Cost-Benefit Analysis of Moving-Target Defense in Power Grids

Subhash Lakshminarayana[*] and David K.Y. Yau [*][†]
[*] Advanced Digital Sciences Center, Illinois at Singapore, Singapore 138602
[†] Singapore University of Technology and Design, Singapore 487372
Email: [*]subhash.l@adsc-create.edu.sg, [†] david_yau@sutd.edu.sg

***Abstract*—We study moving-target defense (MTD) that actively perturbs transmission line reactances to thwart stealthy false data injection (FDI) attacks against state estimation in a power grid. Prior work on this topic has proposed MTD based on randomly selected reactance perturbations, but these perturbations cannot guarantee effective attack detection. To address the issue, we present formal design criteria to select MTD reactance perturbations that are truly effective. However, based on a key optimal power flow (OPF) formulation, we find that the effective MTD may incur a non-trivial operational cost that has not hitherto received attention. Accordingly, we characterize important tradeoffs between the MTD's detection capability and its associated required cost. Extensive simulations, using the MATPOWER simulator and benchmark IEEE bus systems, verify and illustrate the proposed design approach that for the first time addresses both key aspects of cost and effectiveness of the MTD.**

I. Introduction

Cyber attacks against critical infrastructures can lead to severe disruptions. The December 2015 attack against the Ukraine's power grid was a real-world example, which caused power outages for a large number of customers for hours [1]. These attacks were typically crafted by sophisticated attackers, sometimes with national backing, who managed to spend considerable time inside a system to learn its operational details, and accordingly designed the injection of malicious data/control to disrupt its operations [2]. It is thus imperative to design counteracting defense approaches to defeat the knowledgeable attackers. Moving-target defense (MTD) [3] is a defense approach that has received increasing attention. It is based on dynamically changing the system parameters that attackers need to target for customizing their attacks, in order to invalidate the attackers' prior knowledge of the system and render ineffective any of their prior designed strategies. It has the potential to make it extremely difficult or impossible for would-be attackers to keep up with the system dynamics.

In this paper, we focus on *false data injection* (FDI) attacks against state estimation (SE) in power grids. SE is a key method for grid operators to obtain a best estimate of the system state from noisy sensor measurements collected via a supervisory control and data acquisition (SCADA) system, for example. Its output is used in critical applications such as economic dispatch (for profits) and contingency analysis (for reliability). A bad data detector (BDD) associated with the SE is often deployed for identifying bad data (e.g., sensor anomalies and FDI attacks) to ensure trustworthy results. However, it has been shown [4] that FDI attacks crafted using detailed knowledge of a power grid's topology and the reactance settings of its transmission lines can bypass the BDD and remain stealthy. Such an undetected attack can have severe consequences, e.g., trips of transmission line breakers or unsafe frequency excursions [5], [6].

To strengthen the BDD, it has been shown that if a carefully chosen subset of the sensors can be well protected (e.g., by tamper-proof and encryption-enabled PLCs), or if a key subset of the state variables can be independently and reliably verified by phasor measurement units (PMUs) deployed at strategically chosen locations, then a BDD-bypassing FDI attack becomes impossible [7], [8], [9]. However, a major revamp of the basic sensing infrastructure can be quite expensive (e.g., PMU has high cost [10]) or infeasible for the many existing legacy systems whose life cycles often last decades and which are not expected to retire for the foreseeable future. Alternatively, FDI attacks can be significantly mitigated by MTD that invalidates the knowledge attackers used for crafting their prior attacks, specifically by active perturbation of the grid's transmission line reactance settings in our application context [11], [12], [13]. This approach is practical because of current D-FACTS devices capable of active impedance injection [14]. Because of their low cost and ease and flexibility of deployment, they are being increasingly installed in existing alternating-current (ac) transmission networks to control power flows [15].

Prior work on MTD for FDI attacks against power grid SE has two important limitations, which are related. First, the MTD is implemented by selecting a random subset of transmission lines and introducing similarly random perturbations to their reactance settings [13]. There are no known conditions for the MTD perturbations to be truly effective. An important finding of this paper is that the randomly selected perturbations do not necessarily guarantee effective detection. Rather, a perturbation must satisfy certain design criteria that we will make clear (in Section V), or FDI attacks crafted using (outdated) system knowledge before the perturbation will remain stealthy after it. Second, without an adequate

This work was supported by the National Research Foundation (NRF), Prime Minister's Office, Singapore, under its National Cybersecurity R&D Programme (Award No. NRF2014NCR-NCR001-31) and administered by the National Cybersecurity R&D Directorate.

characterization of effective MTD, prior work has not been able to address explicitly the associated cost involved. Rather, it is assumed that the MTD can be always constrained to have negligible or some "low enough" operational cost [13], [11]. However, MTD designed with any absolute cost constraints will not be useful if the MTD does not perform. It is thus critical to understand the inherent cost-benefit tradeoff of the MTD to accordingly inform system operators (SOs) in their choice of security policies, which is a key objective of this paper.

To achieve our goal, we analyze the problem of selecting MTD reactance perturbations that jointly consider their effectiveness (i.e., capability of attack detection) and operational cost (i.e., economic inefficiency). As in prior work, we assume that the attacker has learned the system configuration initially and uses this knowledge to craft stealthy FDI attack vectors, but the attacker cannot track the reactance perturbations without significant delays. In this setting, large MTD perturbations will cause the actual system to deviate significantly from the attacker's prior knowledge, so that a large majority of the previously stealthy FDI attacks will now likely become detectable. Conversely, however, the large perturbations will also cause the power grid to operate significantly away from the optimal state, thereby incurring a significantly higher economic cost. On the other hand, smaller perturbations will be less expensive, but risk more undetected attacks. The general cost-benefit tradeoff is thus interesting.

In this paper, we address the cost-benefit tradeoff of the MTD by formulating its perturbation selection as a constrained optimization problem, namely minimization of the operational cost subject to a given effectiveness constraint. The operational cost is quantified as the increment due to the MTD over the cost achieved at optimal power flow (OPF) of the system without MTD. This cost is always non-negative. The effectiveness is quantified as the fraction of prior stealthy FDI attacks (i.e., those before the MTD perturbation) that will become detectable by the BDD after the perturbation. It is difficult to give an exact analysis of the effectiveness. We will instead employ a heuristic metric that effectively invalidates the attacker's knowledge required to bypass the BDD. Extensive simulation results show that the heuristic metric effectively approximates the true metric.

We use a direct-current (dc) power flow model to approximate power flows in an alternating-current (ac) grid. This approach is widely adopted and well justified in power system research (e.g., [4], [7], [13]). Under the dc model, the OPF cost corresponds mainly to the cost of generation dispatch. Moreover, the sensor measurements are linearly related to the system state through a *measurement matrix*, which in turn depends on the power grid topology and the reactance of the transmission lines. Naturally, perturbing a branch reactance will alter the measurement matrix correspondingly. A key observation in our analysis is that the MTD's effectiveness and operational cost are related to the separation between the column spaces of the measurement matrices before and after the MTD. While the effectiveness is enhanced by increasing the separation between the two column spaces, the operational cost increases. Therefore, different degrees of separation between the two spaces provide a spectrum of balance between the two metrics.

We note that, in light of our deliberate cost analysis of the MTD, the MTD can be viewed as a form of insurance against possible FDI attacks. Such insurance requires an ongoing payment of "premiums" irrespective of whether an attack occurs or not. However, in the event of an attack, which may be accumulatively extremely expensive if allowed to persist indefinitely because of lack of detection, the insurance can provide a much needed hedge against the damage. In actual deployments, whether to procure such insurance (i.e., turn on the MTD or not) is likely a matter of diverse factors such as institutional policies (including the institution's attitude towards risk taking), estimated vulnerability to attacks or likelihood of attacks, and the cost-benefit tradeoff specific to the power grid in question. This paper sheds light on tradeoffs in the key technical problem, which serves as an important reference basis for the other questions. Nevertheless, it does not attempt to answer all the questions, particularly policy questions, that are also interesting.

The main contributions of the paper are summarized as follows:

- We derive conditions for an MTD reactance perturbation to ensure that no FDI attacks crafted based on the outdated (pre-perturbation) system configuration will remain stealthy after the perturbation.
- When the reactance adjustment capability of D-FACTS is insufficient for achieving the above condition, we present heuristic design criteria for selecting MTD perturbations that can still highly likely achieve effective attack detection.
- We characterize the tradeoff between the MTD's effectiveness and its operational cost in a constrained optimization framework. Additionally, we present extensive simulation results using the realistic MATPOWER simulator for benchmark IEEE bus systems to verify and illustrate the tradeoff.

The remainder of this paper is organized as follows. Section II reviews related work. Section III introduces the preliminaries. Section IV explains the attacker and the defender model. Sections V and VI analyze the MTD's effectiveness and its cost-benefit tradeoff. Section VII presents simulation results. Section VIII concludes. The technical proofs can be found in Appendices A,B and C.

II. Prior Work

Recent work [4] analyzed the condition for bypassing the BDD of SE and proposed a technique to construct BDD-bypassing FDI attacks using complete knowledge of the power grid topology and the branch reactances. Subsequent research [16] showed that such attacks can be constructed using partial knowledge of the power grid topology. However, the knowledge of power grid topology is difficult to obtain in practice. Recent work [17], [18] showed that BDD-bypassing attacks can also be crafted using the eavesdropped measurement data only. The impact of such stealthy FDI attacks on system efficiency and safety were investigated. In particular, the economic impact of FDI attacks were studied in [19] and [20]. Reference [6] showed that the attacker can drive the power system frequency to unsafe levels by injecting a sequence of carefully-crafted FDI attacks.

To address BDD's vulnerability, defense mechanisms based on protecting a strategically-selected set of sensors and their data links were proposed [7], [8], [9]. The use of generalized likelihood ratio test was proposed to detect FDI attacks when the adversary has access to only a few meters in [21]. Reference [22] presented a sparse optimization based approach to separate nominal power grid states and anomalies.

The concept of MTD was originally proposed for enterprise networks based on changing the IT features of devices such as end hosts' IP addresses and port numbers, the routing paths between nodes, etc. [23], [24]. More recent work has proposed MTD in power systems by changing its physical characteristics [11], [12], [13]. In particular, *on-going* FDI attacks can be detected by introducing reactance perturbations that are known only to the defender (SO) [11], since the change in sensor measurements (after the perturbations) will be different from its predicted value based on the power flow model (due to the attack). It has also been shown that stealthy FDI attacks can be precluded by actively perturbing the branch reactances to invalidate the attacker's knowledge [13]. We similarly consider MTD for power systems in this paper. Compared with the prior work, ours is the first to jointly consider the MTD's effectiveness and its operational cost. We provide hitherto unavailable formal design criteria for selecting effective MTD reactance perturbations, and expose important tradeoffs between the effectiveness and operational cost.

III. Preliminaries

Power Grid Model

We consider a power network that is characterized by a set $\mathcal{N} = \{1, \dots, N\}$ of buses, $\mathcal{L} = \{1, \dots, L\}$ of transmission lines (an example of the 4 bus power system is shown in Figure 3). The line $l \in \mathcal{L}$ that connects bus i and bus j is denoted by $l = \{i, j\}$. The time of operation is denoted by $t \in \mathbb{R}$.

At bus i, we denote the power generation and load at time t by $G_{i,t}$ and $L_{i,t}$ respectively and the reactance of link l by $x_{l,t}$. We adopt the dc power flow model [25], under which the power flow on line l at time t denoted by $F_{l,t}$, is given by

$$F_{l,t} = \frac{1}{x_{l,t}}(\theta_{i,t} - \theta_{j,t}),$$

where $\theta_{i,t}$ and $\theta_{j,t}$ are the voltage phase angles at buses $i, j \in \mathcal{N}$ respectively at time t. For safe operation, the branch flows must be maintained within the power flow limits $F_k^{\max}$ at all time, i.e.,

$$-F_k^{\max} \leq F_{k,t} \leq F_k^{\max}, \ \forall t.$$

The relationship between branch power flows and the voltage phase angles can be compactly represented as $\mathbf{f}_t = \mathbf{D}_t \mathbf{A}^T \boldsymbol{\theta}_t$, where the matrix $\mathbf{A} \in \mathbb{R}^{N \times L}$ is the branch-bus incidence matrix given by

$$\mathbf{A}_{i,j} = \begin{cases} 1, & \text{if link } j \text{ starts at bus } i, \\ -1, & \text{if link } j \text{ ends at bus } i \ , \\ 0 & \text{otherwise,} \end{cases}$$

and $\mathbf{D}_t \in \mathbb{R}^{L \times L}$ is a diagonal matrix of the reciprocal of link reactances, i.e.,

$$\mathbf{D}_t = \text{diag}\left(\left[\frac{1}{x_{1,t}}, \frac{1}{x_{2,t}}, \dots, \frac{1}{x_{L,t}}\right]\right),$$

and $\mathbf{f}_t = [F_{1,t}, \dots, F_{L,t}]^T$ (similarly $\mathbf{g}_t, \mathbf{l}_t, \boldsymbol{\theta}_t$ denote the vector forms of the corresponding quantities).

We assume that a subset of the links $\mathcal{L}_D \subseteq \mathcal{L}$ are equipped with D-FACTS devices, and the reactances of these links can be changed within the range $[\mathbf{x}^{\min}, \mathbf{x}^{\max}]$, where $\mathbf{x}^{\min}, \mathbf{x}^{\max}$ are the reactance limits achievable by the D-FACTS devices. Naturally, $x_l^{\min} = x_l^{\max} = x_{l,t}$ if $l \notin \mathcal{L}_D$. Denote the vector of branch reactances by $\mathbf{x}_t$.

State Estimation & Bad Data Detection Technique

SE is a technique of estimating the system state from its noisy sensor measurements [25]. Under the dc power flow model, the state at time t corresponds to the nodal voltage phase angles $\boldsymbol{\theta}_t$, which are monitored by a set of M measurements $\mathbf{z}_t \in \mathbb{R}^M$. The measurements correspond to the nodal power injections, and the forward and reverse branch power flows, i.e. $\mathbf{z}_t = [\tilde{\mathbf{p}}_t, \tilde{\mathbf{f}}_t, -\tilde{\mathbf{f}}_t]^T$. We note that the measurements may be different from the actual values of $\mathbf{p}_t$ and $\mathbf{f}_t$ due to sensor measurement noises or cyber-attacks. The measurement vector and the state are related as

$$\mathbf{z}_t = \mathbf{H}_t \boldsymbol{\theta}_t + \mathbf{n}_t,$$

where $\mathbf{n}_t$ is the measurement noise, which is assumed to have Gaussian distribution. $\mathbf{H}_t \in \mathbb{R}^{M \times N}$ is the measurement matrix given by

$$\mathbf{H} = \begin{bmatrix} \mathbf{D}_t \mathbf{A}^T \\ -\mathbf{D}_t \mathbf{A}^T \\ \mathbf{A} \mathbf{D}_t \mathbf{A}^T \end{bmatrix}.$$

The estimate of the system state, $\widehat{\boldsymbol{\theta}}_t$, is computed using a maximum likelihood (ML) estimation technique, given by [25],

$$\widehat{\boldsymbol{\theta}}_t = (\mathbf{H}_t^T \mathbf{W} \mathbf{H}_t)^{-1} \mathbf{H}_t^T \mathbf{W} \mathbf{z}_t,$$

where $\mathbf{W}$ is a diagonal weighting matrix whose elements are reciprocals of the variances of the sensor measurement noise.

A BDD is used to detect faulty sensor measurements. It compares the residual defined by $r_t = ||\mathbf{z}_t - \mathbf{H}_t \widehat{\boldsymbol{\theta}}_t||$ against a pre-defined threshold τ and raises an alarm if $r_t \geq \tau$. The detection threshold τ is determined by the SO to ensure a certain false positive (FP) rate α, where $\alpha > 0$ (usually a small value close to zero).

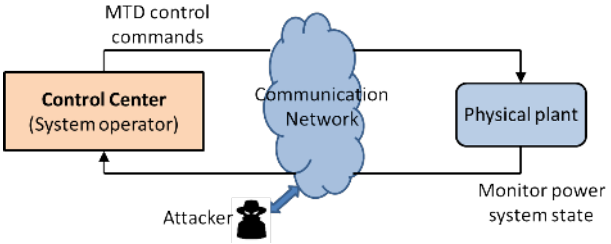


Fig. 1: System block diagram.

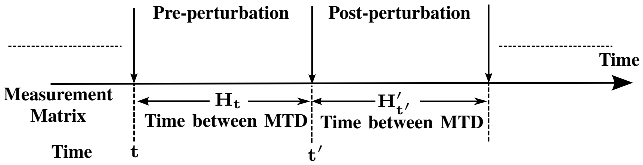


Fig. 2: MTD timeline. The vertical arrows indicate the times at which the system is perturbed.

Undetectable FDI Attacks

We consider FDI attacks against the SE, in which the attacker injects an attack vector $\mathbf{a}_t \in \mathbb{R}^M$ into the sensor measurements, i.e., $\mathbf{z}_t^a = \mathbf{z}_t + \mathbf{a}_t$, where $\mathbf{z}_t^a$ is the measurement vector under an attack. In general, the BDD can detect arbitrary FDI attack vectors. However, it is demonstrated [4] that the BDD's detection probability for attacks of the form $\mathbf{a}_t = \mathbf{H}_t\mathbf{c}$, where $\mathbf{c} \in \mathbb{R}^N$, is no greater than the FP rate α. Such attacks are referred to as *undetectable attacks*.

Optimal Power Flow Problem

OPF is an optimization framework to adjust the power flows in the network (by setting the generator dispatch and the branch reactances) with the objective of minimizing the generation cost for a given load vector $\mathbf{l}_t \in \mathbb{R}^N$, stated as follows[1]:

$$C_{\text{OPF},t} = \min_{\mathbf{g}_t,\mathbf{x}_t} \quad \sum_{i\in\mathcal{N}} C_i(G_{i,t}) \tag{1a}$$

$$s.t. \quad \mathbf{g}_t - \mathbf{l}_t = \mathbf{B}_t\boldsymbol{\theta}_t, \tag{1b}$$

$$-\mathbf{f}^{\max} \leq \mathbf{f}_t \leq \mathbf{f}^{\max}, \tag{1c}$$

$$\mathbf{g}^{\min} \leq \mathbf{g}_t \leq \mathbf{g}^{\max}, \tag{1d}$$

$$\mathbf{x}^{\min} \leq \mathbf{x}_t \leq \mathbf{x}^{\max}, \tag{1e}$$

where $C_i(G_{i,t})$ is the cost of generating $G_{i,t}$ units of power at node $i \in \mathcal{N}$, the matrix $\mathbf{B}_t = \mathbf{A}\mathbf{D}_t\mathbf{A}^T$. In (1), the first constraint (1b) represents the nodal power balance constraint, i.e., the power injected into a node must be equal to the power flowing out of the node. Constraints (1c)-(1e) correspond to the branch power flows, generator limits, and D-FACTS limits, respectively. We denote $\mathbf{g}_t^*, \mathbf{x}_t^* = \arg\max_{\mathbf{g}_t,\mathbf{x}_t}$ OPF. We note that the OPF cost depends on the branch reactances through the matrix $\mathbf{B}_t$ (in addition to the loads).

IV. Moving-Target Defense in Power Grids

A. Attacker and the Defender Model

A block diagram of the system under study is shown in Fig. 1. We consider a strong attacker who has access to the measurement data communicated between the field devices and the control center. Such access could be obtained by exploiting vulnerabilities in power grid communication systems. For example, in modern-day power grids, the field devices (such as remote terminal units) are often IP-accessible [26]. We also assume that the attacker can learn the system's measurement matrix (using the eavesdropped measurements) and craft undetectable FDI attacks accordingly (e.g., see [17], [18]).

Under MTD, the defender (e.g., the SO) tries to thwart the FDI attacks by actively perturbing the transmission line reactances to invalidate the attacker's prior knowledge. We assume that at the time of introducing MTD perturbations, there are no on-going FDI attacks. Note that the power system under consideration is naturally dynamic (even without MTD) since the branch reactances are optimized periodically to reflect temporal changes in the system load (refer to the OPF problem in (1)). However, these natural changes are usually insufficient for effectively negating the attacker's knowledge. Thus, the defender deliberately introduces an additional reactance perturbation to ensure the MTD's detection capability.

The defender implements the MTD reactance perturbations by sending MTD control commands to the remote D-FACTS devices in the grid. Unlike the sensor measurements that support the grid's normal operation (e.g., extensive SCADA measurements collected every few seconds), these commands are much less frequent (e.g., hourly, see the discussion below), have much more restricted scope (i.e., between the control center and the set of D-FACTS devices only), and do not have stringent real-time constraints. Hence, we assume that it is feasible to encrypt the MTD commands to ensure their confidentiality.

We note that although the attacker cannot read the MTD commands directly due to their encryption, in principle he may still infer the MTD perturbations by monitoring their effects on the eavesdropped sensor measurements and estimating the new measurement matrix accordingly. Thus, the secrecy of the MTD generally decays over time. In practice, however, the learning will be time consuming since the attacker must collect an informative sequence of the measurements over a significant duration of time. In this paper, we assume that the time interval between the MTD perturbations is sufficiently small, so that during it the attacker's gain in knowledge (of the measurement matrix) is negligible.

A guiding principle to estimate the perturbation time interval can be obtained from [17], in which it is shown that FDI attacks against an IEEE 14-bus system require about $500-1000$ measurements of the system to successfully bypass the BDD, even if these measurements are assumed to have

[1] In the absence of D-FACTS devices installed within the grid, OPF optimizes over the generator dispatch values only (which is the version of OPF traditionally used [25]).

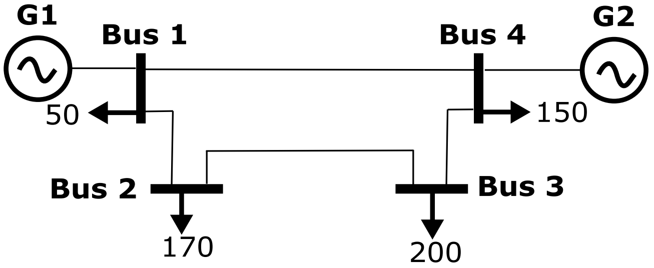


Fig. 3: 4 bus system under consideration. The loads are indicated in MWs.

maximum information diversity in that they are i.i.d. Hence, if we assume optimistically for the attacker that SCADA measurements need to be only $5-10$ seconds apart to achieve the information diversity, their result suggests that the time required by the attacker to learn the system sufficiently well for stealthy attacks is on the order of a few hours. Accordingly, hourly MTD perturbations might be realistic for practical systems. Further, we note that utilities typically solve the OPF more frequently, i.e., every $5-10$ minutes (whereas we only need to update the MTD every hour or so). Thus, between the MTD updates, the OPF will be solved as in (1).

The MTD timeline is illustrated in Fig. 2. We consider two representative time instants t and t' at which the reactances are perturbed for MTD. We denote the branch reactances and the measurement matrix after applying the MTD perturbations by $\mathbf{x}'_{t'} = [x'_{1,t'}, \ldots, x'_{L,t'}]^T$ and $\mathbf{H}'_{t'}$ respectively, and the reactance perturbation vector by $\Delta\mathbf{x}_{t,t'} = \mathbf{x}_t - \mathbf{x}'_{t'}$. We note that in the absence of MTD, the branch reactances and the measurement matrix would be set to $\mathbf{x}_{t'}$ and $\mathbf{H}_{t'}$ by solving (1) at time t'.

In the rest of the paper, we address the question of how to select MTD perturbations that are effective in detecting FDI attacks crafted based on the outdated (i.e., pre-pertubation) knowledge, and examine their cost-benefit tradeoff. We use $a'_{t'}$ to denote the value of a power system parameter a_t after the MTD. E.g., $\theta'_{t'}$ denotes the nodal voltage phase angles after the MTD. To motivate our inquisition, we now illustrate an example to show that certain randomly selected MTD perturbations will remain vulnerable to FDI attacks crafted with the attacker's pre-pertubation knowledge of the system.

B. A Motivating Example

We consider the 4-bus example shown in Fig. 3 [27]. For simplicity, we assume that the system load is fixed (indicated in Fig. 3) and does not change with time. Furthermore, the pre-perturbation system state and the reactance settings $\mathbf{x}_t$ (and $\mathbf{H}_t$) are adjusted by solving (1). The resulting branch flows, generation values and OPF cost are listed in Table II. The attacker is assumed to have learned the pre-perturbation matrix $\mathbf{H}_t$.

	$r'^{(1)}$	$r'^{(2)}$	$r'^{(3)}$	$r'^{(4)}$
Attack 1	2.82	2.87	0	0
Attack 2	0	0	2.87	2.82

TABLE I: BDD residual values.

Line Flow (MWs)				Gen. (MWs)		Cost($)
Line 1	Line 2	Line 3	Line 4	Gen 1	Gen 2	1.15×10^4
126.56	173.44	-43.44	-26.56	350	150	

TABLE II: Pre-perturbation power flows, generator dispatch and OPF cost for 4-bus system.

MTD	Gen. (MWs)		OPF Cost ($)
$\Delta\mathbf{x}_1$	337.37	162.62	1.1626×10^4
$\Delta\mathbf{x}_2$	340.51	159.48	1.595×10^4
$\Delta\mathbf{x}_3$	348.62	151.37	1.1514×10^4
$\Delta\mathbf{x}_4$	345.95	154.02	1.154×10^4

TABLE III: Post-perturbation generator dispatch and OPF cost.

To implement the MTD, we consider four reactance perturbation vectors respectively given by $\Delta\mathbf{x}^{(1)}_{t,t'} = \eta[x_1, 0, 0, 0]^T$, $\Delta\mathbf{x}^{(2)}_{t,t'} = \eta[0, x_2, 0, 0]^T, \Delta\mathbf{x}^{(3)}_{t,t'} = \eta[0, 0, x_3, 0]^T$, $\Delta\mathbf{x}^{(4)}_{t,t'} = \eta[0, 0, 0, x_4]^T$, where η is the percentage change in the reactance relative to its initial value. We assess each of the four MTD perturbations in terms of (i) attack detection and (ii) operational cost.

For attack detection, we inject an attack of the form $\mathbf{a} = \mathbf{H}_t\mathbf{c}$ into the modified power network (after the MTD), and examine its BDD residual. For illustration, we consider two attacks – attack 1 in which $\mathbf{c} = [0, 1, 1, 1]^T$ and attack 2 in which $\mathbf{c} = [0, 0, 0, 1]^T$ – and set $\eta = 0.2$. For simplicity, we ignore measurement noises. The BDD residuals under the four MTD perturbations are listed in Table I. Note that in the absence of measurement noise, a non-zero value of the residual indicates the presence of attack. We observe that for each of the four perturbations, there exist attack vectors of the form $\mathbf{a} = \mathbf{H}_t\mathbf{c}$, which continue to bypass the BDD for the perturbed power network.

We also enlist the post-pertubration OPF cost in Table III. We observe that the OPF cost increases in each of the four cases, compared to its pre-perturbation cost, and the perturbation $\Delta\mathbf{x}_3$ incurs the least cost.

C. MTD Perturbation Selection Challenges

Based on the above illustrating example, we make the following conclusions. First, it is evident that a subset of the attacks of the form $\mathbf{a} = \mathbf{H}_t\mathbf{c}$ continue to bypass the BDD after the MTD. Since the defender does not have prior knowledge of the actual attack vector (note that $\mathbf{c}$ is chosen by the attacker), he cannot make an informed choice of which perturbation to adopt. Without such knowledge, the defender must select the MTD that is capable of detecting a largest subset of the possible attacks. The second design criterion is the MTD's operational cost, i.e., other things being equal, the defender prefers a least-cost MTD. In the following sections, we characterize formally the MTD's effectiveness and its operational cost, and present a framework for choosing appropriate MTD perturbations that balance between the two concerns.

V. MTD's Effectiveness of Attack Detection

In this section, we address the problem of selecting effective MTD reactance perturbations from an attack detection point of view. The goal is to select reactance perturbations within the physical constraints of the D-FACTS devices to effectively invalidate the attacker's knowledge for bypassing the BDD. The section is divided into two parts. In the first part, we devise a metric to quantify the effectiveness of the MTD. In the second part, we derive the conditions and propose design criteria for MTD perturbations to preclude stealthy FDI attacks in practice.

Henceforth, we use the notation "MTD $\mathbf{H}'_{t'}$" to refer to a reactance perturbation that changes the measurement matrix from $\mathbf{H}_t$ to $\mathbf{H}'_{t'}$. We let $\mathcal{A}$ denote the set of all attack vectors of the form $\mathbf{a} = \mathbf{H}_t\mathbf{c}$, i.e.,

$$\mathcal{A} = \{\mathbf{a} : \mathbf{a} = \mathbf{H}_t\mathbf{c}, ||\mathbf{a}|| \leq a_{\max}, \mathbf{c} \in \mathbb{R}^N\}.$$

For an attack vector $\mathbf{a}$, we let $P'_D(\mathbf{a})$ denote its detection probability under MTD $\mathbf{H}'_{t'}$, where $P'_D(\mathbf{a}) = \mathbb{P}(r' \geq \tau)$. We denote by $\mathcal{A}'(\delta) \subseteq \mathcal{A}$ the subset of attacks in $\mathcal{A}$ whose detection probability under MTD $\mathbf{H}'_{t'}$ is greater than a given $\delta \in [0, 1]$, i.e.,

$$\mathcal{A}'(\delta) = \{\mathbf{a} : \mathbf{a} = \mathbf{H}_t\mathbf{c}, ||\mathbf{a}|| \leq a_{\max}, P'_D(\mathbf{a}) > \delta, \mathbf{c} \in \mathbb{R}^N\}.$$

A. Metric to Quantify MTD's Effectiveness

First, we devise a metric to quantify the MTD's effectiveness. Intuitively, an MTD perturbation "A" is more effective than a perturbation "B" if it can detect more FDI attacks in the set $\mathcal{A}$ with high probability. However, $\mathcal{A}$, a subset in the n-dimensional space ($\mathbb{R}^n$), has infinitely many attack vectors. For these sets, the *Lebesgue measure* generalizes the notion of length (one-dimensional), area (two-dimensional), or volume (three-dimensional) to n-dimensions [28]. The effectiveness of an MTD $\mathbf{H}'_{t'}$ for a given $\delta \in [0, 1]$, which we denote by $\eta'(\delta)$, can be quantified as

$$\eta'(\delta) = \frac{\lambda(\mathcal{A}'(\delta))}{\lambda(\mathcal{A})}, \tag{2}$$

where $\lambda(\mathcal{A}'(\delta))$ and $\lambda(\mathcal{A})$ denote the Lebesgue measures of the respective sets. Intuitively, $\eta'(\delta)$ represents the ratio of the number of attack vectors of the form $\mathbf{a} = \mathbf{H}_t\mathbf{c}$ whose detection probability under MTD $\mathbf{H}'_{t'}$ is greater than δ to the total number of attacks in the set $\mathcal{A}$. Since $\mathcal{A}'(\delta) \subseteq \mathcal{A}$, $0 \leq \eta'(\delta) \leq 1$.

Of particular interest are the sets $\mathcal{A}'(\alpha)$ and $\mathcal{A} \setminus \mathcal{A}'(\alpha)$, and the latter is the set of undetectable attacks under MTD $\mathbf{H}'_{t'}$ (refer to Section III for the definition of undetectable attacks). An ideal MTD is one that admits no undetectable attacks of the form $\mathbf{a} = \mathbf{H}_t\mathbf{c}$, i.e., $\mathcal{A}'(\alpha) = \mathcal{A}$ and $\eta'(\alpha) = 1$. In the following subsection, we derive conditions on the MTD $\mathbf{H}'_{t'}$ that can ensure the property.

B. MTD Admitting No Undetectable Attacks

We start by characterizing the condition for an attack $\mathbf{a} = \mathbf{H}_t\mathbf{c}$ to remain undetectable under MTD $\mathbf{H}'_{t'}$.

Proposition 1. *An attack of the form $\mathbf{a} = \mathbf{H}_t\mathbf{c}$ is* undetectable *under MTD perturbation $\mathbf{H}'_{t'}$ if it satisfies the condition rank$(\mathbf{H}'_{t'})$ = rank$([\mathbf{H}'_{t'}\ \ \mathbf{H}_t\mathbf{c}])$, where $[\mathbf{H}'_{t'}\ \ \mathbf{H}_t\mathbf{c}]$ is the augmented matrix.*

The proof of this proposition is presented in Appendix A. Intuitively, the proposition implies that an attack vector of the form $\mathbf{a} = \mathbf{H}_t\mathbf{c}$ is undetectable under MTD $\mathbf{H}'_{t'}$ if it lies in the column spaces of both $\mathbf{H}_t$ and $\mathbf{H}'_{t'}$, since rank$(\mathbf{H}'_{t'})$ = rank$([\mathbf{H}'_{t'}\ \ \mathbf{H}_t\mathbf{c}])$ for the attack vector $\mathbf{a} = \mathbf{H}_t\mathbf{c} \in Col(\mathbf{H}'_{t'})$.

The result allows us to give conditions for the MTD $\mathbf{H}'_{t'}$ to ensure no undetectable attacks of the form $\mathbf{a} = \mathbf{H}_t\mathbf{c}$. In particular, to achieve the aforementioned property, MTD $\mathbf{H}'_{t'}$ must be selected such that no attack vector $\mathbf{a}$ in the column space of $\mathbf{H}_t$ lies in the column space of $\mathbf{H}'_{t'}$. The following theorem states the condition.

Theorem 1. *An MTD $\mathbf{H}'_{t'}$ has no undetectable attacks of the form $\mathbf{a} = \mathbf{H}_t\mathbf{c}$ if $Col(\mathbf{H}'_{t'})$ is the orthogonal complement of $Col(\mathbf{H}_t)$. Furthermore, for a given attack vector $\mathbf{a}$, such an MTD achieves the maximum value of $P'_D(\mathbf{a})$ among all the possible MTD perturbations.*

The proof is presented in Appendix B. The first statement of this theorem implies that for the MTD $\mathbf{H}'_{t'}$ satisfying the orthogonality condition, there are no attacks of the form $\mathbf{a} = \mathbf{H}_t\mathbf{c}$ for which $P'_D(\mathbf{a})$ is as low as the FP rate α (in general, α is chosen by the SO to be a small value). However, this result does not automatically imply that the attacks will also be detected with high probability, which is the desired outcome. But the second statement of Theorem 1 shows that this is indeed the case, since such an MTD also maximizes $P'_D(\mathbf{a})$ among all possible MTD perturbations.

From Theorem 1, we conclude that purely from an attack detection point of view, an MTD perturbation should be selected to achieve the stated orthogonality condition. However, this may not always be feasible due to practical limitations, e.g., the D-FACTS devices may only allow the reactances to be perturbed within a certain range. In these cases, we require an additional design criterion to select the MTD perturbations, which is the subject of the following subsection.

C. Heuristic Design Criteria for Selecting MTD Perturbation

Intuitively, if the reactance adjustment capability of D-FACTS is insufficient to meet the orthogonality condition of Theorem 1, the MTD perturbation should be selected to make $Col(\mathbf{H}'_{t'})$ as orthogonal to $Col(\mathbf{H}_t)$ as possible within the constraints of the D-FACTS device. To formalize this notion, we introduce the concept of *principal angle* between subspaces, defined as follows:

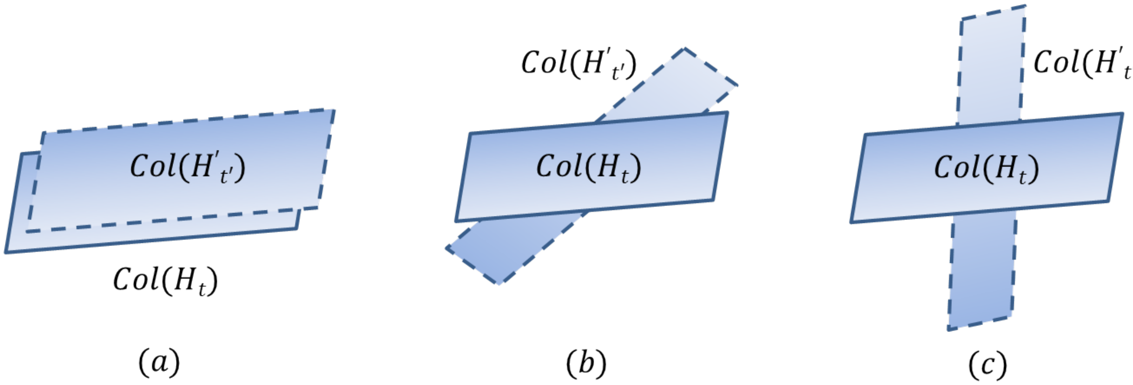


Fig. 4: Orientation of $Col(H'_{t'})$ with respect to $Col(H_t)$, (a) $\gamma(\mathbf{H}_t, \mathbf{H}'_{t'}) = 0$ (perfectly aligned column spaces), (b) $0 \leq \gamma(\mathbf{H}_t, \mathbf{H}'_{t'}) \leq \pi/2$, and (c) $\gamma(\mathbf{H}_t, \mathbf{H}'_{t'}) = \pi/2$ (orthogonal column spaces).

Definition V.1 ([29])**.** *The smallest principal angle (SPA)* $0 \leq \theta \leq \pi/2$ *between the subspaces* $\mathcal{F}, \mathcal{G} \subseteq \mathbb{C}^N$ *is defined as*

$$\cos(\theta) = \max_{\substack{\mathbf{u} \in \mathcal{F}, \mathbf{u} \in \mathcal{G} \\ ||\mathbf{u}||=1, ||\mathbf{v}||=1}} |\mathbf{u}^H \mathbf{v}|.$$

The SPA generalizes the concept of angle between a pair of vectors to a pair of n-dimensional subspaces. Let $\gamma(\mathbf{H}_t, \mathbf{H}'_{t'})$ denote the SPA between $Col(\mathbf{H}_t)$ and $Col(\mathbf{H}'_{t'})$. We conjecture that MTD perturbations with a higher value of $\gamma(\mathbf{H}_t, \mathbf{H}'_{t'})$ are more effective in terms of attack detection. Thus, SPA can be utilized as a design criterion for selecting good MTD perturbations.

The conjecture is based upon the following observations. (i) In Appendix C, we present arguments which suggest that the attack detection probability $P'_D(\mathbf{a})$ increases as we select MTD perturbations with higher $\gamma(\mathbf{H}_t, \mathbf{H}'_{t'})$. (ii) In the following, we give some observations to suggest that the measure of the set of undetectable attacks decreases by selecting MTD perturbations with higher $\gamma(\mathbf{H}_t, \mathbf{H}'_{t'})$.

We examine MTD perturbations in two extreme cases as illustrated in Fig. 4. First, consider MTD $\mathbf{H}'_{t'} = (1+\eta)\mathbf{H}_t$, for which it can be verified that $\gamma(\mathbf{H}_t, \mathbf{H}'_{t'}) = 0$. For such an MTD, the column spaces of the matrices $\mathbf{H}_t$ and $\mathbf{H}'_{t'}$ are perfectly aligned. Hence all attacks of the form $\mathbf{a} = \mathbf{H}_t\mathbf{c}$ remain undetectable after the MTD (i.e., $\mathcal{A}'(\alpha) = \emptyset$ and $\lambda(\mathcal{A}'(\alpha)) = 0$). Thus, an MTD perturbation with $\gamma(\mathbf{H}_t, \mathbf{H}'_{t'}) = 0$ is the least effective in detecting FDI attacks. Second, for MTD $\mathbf{H}'_{t'}$ satisfying the orthogonality condition of Theorem 1, it can be verified that $\gamma(\mathbf{H}_t, \mathbf{H}'_{t'}) = \pi/2$. As shown in the previous subsection, in this case, $\mathcal{A}'(\alpha) = \mathcal{A}$ and there are no undetectable attacks of the form $\mathbf{a} = \mathbf{H}\mathbf{c}$.

These arguments suggest that MTD perturbations for which $\gamma(\mathbf{H}_t, \mathbf{H}'_{t'})$ is closer to $\pi/2$ are more effective in detecting FDI attacks, a trend that is also confirmed by our simulation results using the IEEE 14-bus system (see Section VII). A natural follow up question is how to select the reactance perturbation vector $\Delta\mathbf{x}_{t,t'}$ to achieve the aforementioned design criteria. In the next section, we present an optimization framework to numerically compute $\Delta\mathbf{x}_{t,t'}$ while also considering the MTD's operational cost.

VI. MTD's Cost-Benefit Tradeoff

Thus far, we have investigated the MTD from an attack detection point of view only. In this section, we formally define the operational cost of MTD in an optimization framework.

MTD Operational Cost

We quantify MTD's cost in terms of the increase in OPF cost due to the MTD relative to its value without MTD, i.e.,

$$C_{\text{MTD},t'} = \frac{C'_{\text{OPF},t'} - C_{\text{OPF},t'}}{C_{\text{OPF},t'}}, \tag{3}$$

where $C_{\text{OPF},t'}$ is the OPF cost of the system corresponding to the measurement matrix $\mathbf{H}_{t'}$ computed using (1) (at time t'), and $C'_{\text{OPF},t'}$ is the OPF cost of the system with MTD (corresponding to the measurement matrix $\mathbf{H}'_{t'}$). Note that $C_{\text{MTD},t'}$ is always non-negative since the additional perturbation due to MTD will increase the OPF cost.

From (3), we note that $C_{\text{MTD},t'}$ depends on the separation between the column spaces of $\mathbf{H}_{t'}$ and $\mathbf{H}'_{t'}$. In particular, if the two matrices are identical, then $C_{\text{MTD},t'}$ is zero. As the separation between the column spaces of the two matrices $\gamma(\mathbf{H}_{t'}, \mathbf{H}'_{t'})$ is increased, the power flows within the two systems and the corresponding generation dispatch will be different (due to the reactance perturbation). Consequently, the OPF cost in the system with MTD perturbation will increase.

Our observation is that $\gamma(\mathbf{H}_t, \mathbf{H}'_{t'})$ closely approximates $\gamma(\mathbf{H}_{t'}, \mathbf{H}'_{t'})$. Hence, MTD's operational cost increases as we choose perturbations with higher $\gamma(\mathbf{H}_t, \mathbf{H}'_{t'})$. The approximation can be explained as follows. Recall that $\mathbf{H}_t$ and $\mathbf{H}_{t'}$ differ only due to temporal variations in the system load. Since the power system load is temporally correlated, the matrices $\mathbf{H}_t$ and $\mathbf{H}_{t'}$ will not differ significantly and their column spaces are nearly aligned. Thus, $\gamma(\mathbf{H}_t, \mathbf{H}'_{t'})$ can be used as an approximate measure of the SPA between the column spaces of $\mathbf{H}_{t'}$ and $\mathbf{H}'_{t'}$. Extensive simulation results driven by real-world data load traces presented in Section VII confirm the validity of this approximation.

MTD Tradeoff

Following the above arguments, we note that the defender faces conflicting objectives. On the one hand, for the MTD to be effective from an attack detection point of view, the column spaces of the matrices $\mathbf{H}_t$ and $\mathbf{H}'_{t'}$ should be as orthogonal as possible. On the other hand, the MTD's operational cost increases with $\gamma(\mathbf{H}_t, \mathbf{H}'_{t'})$. Thus, there exists a trade-off between the MTD's effectiveness and its operational cost. To balance the two aspects, we formulate the MTD reactance selection problem as a constrained optimization problem with the objective of minimizing the operational cost subject to a constraint on the MTD's effectiveness. The problem is stated as:

$$\begin{aligned}
C'_{\text{OPF},t'} = \quad & \min_{\mathbf{g}'_{t'}, \mathbf{x}'_{t'}} \quad \sum_{i \in \mathcal{N}} C_i(G'_{i,t'}) && \text{(4a)}\\
s.t. \quad & \gamma(\mathbf{H}_t, \mathbf{H}'_{t'}) \geq \gamma_{\text{th}}, && \text{(4b)}\\
& \mathbf{g}'_{t'} - \mathbf{l}_{t'} = \mathbf{B}'_{t'} \boldsymbol{\theta}'_{t'}, && \text{(4c)}\\
& -\mathbf{f}^{\max} \leq \mathbf{f}'_{t'} \leq \mathbf{f}^{\max}, && \text{(4d)}\\
& \mathbf{g}^{\min} \leq \mathbf{g}'_{t'} \leq \mathbf{g}^{\max}, && \text{(4e)}\\
& \mathbf{x}^{\min} \leq \mathbf{x}'_{t'} \leq \mathbf{x}^{\max}. && \text{(4f)}
\end{aligned}$$

In (4), the SPA between the column spaces of $\mathbf{H}_t$ and $\mathbf{H}'_{t'}$ is used as a heuristic metric to approximate the effectiveness of the attack detection $\eta'(\delta)$ (based on the conjecture stated in Section V-C). In (4b), we impose a constraint on the SPA, where $\gamma_{\text{th}} \in [0, \pi/2]$ is a threshold that must be tuned numerically (see Section VII for more details). Simulation results show that different values of the threshold γ_{th} provide a spectrum of trade-offs between the MTD's effectiveness and its operational cost. We propose to solve (4) numerically using existing constrained non-linear optimization solvers (e.g., the *fmincon* function of MATLAB).

Note that the attacker does not have sufficient information to solve (4) and thus cannot anticipate the MTD perturbations. In particular, at time t', the attacker does not know $\mathbf{H}_t$, since there is not sufficient time to learn it given the frequency of perturbations (see the discussion in Sec. IV-A). Hence, the secrecy of the MTD is satisfied.

VII. Simulation Results

In this section, we present simulation results to evaluate the MTD's effectiveness and its operational cost.

A. Simulation Settings & Methodology

The simulations are carried out in MATLAB. All the constrained optimization problems involved in the simulations are solved using the *fmincon* function of MATLAB with the *MultiStart* algorithm.

We perform simulations using the IEEE 14-bus system. The bus topology is shown in Fig. 5. We obtain its configuration data from the MATPOWER package [27]. As shown in Fig. 5, the generators are installed at buses $1, 2, 3, 6, 8$ and their parameters are listed in Table IV. We use the linear generation cost model given by $C_i(G_{i,t}) = c_i G_{i,t}$. We assume that D-FACTS devices are installed on 6 branches indexed by $\mathcal{L}_D = \{1, 5, 9, 11, 17, 19\}$. The D-FACTS limits are set to $\mathbf{x}_{\min} = (1 - \eta_{\max})\mathbf{x}$ and $\mathbf{x}_{\min} = (1 + \eta_{\max})\mathbf{x}$, where $\mathbf{x}$ is the default values (obtained from the IEEE 14-bus case file) and $\eta_{\max}$ is set to 0.5. Further, the branch flow limits are chosen to be 160 MWs for link 1, and 60 MWs for all other links of the power system. The rest of the settings are obtained from the MATPOWER configuration case file.

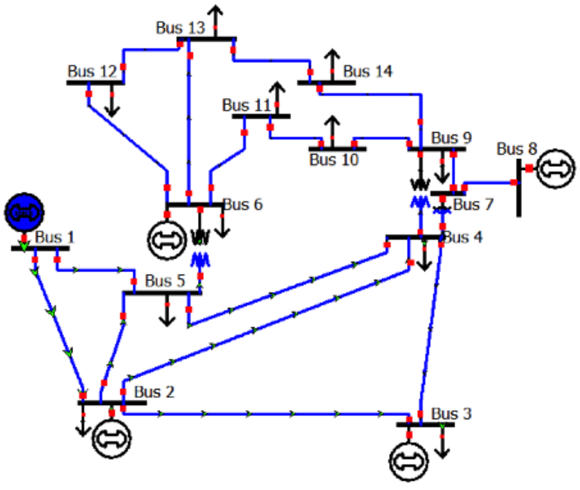


Fig. 5: IEEE 14-bus system. (Figure source: [30])

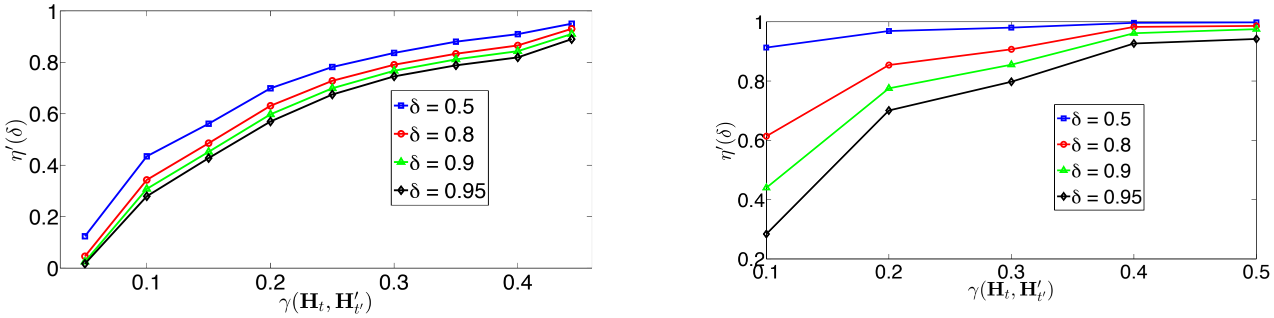
TABLE IV: Generator parameters.

Gen. bus	1	2	3	6	8
$P_{\max}$ (MWs)	300	50	30	50	20
c_i (\$/MWh)	20	30	40	50	35

B. Simulation Results with Static Load

In the first set of simulations, we assume that the system load is static (we use default values from the IEEE 14-bus MATPOWER case file). The pre-perturbation reactances $\mathbf{x}_t$ (and $\mathbf{H}_t$) are adjusted by solving (1). The defender designs MTD $\mathbf{H}'_{t'}$ assuming that the attacker has acquired the knowledge of $\mathbf{H}_t$, and that he injects attacks of the form $\mathbf{a} = \mathbf{H}_t\mathbf{c}$.

Effectiveness of Attack Detection: First, we examine the MTD's effectiveness ($\eta'(\delta)$) for different values of $\gamma(\mathbf{H}, \mathbf{H}')$. We choose $\gamma(\mathbf{H}_t, \mathbf{H}'_{t'}) \in [0, 0.45]$ radians in steps of 0.05 radians. For each value of $\gamma(\mathbf{H}_t, \mathbf{H}'_{t'})$, we solve the optimization problem (4) by setting γ_{th} to the corresponding value, and evaluate $\eta'(\delta)$ using Monte Carlo simulations as follows. We consider 1000 attack vectors of the form $\mathbf{a} = \mathbf{H}_t\mathbf{c}$, where the vector $\mathbf{c}$ is chosen as a random vector drawn from the Gaussian distribution, and scale its magnitude such that $||\mathbf{a}||_1/||\mathbf{z}||_1 \approx 0.08$ (the scaling adjusts the magnitude of attack injections to be relatively small in comparison to the actual measurements). We then evaluate $P'_D(\mathbf{a})$ for each of the attack vectors (the details will be presented shortly), and count the fraction of attack vectors for which $P'_D(\mathbf{a}) \geq \delta$, for a given value of $\delta \in [0, 1]$. For each attack vector, the detection probability $P'_D(\mathbf{a})$ is computed by generating 1000



(a) IEEE 14-Bus System

(b) IEEE 30-Bus System

Fig. 6: MTD effectiveness for different values of $\gamma(\mathbf{H}_t, \mathbf{H}'_{t'})$ (radians). FP rate is set to 5×10^{-4}.

instantiations of measurement noise (according to the Gaussian distribution), and counting the number of times the BDD alarm is triggered. The BDD threshold is adjusted such that the FP rate is set to 5×10^{-4}. We note that MTD does not alter the FP rate of the BDD.

In Fig. 6 (a), we plot the variation of $\eta'(\delta)$ as a function of $\gamma(\mathbf{H}_t, \mathbf{H}'_{t'})$ for different values of δ. In this figure, the y-axis represents the fraction of attacks for which $P'_D(\mathbf{a}) \geq \delta$, for a given $\gamma(\mathbf{H}_t, \mathbf{H}'_{t'})$. We observe that $\eta'(\delta)$ monotonically increases with $\gamma(\mathbf{H}_t, \mathbf{H}'_{t'})$, thus confirming our intuition that MTD perturbations with higher values of $\gamma(\mathbf{H}_t, \mathbf{H}'_{t'})$ are more effective in attack detection. E.g., for $\gamma = 0.44$, 97% of the attacks have a detection probability greater than 0.95. In practice, the defender can run these simulations to determine an appropriate γ_{th} for meeting a desired level of attack detection.

Comparison With Existing Work: We also perform simulations to compare our MTD selection approach with state of the art [11], [12], [13]. Similar to the related work, we implement MTD by selecting random MTD perturbations that are constrained to be within 2% of the optimal value. We plot $\eta'(\delta)$ as a function of δ for five such randomly-chosen perturbations in Fig. 7. It can be seen that $\eta'(\delta)$ exhibits high variability across the trials, implying that the randomly chosen MTD perturbations cannot always guarantee effective attack detection.

Further, out of 500 such randomly chosen perturbations (known also as the *keyspace* [11], [12]), we count the fraction of perturbations which satisfy $\eta'(\delta) \geq 0.9$ for different values of δ, and plot the results in Fig. 8. We observe that less 10% of the randomly-selected MTD perturbations satisfy $\eta'(0.9) \geq 0.9$. In contrast, the MTD perturbations chosen according to our approach can always guarantee a certain effectiveness, once the subspace angle threshold γ_{th} is adjusted to an appropriate value. This highlights the importance of designing the MTD according to the formal design criterion advanced in this work.

To show the scalability of the proposed approach to larger bus systems, we plot the $\eta'(\delta)$ as a function of $\gamma(\mathbf{H}_t, \mathbf{H}'_{t'})$ for the IEEE 30-bus system in Fig. 6 (b). We use default settings provided in the MATPOWER case file. We observe results similar to those for the IEEE 14-bus system, i.e., perturbations which have a higher value of $\gamma(\mathbf{H}_t, \mathbf{H}'_{t'})$ are more effective in terms of attack detection.

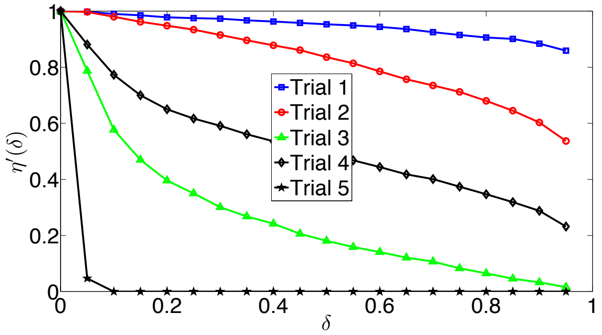


Fig. 7: MTD effectiveness under five randomly chosen MTD perturbations in IEEE 14-bus system. FP rate is set to 5×10^{-4}.

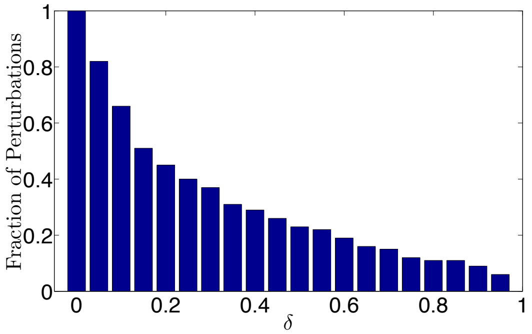


Fig. 8: Fraction of randomly-chosen MTD perturbations that satisfy $\eta'(\delta) \geq 0.9$.

C. Simulation Results With Dynamic Load

In the next set of simulations, we consider dynamic load. We use a load data trace from New York state for one day (25-JAN-2016) [31] sampled hourly, and feed it to the IEEE 14-bus system. The simulations are performed every hour. At each hour, $C_{\text{OPF},t}$ is computed by solving (1) with the load input of the corresponding hour. On the other hand, $C'_{\text{OPF},t'}$ is computed by solving (4) assuming that the attacker's knowl-

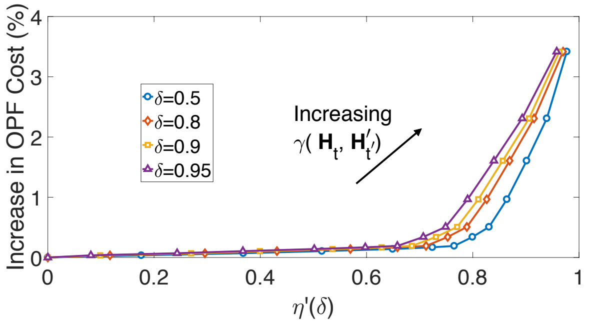


Fig. 9: Tradeoff between MTD's effectiveness and operational cost in IEEE 14-bus system. The data corresponds to 6 PM.

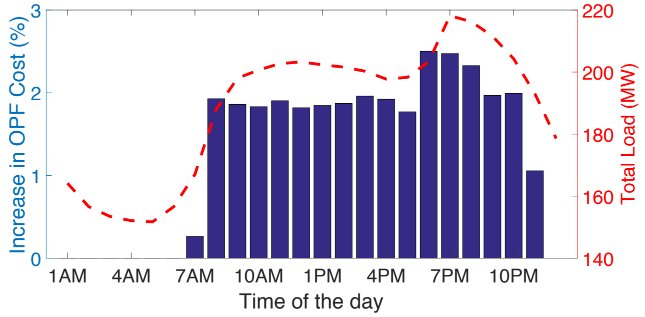


Fig. 10: MTD operational cost over a day computed using New York state hourly load data trace (25-JAN-2016).

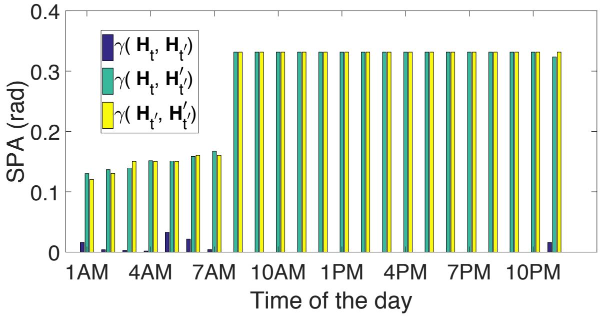


Fig. 11: Smallest principal angle (in radians) between pre-perturbation and post-perturbation measurement matrices.

edge is outdated by 1 hour. For example, while computing the MTD $\mathbf{H}'_{t'}$ at 9 AM, we assume that the attacker has acquired the knowledge of the measurement matrix $\mathbf{H}_t$ at 8 AM. (Recall from our previous discussion in Sec. IV-A that hourly MTD perturbations are realistic for practical systems.)

MTD Tradeoff: In Fig. 9, we plot of the tradeoff between $\eta'(\delta)$ and the operational cost for data corresponding to 6 PM. We make the following observations. For low values of $\eta'(\delta)$, the operational cost is nearly zero. However, as $\gamma(\mathbf{H}_t, \mathbf{H}'_{t'})$ and consequently $\eta'(\delta)$ is increased, the MTD incurs a non-trivial operational cost. In particular, the cost increases steeply for values of $\eta'(\delta)$ very close to 1. E.g., for $\delta = 0.9$, an increase in the value of $\eta'(\delta)$ from 0.8 to 0.9 changes the MTD operational cost from 0.96% to 2.31%. These results suggest that the defender must carefully choose an appropriate level of attack detection while taking into account the increase in operational cost.

MTD Operational Cost Over a Day: We also perform simulations to show how the cost varies over the day. At each hour, we adjust the subspace angle threshold γ_{th} numerically such that the MTD perturbation achieves effectiveness of $\eta'(0.9) \geq 0.9$. The corresponding value of $\gamma(\mathbf{H}_{t'}, \mathbf{H}'_{t'})$ is shown in Fig. 11. The rest of the bus settings is identical to the previous simulation. The variation of MTD operational cost and the aggregate load are shown in Fig. 10. It can be observed that the MTD operational cost increases at higher load. This can be explained as follows. When the system load is low, there will be a significant buffer capacity between the branch power flows and the corresponding flow limits. If the difference in power flows between the two systems (with and without MTD) is within the buffer capacity, then the generator dispatch in the two systems will be identical (or close to each other). Thus, the corresponding MTD cost is low. At higher loads, the power system is significantly congested, and the branch power flows of the two systems (with and without MTD) will differ significantly. Consequently the generator dispatch in the two systems will be different leading to an increase in the OPF cost.

We also plot the quantities $\gamma(\mathbf{H}_t, \mathbf{H}_{t'})$ and $\gamma(\mathbf{H}_{t'}, \mathbf{H}'_{t'})$ for every hour in Fig. 11. We observe that $\gamma(\mathbf{H}_t, \mathbf{H}_{t'})$ is nearly zero for all the simulation instants. This is because the matrices $\mathbf{H}_t$ and $\mathbf{H}_{t'}$ do not differ significantly due to the temporal correlation of the system load between different simulation instants and their column spaces are nearly aligned. These results also validate the approximation $\gamma(\mathbf{H}_t, \mathbf{H}'_{t'}) \approx \gamma(\mathbf{H}_{t'}, \mathbf{H}'_{t'})$.

D. Discussion

To put the MTD operational cost in perspective, we can compare it against the potential cost of damage due to a BDD-bypassing attack. For example, prior work [5], [20] suggests that such an attack can increase the OPF cost by up to 28%, and additionally cause transmission line trips (considering IEEE 14-bus system with similar simulation settings). Our numbers suggest that the MTD's operational cost is comparatively significantly smaller. In practice, based on its own deployment scenario and other factors like estimated likelihood of attacks, the SO can make similar comparisons to assess the merits of adopting the MTD defense.

VIII. Conclusions

We addressed the problem of selecting MTD reactance perturbations that are truly effective in thwarting stealthy FDI attacks against SE in power grids. We devised a novel metric to quantify the MTD's effectiveness, and identified key design criteria to compute effective MTD perturbations in practice. We also showed that the effective MTD may incur a non-trivial operational cost, and provided analysis to expose the cost-benefit tradeoff of the MTD in an OPF framework. Our result offers MTD to system operators as an insurance against possible FDI attacks, and minimizes the cost of such insurance subject to an effectiveness constraint.

References

[1] "Confirmation of a coordinated attack on the Ukrainian power grid," http://bit.ly/1OmxfnG.

[2] "Analysis of the cyber attack on the Ukrainian power grid," http://bit.ly/2ohNwJ1.

[3] U.S. Department of Homeland Security, "Moving target defense," http://bit.ly/1pWSSVZ.

[4] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," in *Proc. ACM CCS*, 2009, pp. 21–32.

[5] Y. Yuan, Z. Li, and K. Ren, "Quantitative analysis of load redistribution attacks in power systems," *IEEE Trans. Parallel Distrib. Syst.*, vol. 23, no. 9, pp. 1731–1738, Sept 2012.

[6] R. Tan, H. H. Nguyen, E. Y. S. Foo, X. Dong, D. K. Y. Yau, Z. Kalbarczyk, R. K. Iyer, and H. B. Gooi, "Optimal false data injection attack against automatic generation control in power grids," in *ACM/IEEE ICCPS*, Apr. 2016, pp. 1–10.

[7] R. B. Bobba, K. M. Rogers, Q. Wang, H. Khurana, K. Nahrstedt, and T. J. Overbye, "Detecting false data injection attacks on DC state estimation," in *Proc. Workshop on Secure Control Systems (SCS)*, Apr. 2010. [Online]. Available: http://bit.ly/2fYcLZ4

[8] G. Dan and H. Sandberg, "Stealth attacks and protection schemes for state estimators in power systems," in *Proc. IEEE SmartGridComm*, Oct 2010, pp. 214–219.

[9] T. T. Kim and H. V. Poor, "Strategic protection against data injection attacks on power grids," *IEEE Trans. on Smart Grid*, vol. 2, no. 2, pp. 326–333, June 2011.

[10] U.S. Department of Energy, "Factors affecting PMU installation costs," https://tinyurl.com/kz24nyb.

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

[12] K. R. Davis, K. L. Morrow, R. Bobba, and E. Heine, "Power flow cyber attacks and perturbation-based defense," in *SmartGridComm*, Nov 2012, pp. 342–347.

[13] M. A. Rahman, E. Al-Shaer, and R. B. Bobba, "Moving target defense for hardening the security of the power system state estimation," in *Proc. First ACM Workshop on Moving Target Defense*, 2014, pp. 59–68.

[14] D. Divan and H. Johal, "Distributed FACTS; A new concept for realizing grid power flow control," *IEEE Trans. Power Syst.*, vol. 22, no. 6, pp. 2253–2260, Nov 2007.

[15] K. M. Rogers and T. J. Overbye, "Some applications of distributed flexible AC transmission system (D-FACTS) devices in power systems," in *North American Power Symposium (NAPS)*, Sept 2008, pp. 1–8.

[16] A. Teixeira, S. Amin, H. Sandberg, K. H. Johansson, and S. S. Sastry, "Cyber security analysis of state estimators in electric power systems," in *IEEE CDC*, Dec 2010, pp. 5991–5998.

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

[18] X. Li, H. V. Poor, and A. Scaglione, "Blind topology identification for power systems," in *Proc. SmartGridComm*, Oct 2013, pp. 91–96.

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

[20] Y. Yuan, Z. Li, and K. Ren, "Modeling load redistribution attacks in power systems," *IEEE Trans. Smart Grid*, vol. 2, no. 2, 2011.

[21] O. Kosut, L. Jia, R. J. Thomas, and L. Tong, "Malicious data attacks on the smart grid," *IEEE Trans. Smart Grid*, vol. 2, no. 4, pp. 645–658, Dec 2011.

[22] L. Liu, M. Esmalifalak, Q. Ding, V. A. Emesih, and Z. Han, "Detecting false data injection attacks on power grid by sparse optimization," *IEEE Trans. Smart Grid*, vol. 5, no. 2, pp. 612–621, March 2014.

[23] S. Antonatos, P. Akritidis, E. P. Markatos, and K. G. Anagnostakis, "Defending against hitlist worms using network address space randomization," *Comput. Netw.*, vol. 51, no. 12, pp. 3471–3490, Aug. 2007.

[24] J. H. Jafarian, E. Al-Shaer, and Q. Duan, "Openflow random host mutation: Transparent moving target defense using software defined networking," in *Proc. HotSDN*, 2012, pp. 127–132.

[25] A. Wood and B. Wollenberg, *Power Generation, Operation, and Control*. A Wiley-Interscience, 1996.

[26] "Shodan," https://www.shodan.io/.

[27] 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 Trans. Power Syst.*, vol. 26, no. 1, pp. 12–19, Feb 2011.

[28] T. Tao, *An introduction to measure theory*, ser. Graduate Studies in Mathematics. American Mathematical Society, 2011.

[29] A. Bjoerck and G. H. Golub, "Numerical methods for computing angles between linear subspaces," Stanford, CA, USA, Tech. Rep., 1971.

[30] "IEEE 14-Bus System," http://icseg.iti.illinois.edu/ieee-14-bus-system/.

[31] "NYISO load data," https://tinyurl.com/kx3h82t.

[32] C. D. Meyer, Ed., *Matrix Analysis and Applied Linear Algebra*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2000.

[33] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*. John Wiley & Sons, 1982.

Appendix A: Proof of Proposition 1

To simplify notation, in this appendix, we drop the time subscripts t and t' from the relevant quantities.

A sketch of the proof is as follows. First, we express the residual r' as the sum of two components, a noise component $\mathbf{r}'_n$ and an attack component $\mathbf{r}'_a$, given by $r' = ||\mathbf{r}'_n + \mathbf{r}'_a||$. We then show that for attacks that satisfy the condition of Proposition 1, $\mathbf{r}'_a = \mathbf{0}$, and hence their detection probability is no greater than the FP rate.

We proceed with the first step of the proof. Recall the expression of $r' = ||\mathbf{z}' - \mathbf{H}'\widehat{\boldsymbol{\theta}'}||$, where $\mathbf{z}' = \mathbf{H}'\boldsymbol{\theta}' + \mathbf{n} + \mathbf{H}\mathbf{c}$, $\widehat{\boldsymbol{\theta}'} = (\mathbf{H}'^T\mathbf{W}\mathbf{H}')^{-1}\mathbf{H}'^T\mathbf{W}\mathbf{z}'$. It can be simplified as

$$\begin{aligned} r' &= ||\mathbf{z}' - \mathbf{H}'(\mathbf{H}'^T\mathbf{W}\mathbf{H}')^{-1}\mathbf{H}'^T\mathbf{W}\mathbf{z}'|| \\ &= ||\mathbf{H}'\boldsymbol{\theta}' + \mathbf{n} + \mathbf{H}\mathbf{c} \\ &\qquad - \mathbf{H}'(\mathbf{H}'^T\mathbf{W}\mathbf{H}')^{-1}\mathbf{H}'^T\mathbf{W}(\mathbf{H}'\boldsymbol{\theta}' + \mathbf{n} + \mathbf{H}\mathbf{c})|| \\ &= ||(\mathbf{I} - \boldsymbol{\Gamma}')\mathbf{n} + (\mathbf{I} - \boldsymbol{\Gamma}')\mathbf{H}\mathbf{c}||, \end{aligned} \tag{5}$$

where $\boldsymbol{\Gamma}' = \mathbf{H}'(\mathbf{H}'^T\mathbf{W}\mathbf{H}')^{-1}\mathbf{H}'^T\mathbf{W}$. We note that r' consists of two components, a noise component $\mathbf{r}'_n \triangleq (\mathbf{I} - \boldsymbol{\Gamma}')\mathbf{n}$, and an attack component $\mathbf{r}'_a \triangleq (\mathbf{I} - \boldsymbol{\Gamma}')\mathbf{H}\mathbf{c}$. If $\mathbf{r}'_a = \mathbf{0}$, then the detection probability of $\mathbf{a}$ is no greater than the FP rate α, and hence, the attack is undetectable under the MTD perturbation $\mathbf{H}'$. Note that for all the attacks $\mathbf{a} = \mathbf{H}\mathbf{c} \in Col(\mathbf{H}')$, $\mathbf{r}'_a = \mathbf{0}$. In other words, the system of equations $\mathbf{H}\mathbf{c} = \mathbf{H}'\mathbf{c}'$ must be consistent, for some $\mathbf{c}' \in \mathbb{R}^N$. This condition holds true if and only if rank$(\mathbf{H}')$ = rank$([\mathbf{H}' \ \mathbf{H}\mathbf{c}])$ [32].

Appendix B: Proof of Theorem 1

A sketch of the proof is as follows. We prove the first statement by showing that for an MTD $\mathbf{H}'$ satisfying the orthogonality condition, $\mathbf{r}'_a = \mathbf{0}$ if an only if $\mathbf{c} = \mathbf{0}$. Thus it follows that there are no non-zero attacks that are undetectable under such an MTD. To prove the second statement, we show that $P'_D(\mathbf{a})$ increases as we increase $||\mathbf{r}'_a||$. Furthermore, we show that $||\mathbf{r}'_a||$ achieves its maximum value under the MTD perturbation that satisfies the conditions of Theorem 1.

We begin with the proof of the first statement of Theorem 1. If $Col(\mathbf{H}')$ is the orthogonal complement of $Col(\mathbf{H})$, then

$\mathbf{H'}^T\mathbf{WHc} = \mathbf{0}$, $\forall \mathrm{c} \in \mathbb{R}^N$, since $\mathbf{Hc} \in Col(\mathbf{H})$. In this case, $\mathbf{r}'_a$ becomes

$$\mathbf{r}'_a = \mathbf{Hc} - \mathbf{H}'(\mathbf{H'}^T\mathbf{WH}')^{-1}\mathbf{H'}^T\mathbf{WHc} = \mathbf{Hc}.$$

Recall that an attack is undetectable if $\mathbf{r}'_a = \mathbf{0}$. For MTD $\mathbf{H}'$ that satisfies the orthogonality condition, substituting for $\mathbf{r}'_a$ from (6), we have that $\mathbf{Hc} = \mathbf{0}$. Since $\mathbf{H}$ is a full rank matrix, the set of equations $\mathbf{Hc} = \mathbf{0}$ has a unique solution $\mathbf{c} = \mathbf{0}$ [32]. Hence, there are no non-zero undetectable attacks of the form $\mathbf{a} = \mathbf{Hc}$.

Next, we prove the second statement of Theorem 1. First, note that under any MTD $\mathbf{H}'$, $||\mathbf{r}'_a||$ can be bounded as $0 \leq ||\mathbf{r}'_a|| \leq ||\mathbf{a}||$. The lower bound is true in a straightforward manner. The upper bound follows from

$$||\mathbf{r}'_a|| = ||(\mathbf{I} - \mathbf{\Gamma}')\mathbf{a}|| \leq ||(\mathbf{I} - \mathbf{\Gamma}')||\ ||\mathbf{a}|| = ||\mathbf{a}||, \qquad (6)$$

where the last equality is due to the fact that $\mathbf{I} - \mathbf{\Gamma}'$ is a projection matrix and hence has unit norm. Furthermore, under any MTD $\mathbf{H}'$, $r' = ||\mathbf{r}'_n + \mathbf{r}'_a||$ follows a *noncentral chi-square* distribution [33] with its noncentrality parameter equal to $||\mathbf{r}'_a||$ (since $\mathbf{r}'_n + \mathbf{r}'_a$ is a Gaussian random variable with $\mathbf{r}'_a$ as its mean).

For a non-central chi-square distributed random variable X, $\mathbb{P}(X \geq \tau)$ increases by increasing the noncentrality parameter. Hence, we can conclude that the quantity $P'_D(\mathbf{a}) = \mathbb{P}(r' \geq \tau)$ increases by increasing $||\mathbf{r}'_a||$. For an attack vector $\mathbf{a}$, the quantity $||\mathbf{r}'_a||$ depends on the choice of MTD $\mathbf{H}'$. Thus, we can conclude that MTD perturbations that yield a greater value of $||\mathbf{r}'_a||$ can detect the attack vector $\mathbf{a}$ with higher probability (i.e., $P'_D(\mathbf{a})$ is higher).

In particular, for MTD $\mathbf{H}'$ that satisfies the conditions of Theorem 1, from (6), we note that $||\mathbf{r}'_a|| = ||\mathbf{a}||$, which is also the maximum value of $||\mathbf{r}'_a||$. Therefore, such an MTD achieves the maximum possible value of $P'_D(\mathbf{a})$.

Appendix C: Conjecture of Section 5.3

In this appendix, we present arguments that the attack detection probability $P'_D(\mathbf{a})$ increases as we select MTD perturbations with higher $\gamma(\mathbf{H}, \mathbf{H}')$. We use the short-hand notation $f(\mathbf{u}, \mathbf{v})$ to represent the quantity $\max_{\substack{\mathbf{u}\in\mathcal{F},\mathbf{u}\in\mathcal{G}\\ ||\mathbf{u}||=1,||\mathbf{v}||=1}} |\mathbf{u}^H\mathbf{v}|$.

The conjecture can be argued by examining the dependence of $||\mathbf{r}'_a||$ on $\gamma(\mathbf{H}, \mathbf{H}')$ in the following three cases:

- Case 1: When $Col(\mathbf{H}')$ is the orthogonal complement of $Col(\mathbf{H})$, we have that $f(\mathbf{u}, \mathbf{v}) = 0$ (since $\mathbf{u}^H\mathbf{v} = 0$, $\forall \mathbf{u} \in Col(\mathbf{H}), \mathbf{v} \in Col(\mathbf{H}')$), and $\gamma(\mathbf{H}, \mathbf{H}') = cos^{-1}(0) = \pi/2$. From the arguments in Appendix B, recall that in this case, $||\mathbf{r}'_a|| = ||\mathbf{a}||$.
- Case 2: When $Col(\mathbf{H})$ and $Col(\mathbf{H}')$ are identical (e.g. when $\mathbf{H}' = (1 + \eta)\mathbf{H}$), we have that $f(\mathbf{u}, \mathbf{v}) = 1$, and $\gamma(\mathbf{H}, \mathbf{H}') = cos^{-1}(1) = 0$. In this case, after straightforward simplification, it can be shown that $||\mathbf{r}'_a|| = 0$.
- Case 3: For $0 \leq \gamma \leq \pi/2$, from reference [16], we have the following bound

  $$||\mathbf{r}'_a|| \leq \sin(\gamma(\mathbf{H}, \mathbf{H}'))||\mathbf{a}||. \qquad (7)$$

  Note that the bound of (7) increases as $\gamma(\mathbf{H}, \mathbf{H}')$ increases, which suggests that $||\mathbf{r}'_a||$ also increases.

The conjecture can be justified from the observation in these three cases and using the fact that $P'_D(\mathbf{a})$ increases as $||\mathbf{r}'_a||$ increases (Appendix B).