pp. 2208–2223, Mar. 2019, doi: 10.1109/TSG.2018.2791512.

[7] M. Liu, C. Zhao, Z. Zhang, and R. Deng, Explicit Analysis on Effectiveness and Hiddenness of Moving Target Defense in AC Power Systems, *IEEE Transactions on Power Systems*, pp. 1–1, 2022, doi: 10.1109/TPWRS.2022.3152801.

[8] H. Zhang, B. Liu, X. Liu, A. Pahwa, and H. Wu, Voltage Stability Constrained Moving Target Defense against Net Load Redistribution Attacks, *IEEE Transactions on Smart Grid*, pp. 1–1, 2022, doi: 10.1109/TSG.2022.3170839.

[9] Z. Zhang, Y. Tian, R. Deng and J. Ma, A Double-Benefit Moving Target Defense Against Cyber–Physical Attacks in Smart Grid, *IEEE Internet of Things Journal*, vol. 9, no. 18, pp. 17912-17925, 15 Sept.15, 2022, doi: 10.1109/JIOT.2022.3161790.

[10] R. Deng, P. Zhuang, and H. Liang, CCPA: Coordinated Cyber-Physical Attacks and Countermeasures in Smart Grid, *IEEE Transactions on Smart Grid*, vol. 8, no. 5, pp. 2420–2430, Sep. 2017, doi: 10.1109/TSG.2017.2702125.

[11] J. Tian, R. Tan, X. Guan, Z. Xu, and T. Liu, Moving Target Defense Approach to Detecting Stuxnet-Like Attacks, *IEEE Transactions on Smart Grid*, vol. 11, no. 1, pp. 291–300, Jan. 2020, doi: 10.1109/TSG.2019.2921245.

[12] B. Liu and H. Wu, Optimal D-FACTS Placement in Moving Target Defense against False Data Injection Attacks, *IEEE Transactions on Smart Grid*, pp. 1–1, 2020, doi: 10.1109/TSG.2020.2977207.

[13] C. Liu, J. Wu, C. Long, and D. Kundur, Reactance Perturbation for Detecting and Identifying FDI Attacks in Power System State Estimation, *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 4, pp. 763–776, Aug. 2018, doi: 10.1109/JSTSP.2018.2846542.

[14] Z. Zhang, R. Deng, P. Cheng, and M.-Y. Chow, Strategic Protection Against FDI Attacks With Moving Target Defense in Power Grids, *IEEE Transactions on Control of Network Systems*, vol. 9, no. 1, pp. 245–256, Mar. 2022, doi: 10.1109/TCNS.2021.3100411.

[15] S. Lakshminarayana and D. K. Y. Yau, Cost-Benefit Analysis of Moving-Target Defense in Power Grids, *IEEE Transactions on Power Systems*, pp. 1–1, 2020, doi: 10.1109/TPWRS.2020.3010365.

[16] B. Liu, Q. Yang, H. Zhang, and H. Wu, An Interior-Point Solver for AC Optimal Power Flow Considering Variable Impedance-Based FACTS Devices, *IEEE Access*, vol. 9, pp. 154460–154470, 2021, doi: 10.1109/ACCESS.2021.3128035.

[17] Z. Zhang, R. Deng, D. K. Y. Yau, P. Cheng, and J. Chen, Analysis of Moving Target Defense Against False Data Injection Attacks on Power Grid, *IEEE Transactions on Information Forensics and Security*, pp. 1–1, 2019, doi: 10.1109/TIFS.2019.2928624.

[18] M. Liu, C. Zhao, Z. Zhang, R. Deng and P. Cheng, Analysis of Moving Target Defense in Unbalanced and Multiphase Distribution Systems Considering Voltage Stability, *2021 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, Aachen, Germany, 2021, pp. 207-213, doi: 10.1109/SmartGridComm51999.2021.9632320.

[19] Z. Zhang, R. Deng, D. K. Y. Yau, P. Cheng, and M.-Y. Chow, Security Enhancement of Power System State Estimation With an Effective and Low-Cost Moving Target Defense, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, pp. 1–16, 2022, doi: 10.1109/TSMC.2022.3222793.

[20] B. Liu, H. Wu, A. Pahwa, F. Ding, E. Ibrahim, and T. Liu, Hidden Moving Target Defense against False Data Injection in Distribution Network Reconfiguration, *2018 IEEE Power Energy Society General Meeting (PESGM)*, Aug. 2018, pp. 1–5. doi: 10.1109/PESGM.2018.8586470.

[21] M. Higgins, F. Teng, and T. Parisini, Stealthy MTD Against Unsupervised Learning-Based Blind FDI Attacks in Power Systems, *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 1275–1287, 2021, doi: 10.1109/TIFS.2020.3027148.

[22] C. Murguia and J. Ruths, CUSUM and Chi-squared attack detection of compromised sensors, in *2016 IEEE Conference on Control Applications (CCA)*, Buenos Aires, Argentina, Sep. 2016, pp. 474–480. doi: 10.1109/CCA.2016.7587875.

[23] Z. Zhang, R. Deng, D. K. Y. Yau, P. Cheng, and J. Chen, On Hiddenness of Moving Target Defense against False Data Injection Attacks on Power Grid, *ACM Trans. Cyber-Phys. Syst.*, vol. 4, no. 3, p. 25:1-25:29, 2020, doi: 10.1145/3372751.

[24] J. Zheng and A. S. Namin, A Survey on the Moving Target Defense Strategies: An Architectural Perspective, *J. Comput. Sci. Technol.*, vol. 34, no. 1, pp. 207–233, Jan. 2019, doi: 10.1007/s11390-019-1906-z.

[25] S. Lakshminarayana, S. Sthapit, and C. Maple, A Comparison of Data-Driven Techniques for Power Grid Parameter Estimation. arXiv, Jul. 08, 2021. doi: 10.48550/arXiv.2107.03762.

[26] H. Yang, X. He, Z. Wang, R. C. Qiu, and Q. Ai, Blind False Data Injection Attacks Against State Estimation Based on Matrix Reconstruction, *IEEE Transactions on Smart Grid*, vol. 13, no. 4, pp. 3174–3187, Jul. 2022, doi: 10.1109/TSG.2022.3164874.

[27] Z.-H. Yu and W.-L. Chin, Blind False Data Injection Attack Using PCA Approximation Method in Smart Grid, *IEEE Transactions on Smart Grid*, vol. 6, no. 3, pp. 1219–1226, May 2015, doi: 10.1109/TSG.2014.2382714.

[28] J. Kim, L. Tong, and R. J. Thomas, Subspace Methods for Data Attack on State Estimation: A Data Driven Approach, *IEEE Transactions on Signal Processing*, vol. 63, no. 5, pp. 1102–1114, Mar. 2015, doi: 10.1109/TSP.2014.2385670.

[29] R. D. Zimmerman, C. E. Murillo-Sanchez, and R. J. Thomas, MATPOWER: Steady-State Operations, Planning, and Analysis Tools for Power Systems Research and Education, *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 12–19, Feb. 2011, doi: 10.1109/TPWRS.2010.2051168.

[30] B. Liu and H. Wu, Systematic planning of moving target defence for maximising detection effectiveness against false data injection attacks in smart grid, *IET Cyber-Physical Systems: Theory & Applications*, *2021, vol. 6, no. 3, pp. 151–163, doi:* 10.1049/cps2.12012.