Stealthy MTD Against Unsupervised Learning-Based Blind FDI Attacks in Power Systems

Martin Higgins, *Graduate Student Member, IEEE*, Fei Teng, *Member, IEEE*, and Thomas Parisini, *Fellow, IEEE*

***Abstract*— This paper examines how moving target defenses (MTD) implemented in power systems can be countered by unsupervised learning-based false data injection (FDI) attack and how MTD can be combined with physical watermarking to enhance the system resilience. A novel intelligent attack, which incorporates dimensionality reduction and density-based spatial clustering, is developed and shown to be effective in maintaining stealth in the presence of traditional MTD strategies. In resisting this new type of attack, a novel implementation of MTD combining with physical watermarking is proposed by adding Gaussian watermark into physical plant parameters to drive detection of traditional and intelligent FDI attacks, while remaining hidden to the attackers and limiting the impact on system operation and stability.**

***Index Terms*— Cybersecurity, false data injection attacks, power systems state estimation, moving target defense, and physical watermarking.**

I. Introduction

THE modern power system is increasingly dependent on communication integrated devices for efficiency, reliability and control. The higher levels of inter-connectivity in the infrastructure and a ubiquitous use of communications have resulted in new types of vulnerabilities which have not been fully covered by the existing defense frameworks. Occurrences such as the 2015 cyber-attack against distribution companies in Ukraine [1] have drawn attention to the field of defense against cyber-threats. The Ukraine attack took many months of infiltration and was successful in compromising the SCADA system and de-energizing a portion of the grid for a few hours. However, the attack itself was discovered almost instantly once implemented. If the attackers had opted for a stealthy attack type, such as FDI attacks, they may have been able to continue attacking for months or years without being detected and the eventual consequences could have been much greater.

FDI attacks, first outlined in [2], involve altering system measurements to corrupt a network operator's state estimation process and cause negative consequences such as line overloading or outage masking [3]. A comprehensive review of FDI attacks can be found in [4]. FDI attacks need to remain undetected by the network operator to be effective. To this end, FDI attacks compete with bad data detectors (BDD) within state estimation processes. In modern energy management systems (EMS), the BDD at the power system level relies on weighted-least squares (WLS) and chi-squared error testing [5], meaning an attacker needs to structure the attack based on the system model in order to remain undetected. Initial models for FDI attacks assumed full knowledge of the system and full access to meter measurements within the system [2]. An incomplete knowledge attack was introduced in [6], which showed a system could be attacked with only partial knowledge of the system topology and a subset of meter measurements. In [7], the blind FDI attack is introduced, which requires no system knowledge provided the attacker has access to all meters within the attacked grid system. The blind FDI attack uses independent component analyses (ICA) to map the inter-correlations of the visible meter measurements to create an approximation for the power flow model. A more effective version of the attack which utilizes partial susceptance knowledge was developed in [8], allowing an islanded approach where the visible or 'high knowledge' parts of the system could be attacked by the standard-FDI attack while low information areas by the blind approach. Some recent studies enhanced FDI attacks by combining with other forms of attacks, such as denial of service (DoS) attack [9].

In addition, data-driven approaches have recently been applied to FDI attacks, although mostly from the defenders perspective [10], [11]. In [12], singular value decomposition is used to construct attack vectors without knowing the underlying system measurement matrix. In [13], two strategies using subspace separation are suggested: one aims to use estimated system subspace to hide attack vectors and another aims to

Manuscript received April 14, 2020; revised July 24, 2020 and September 11, 2020; accepted September 16, 2020. Date of publication September 28, 2020; date of current version November 9, 2020. This work was supported in part by the ESRC under Grant ES/T000112/1, in part by the EPSRC Centre for Doctoral Training in Future Power Networks and Smart Grids under Grant EP/L015471/1, in part by the European Union's Horizon 2020 research and innovation programme under Grant 739551 (KIOS CoE), and in part by the Italian Ministry for Research in the framework of the 2017 Program for Research Projects of National Interest (PRIN), under Grant 2017YKXYXJ. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Wei Yu. *(Corresponding author: Fei Teng.)*

Martin Higgins and Fei Teng are with the Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, U.K. (e-mail: f.teng@imperial.ac.uk).

Thomas Parisini is with the Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, U.K., also with the KIOS Research and Innovation Centre of Excellence, University of Cyprus, 20537 Nicosia, Cyprus, and also with the Department of Engineering and Architecture, University of Trieste, 34127 Trieste, Italy (e-mail: t.parisini@gmail.com).

Digital Object Identifier 10.1109/TIFS.2020.3027148

mislead BDD so that non-attacked measurements are removed. These methods allow for admittance values to be estimated but require a large number of historical measurements. In [14], sparse FDI attacks against wide area measurement systems and defense methods are explored. Using historical data to mount FDI by using multiple linear regression model was outlined in [15]. However, the above literature all focus on fixed network topology, while whether and how data-driven approaches can be applied to design FDI attacks under intentional or unintentional topology changes has not yet been investigated.

In fact, as FDI attacks are dependent on the characteristics of the physical system, a body of work has emerged to utilize the physical system to actively defend against the attacks. In particular, MTD is proposed through either transmission switching [16] or admittance perturbation via distributed flexible AC transmissions (D-FACTS) devices [17], [18] to change physical system topology to proactively drive BDD. An analysis of MTD against FDI attacks is offered in [19] where they prove the susceptibility of isolated state measurements and design an algorithm for branch perturbation selection. Some limitations of MTD were explored in [20]. With the increasing capability of the attackers, there are growing interests in the research community to design new forms of MTD which can hide its existence to the attacker. One of the key state-of-the-art papers in this field is [21], which presents an enhanced hidden MTD model to make the topology change invisible to an attacker via identifying alternative topology and state combinations under the same power flow profile. Whilst this method is clearly effective, it relies on being able to find alternative topology and states to maintain constant power flows, which can be computationally expensive and even infeasible in a system with limited acceptable state ranges.

In this context, this paper examines the vulnerability of current MTD strategies under unsupervised learning-based FDI attacks and develops a new form of stealthy MTD to increase system resilience. Our main contributions are twofold:

- On the attacking front, this work introduces a novel new counter-MTD technique. Where previous FDI attacks have been designed against static systems, we seek to offer new attacking considerations in the presence of dynamic systems with MTD. The proposed intelligent attack under zero system knowledge assumption combines dimensionality reduction and unsupervised learning to identify underlying clusters associated with network topology and design the corresponding attack vector. The method is shown to be effective and stealthy against traditional MTD.
- From the defensive perspective, we introduce a new implementation of MTD to drive detection against traditional and intelligent FDI attacks. The proposed defense strategy combines MTD and physical watermarking concept [22], for the first time, to add a Gaussian watermark into physical plant parameters. As the added watermark mimics the underlying noise of the system, the physical changes driven by MTD stay hidden. The physical watermarking is combined with cumulative error monitoring to spot minor but sustained changes in the system to trigger an alarm.

The rest of this paper is organized as follows. The problem formulation and underlying basis for FDI attacks and MTD through topology and parameter changes are outlined in section II. Section III details the design of the proposed intelligent attack, justification for algorithm selection and demonstration of its effectiveness in circumventing MTD. Section IV proposes the Gaussian style physical watermark in physical system parameters with cumulative error detection approach. Section V contains the results and analysis of the different types of MTD as applied to blind FDI attacks and Section VI concludes the paper.

II. Problem Formulation

A. State Estimation

A static power system problem is considered, consisting of a set of n state variables $\mathbf{x} \in \mathbb{R}^{n\times 1}$ estimated by analysing a set of m meter measurements $\mathbf{z} \in \mathbb{R}^{m\times 1}$ and corresponding error vector $\mathbf{e} \in \mathbb{R}^{m\times 1}$. The non-linear vector function $\mathbf{h(.)}$ relating meter measurements $\mathbf{z}$ to states $\mathbf{h(x)} = (h_1(\mathbf{x}), h_2(\mathbf{x}), \ldots, h_m(\mathbf{x}))^T$ is shown by

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

With real power flow measurements under the non-linear expression defined by

$$P_{ij} = V_i^2 g_{ij} - V_i V_j g_{ij} \cos \Delta\theta_{ij} - V_i V_j b_{ij} \sin \Delta\theta_{ij}. \tag{2}$$

For simplicity and clarity, we first derive the initial formulation and condition based on the linear DC approximation of AC state estimation. A mathematical extension and simulations on original system are then preformed in later sections to demonstrate the applicability of the proposed methods in full AC state estimation.

As a result, the matrix formulation, represented by a linear regression model as a function of the Jacobian $\mathbf{H} \in \mathbb{R}^{m\times n}$ matrix and the state vector, can be expressed as:

$$\mathbf{z} = \mathbf{Hx} + \mathbf{e}. \tag{3}$$

The state estimation problem is to find the best fit estimate of $\hat{\mathbf{x}}$ corresponding to the measured power flow values of $\mathbf{z}$. Under the most widely used estimation approach, the state variables are determined by minimization of a WLS optimization problem as

$$\min_x J(\mathbf{x}) = (\mathbf{z} - \mathbf{Hx})^T \mathbf{W} (\mathbf{z} - \mathbf{Hx}). \tag{4}$$

$\mathbf{W}$ is a diagonal $m \times m$ matrix consisting of the measurement weights.

A solution for a minimal $\mathbf{J(x)}$ can be analytically obtained by taking the 1st derivative with respect to $\mathbf{x}$ and solving for 0, yielding $\hat{\mathbf{x}}$ defined by

$$\hat{\mathbf{x}} = (\mathbf{H}^T \mathbf{WH})^{-1} \mathbf{H}^T \mathbf{Wz}. \tag{5}$$

B. Bad Data Detection

The current approach in power systems operation for bad data detection is to use the 2-norm of the measurement residual with a detection threshold η [23]. The residual $\mathbf{r}$ is defined by the difference between the measured power flow values of $\mathbf{z}$

and the value calculated from the estimated state values $\hat{\mathbf{x}}$ and the known topology matrix $\mathbf{H}$

$$r = ||\mathbf{z} - \mathbf{H}\hat{\mathbf{x}}||_2. \tag{6}$$

Assuming the errors of state variable $\mathbf{x}$ are random, independent and follow a normal distribution with mean zero and unit $\mathcal{N}(0, \sigma^2)$, a chi-squared distribution model $\chi^2_{m-n,\alpha}$ with $m-n$ degrees of freedom and confidence interval α (typically 0.95 or 0.99) can be used to define the detection threshold as

$$\eta = \sigma\sqrt{\chi^2_{m-n,\alpha}}. \tag{7}$$

If $r_t > \eta$ BDD alarms will trigger and the system operator will discard the result, removing the elements from the residual calculation with large values and replacing with an appropriate pseudo-measurement, based on historical data.

C. Constructing Attack Vectors

In the case of an infinitely resourced and knowledgeable attack, the attacker can gain full access to the metering infrastructure and change measured power flows in any desired manner. In this case, it is inconsequential to design the attack that maintains the residual at a given value. The attacker can choose any linear combination of $\mathbf{Hc}$ where $\mathbf{c} \in \mathbb{R}^{n\times 1}$. The vector $\mathbf{c}$ can be selected to have the desired impact on the state vector $\mathbf{x}$:

$$\mathbf{z}_a = \mathbf{z} + \mathbf{a} = \mathbf{z} + \mathbf{Hc}. \tag{8}$$

The 2-norm residual remains unchanged as shown below:

$$r_a = \|(\mathbf{z} + \mathbf{a}) - \mathbf{H}(\hat{\mathbf{x}} + \mathbf{c})\|_2 = \|\mathbf{z} - \mathbf{H}\hat{\mathbf{x}}\|_2. \tag{9}$$

In a more realistic scenario, where the attacker has full access to the metering infrastructure but no understanding of how the network components interconnect or the branch admittance, the attacker has to commit a "blind" form of attack by estimating plausible attack vector models based on historical measurements. One way of achieving this is to utilize Blind Source Separation (BSS) techniques. This scenario has been outlined previously in [7]. The relationship between the state variables in a power system and latent independent variables $\mathbf{y}$ under a fixed topology $\mathbf{H}$ can be described by

$$\mathbf{x} = f(\mathbf{H}, \mathbf{y}). \tag{10}$$

In practice $\mathbf{y}$ represents the loads of power system which vary independently while the topology is fixed but other underlying latent variables may exist for some systems. The state vector x can be approximated as the first-order coefficient of the Taylor expansion $\mathbf{A}$ around y.

$$\mathbf{x} \approx \mathbf{Ay}. \tag{11}$$

Returning to the state estimation problem, the system states can then be expressed in terms of load such that

$$\mathbf{z} \approx \mathbf{HAy} + \mathbf{e}. \tag{12}$$

If the attacker can acquire $\mathbf{HA}$, an attack vector can be constructed with a value selected for a change in power flows $\delta\mathbf{y}$ shown by

$$\mathbf{z}_b = \mathbf{z} + \mathbf{HA}\delta\mathbf{y}. \tag{13}$$

A generalized form of blind source separation $\mathbf{u} = \mathbf{Gv}$ can be used, where $\mathbf{u}$ is the vector that can be directly observed, $\mathbf{G}$ is the fixed vector known as the mixing matrix and $\mathbf{v}$ the underlying vector of signals. The state estimation can be constructed in an equivalent manner such that:

$$\mathbf{z} = \mathbf{HAy} = \mathbf{Gy}. \tag{14}$$

Provided the errors follow a Gaussian distribution and do not contain gross errors, $\mathbf{HA}$ can be extracted using independent component analysis as shown previously in [7], [24].

D. AC Extension of Blind Attack

Similar to the DC attack, AC FDI attacks must satisfy the system model to remain hidden such that

$$\mathbf{z}_a = \mathbf{z} + \mathbf{a} = \mathbf{h}(\mathbf{x} + c). \tag{15}$$

This can be done without topology information either using the geometric approach [25] or a historical measurement based replay approach. *Chin et al* showed that where the vector angle between the normal power flows and attacking vector was defined by

$$\mathbf{z}^T\mathbf{a} = cos(\psi) \tag{16}$$

the attack can bypass AC detection provided the vector space angle between the attacking vector and measurement vector was close to zero such that

$$\mathbf{z}^T\mathbf{z}_a = 1. \tag{17}$$

Under these considerations, a regression model can be extracted to attack the system. Alternatively, in the case of limited information, the attacker can implement a replay style attack which reuses a previous vector from historical measurements such that

$$\mathbf{z}_t^a = \mathbf{z}_{t-q} \tag{18}$$

where q is used to denote a vector from a previous time period. Our AC simulations were built with this replay case in mind, but it should be noted both methods are susceptible to conventional MTD.

E. MTD Through Topology Changes

Under AC state estimation, system measurements will consist of real power flows defined by (2) and reactive power by

$$\begin{aligned} Q_{ij} = &-V_i^2(b_{ij} + b_{ij}^{sh}) + V_iV_jg_{ij}\cos\Delta\theta_{ij} \\ &- V_iV_jb_{ij}\sin\Delta\theta_{ij}. \end{aligned} \tag{19}$$

For real power residual, error at the individual measurement level will be the difference between the measured flows and

estimated value from the system model such that real power residual can be expressed as

$$r_{ij}^{P} = -P_{ij}^{m} + V_i^2 g_{ij} - V_i V_j g_{ij} \cos \Delta\theta_{ij} - V_i V_j b_{ij} \sin \Delta\theta_{ij}. \quad (20)$$

and reactive power flow residual can be expressed as

$$r_{ij}^{Q} = -Q_{ij}^{m} - V_i^2 (b_{ij} + b_{ij}^{sh}) + V_i V_j g_{ij} \cos \Delta\theta_{ij} - V_i V_j b_{ij} \sin \Delta\theta_{ij}. \quad (21)$$

In the AC state estimation, MTD can employ resistive as well as inductive components to introduce change. Alternatively, the SO can aim to force a state of non-convergence in the case of FDI which is done by violating the non-convergence criteria of the Newton-Raphson principle for power systems. Again, the alarm criteria will be the 2-norm value of the residual vector calculated by

$$\mathbf{r}_{ac} = \|\mathbf{z} - \mathbf{h}(\hat{\mathbf{x}})\|_2. \quad (22)$$

We derive here the analytical expression of the impact on residual of topology change for a linear system under attack vector $\mathbf{a} = \mathbf{Hc}$. Using the WLS formulation, r_a can be expressed as

$$r_a = \|(\mathbf{z} + \mathbf{Hc}) - \mathbf{H}(\mathbf{H}^T\mathbf{WH})^{-1}\mathbf{H}^T\mathbf{W}(\mathbf{z} + \mathbf{Hc})\|_2. \quad (23)$$

The attacker is assumed to have static topology knowledge and construct the injected attack vector $\mathbf{z}_a$ as a function of the original topology $\mathbf{H}_o$. The new topology with MTD applied is $\mathbf{H}_n$, which is only known by the SO. As a result, the measurement vector under attack $\mathbf{z}_a$ will be

$$\mathbf{z}_a = \mathbf{z} + \mathbf{H}_o\mathbf{c}. \quad (24)$$

The SO estimates $\hat{\mathbf{x}}$ via the WLS minimization using the visible $\mathbf{z}_a$ and $\mathbf{H}_n$. The min error estimate of $\hat{\mathbf{x}}_n$ will utilize the new topology $\mathbf{H}_n$ while the attack vector is developed based on the old topology $\mathbf{H}_o$. Consequently, the new residual will be a product of the attack vector based on old topology $\mathbf{H}_o\mathbf{c}$ and the WLS estimation based on the new topology as

$$r_n = \|\mathbf{z} + \mathbf{H}_o\mathbf{c} - \mathbf{H}_n(\mathbf{H}_n^T\mathbf{WH}_n)^{-1}\mathbf{H}_n^T\mathbf{W}(\mathbf{z} + \mathbf{H}_o\mathbf{c})\|_2. \quad (25)$$

Defining WLS minimization factor for the new topology as $\mathbf{F}_n$, which is fixed for a given topology as $\mathbf{F}_n = (\mathbf{H}_n^T\mathbf{WH}_n)^{-1}\mathbf{H}_n^T\mathbf{W}$, the residual 2-norm can be rewritten as

$$r_n = \|\mathbf{z} + \mathbf{H}_o\mathbf{c} - \mathbf{H}_n\mathbf{F}_n(\mathbf{z} + \mathbf{H}_o\mathbf{c})\|_2. \quad (26)$$

Considering the old topology $\mathbf{H}_o$ as a function of the new and system change $\mathbf{H}_n + \Delta\mathbf{H}$, the residual in terms of the new topology can hence be calculated as

$$r_n = \|\mathbf{z} + (\mathbf{H}_n + \Delta\mathbf{H})\mathbf{c} - \mathbf{H}_n\mathbf{F}_n(\mathbf{z} + (\mathbf{H}_n + \Delta\mathbf{H})\mathbf{c})\|_2. \quad (27)$$

$\mathbf{H}_n\mathbf{F}_n\mathbf{H}_n$ is the idempotent matrix of $\mathbf{H}$ and therefore $\mathbf{H}_n\mathbf{F}_n\mathbf{H}_n\mathbf{c} = \mathbf{H}_n\mathbf{c}$ the expression can be rearranged into

$$r_n = \|(1 - \mathbf{H}_n\mathbf{F}_n)\mathbf{z} + (1 - \mathbf{H}_n\mathbf{F}_n)\Delta\mathbf{Hc}\|_2. \quad (28)$$

As shown in (28), any $\Delta\mathbf{H}$ will change the residual value $\mathbf{r}_n$. The aim of defender is to select a value for $\Delta\mathbf{H}$ such that under attack vector $\mathbf{H}_o\mathbf{c}$, the new residual exceeds the alarm criteria (usually chi-squared criteria) $r_n > \sigma\sqrt{\chi^2_{m-n,\alpha}}$.

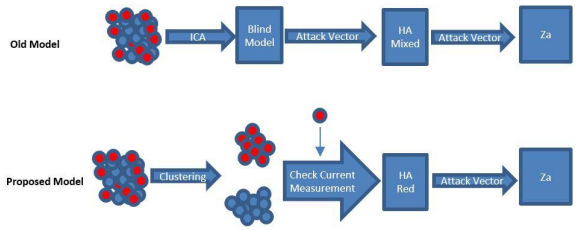

Fig. 1. Proposed algorithm process to circumvent MTD. Red and blue points corresponded to observed power flows from 2 different network configurations.

III. Clustering to Circumvent MTD

This section investigates how the data-driven approach can be applied to explore the vulnerability of existing MTD. In particular, an efficient method is proposed to identify changes in the network caused by the implementation of D-FACTS or switching through analysing the resultant power flow profiles. By doing so, the attacker can ensure only data points corresponding to the current configuration are used to create the blind attack. The proposed attack flows are as follows:

1) Observations of historical power flows are clustered into groups.
2) The clustering algorithm identifies the current power flow set to find corresponding measurements for the attack model.
3) The blind attack model is developed using only the data corresponding to the current power flow profile cluster.

A simple example of this process is illustrated and compared with the normal blind attack in Figure 1. To achieve this, we propose a combination of data prepossessing via T-distributed stochastic neighbour embedding (T-SNE) for dimensionality reduction followed by density based spatial clustering of application with noise (DBSCAN) to classify the data sets. We outline the justification for our chosen methods below.

A. Attack Design Considerations

Power transmission systems are by their very nature large. To design such data-driven attacks, one of the key considerations is to maintain the feasibility of implementation in real-time operation of large scale systems. Therefore, it is essential to circumvent the curse of dimensionality (CoD) within the context of this attack. We hence explore the use of T-SNE to reduce the dimensionality of data sets before applying the clustering algorithm. In addition, due to the blind nature of the attack, no prior knowledge of the number of underlying topologies can be assumed and therefore an unsupervised learning method, DBSCAN in this case, is developed.

1) T-SNE for Dimensionality Reduction: T-SNE is a form of dimensionality reduction which works by constructing probability distributions over pairs of objects containing high dimensionality [26]. T-SNE considers a set of N high dimension objects. d_i and d_j are two points within this set. σ_i is the variance of the Gaussian centred on data point d_i. The closeness of these data points is defined by the conditional

probability $p_{j|i}$ that point d_j would select d_i as a neighbour given that the neighbours are picked proportionately to a Gaussian centred around d_j. This is given by

$$p_{j|i} = \frac{\exp(-||d_i - d_j||^2/2\sigma_i^2)}{\sum_{k\neq i}\exp(-||d_i - d_j||^2/2\sigma_i^2)}. \tag{29}$$

The aim of T-SNE is to reduce these points into their low dimensional counterparts g_j and g_i. These have an equivalent conditional probability $q_{j|i}$ defined by

$$q_{j|i} = \frac{\exp\left(-||g_i - g_j||^2\right)}{\sum_{k\neq i}\exp((-||g_i - g_j||^2))}. \tag{30}$$

If the map points g_j and g_i correctly model the similarity between the high dimensional sets, the conditional probabilities $p_{j|i}$ and $q_{j|i}$ will be equal. The positions of g_i and g_j are determined via gradient descent between the distributions p and q, and is used to minimize the Kullback-Leiber (KL) divergence via cost function C [27] shown by

$$C = \sum_i KL(P_i||Q_i) = \sum_i\sum_j p_{j|i}\log\frac{p_{j|i}}{q_{j|i}}, \tag{31}$$

where P_i is the conditional probability distribution over all data points given data point d_i and Q_i is the conditional probability distribution over every other map point, given map point g_i.

Native T-SNE itself has a time complexity of $O(n^2)$ but this can be reduced to $O(n)$ by using optimization techniques as discussed in [28]. The brunt of the computational load is therefore taken by T-SNE which reduces the measurements of the network power flows into 2-dimensional space.

There are other possible unsupervised approaches for dimensionality reduction. Linear reduction algorithms such as principle component analysis (PCA) are one such example. PCA performs linear mapping to lower dimensional spaces and unlike T-SNE is deterministic rather than probabilistic. PCA being a linear algorithm means that it does have some computational benefits. However, PCA cannot represent complex polynomial relationships in the same way T-SNE can. Also, the KL divergence minimisation that T-SNE employs means much of the local structure of data is preserved in T-SNE which it is not to the same degree in PCA. We also consider that with the stated purpose of identifying like groupings of points T-SNE is also the most appropriate choice. The probabilistic neighbour assessment approach of T-SNE seeks to identify neighbours specifically which makes the output trend towards close and distinct cluster groups emerging (as shown in Figure 2). This makes it easy to identify groups for the next section of the attack algorithm (clustering and model building) to operate over.

2) DBSCAN for Unsupervised Learning: When the dimension reduced data set is received, a cluster algorithm will be applied to identify the underlining clusters of the data. Due to the blind nature of the attack, we propose using DBSCAN for the unsupervised clustering portion of this attack. The DBSCAN algorithm works as follows:

1) An initial starting point is randomly selected. This point is then marked as visited.

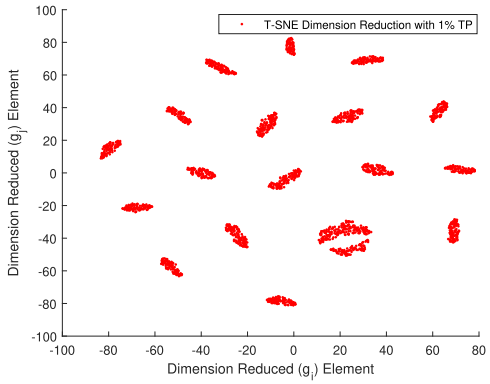

Fig. 2. Power flow profile observations of 1% admittance perturbation MTD applied to 19 lines intermittently under T-SNE dimensionality reduction. The X and Y axis values are non-dimensional probabilistic reductions (g_i & g_j from equation 30). The system is reduced from a 34 dimension meter IEEE 14-bus system.

2) The points adjacent to this point defined by ϵ are counted and added to a set
3) If the number of points exceeds the defined min point value the initial point is defined as a new cluster. This process is continued for all points in the neighbourhood
4) If the number of points is less than the min the point is defined as noise
5) These steps are repeated until the whole set has been clustered.

DBSCAN shows good benchmark performance against other forms of unsupervised learning [29] and also offers several relevant advantages to this form of the attack. DBSCAN has a time complexity of $O(n^2)$ but this can be reduced to $O(n\log n)$ with parameter optimisation [30], unlike hierarchical clustering which has a time complexity of $O(n^3)$ and is highly computationally intensive for large systems by comparison. DBSCAN also does not require pre-specification of the number of clusters (making it more appropriate for a blind style attack) and is robust against outlying data points and noise. Density-based Local Outlier Factor (LOF) was also considered and has previously been seen for FDI attack detection in [31]. LOF and DBSCAN are both density based methods. However, the stated aim of LOF is for anomaly detection while DBSCAN is more appropriate for direct clustering and finding groups. Also, the time complexity of DBSCAN is preferable to LOF. LOF has a time complexity of $O(n^2)$ [32] by comparison this represents the worst case time complexity of DBSCAN.

B. Intelligent Blind FDI Attack

This sub-section details the proposed intelligent blind FDI, as outlined in Algorithm 1. Once the attacker has obtained an adequate amount of measurement data, they can initiate the attack algorithm. When the latest measurement data arrive, initially, T-SNE is applied for dimensionality reduction of the sets of power flow observations into a two dimensional space. The reduced form of the data set is then clustered via the DBSCAN algorithm into distinct subgroups of like

measurements and the one corresponding to the current system topology is identified. The mixing matrix is subsequently derived based on this subgroup of data by using independent component analysis as per the normal blind attack. A vector of false data $\mathbf{z}_a$ containing the desired attack bias will be then calculated based on the mixing matrix.

Ultimately, the attacker does not know which model corresponds to the base case or the case with MTD implemented. The attacker simply knows there are multiple distinct underlying models and creates a series of models equal to the number of clusters. The attacker may be able to guess based on how the topologies represent in terms of timing which is the base (no MTD case) but this is largely irrelevant for the attack. A minimum cluster size check will also be implemented to ensure the attack has sufficient data to create the blind model.

Algorithm 1 DBSCAN Blind-ICA Attack

Input: A set of power flow observations $\mathbf{z}_{obvs}$

1: $\mathbf{Y}$ = tsne ($\mathbf{z}_{obvs}$) % Dimensionality reduction
2: **idx** = dbscan(Y,mpts,ϵ) % Cluster power flows
3: **For** i = 1:length(unique(idx)) % Assign pf to cell
j = [j,idx] % Assign cluster to obvs
A{i} = j(j(:,l)==i,:)% Assign obvs to cell
4: c = j(end) %check what profile current **z** is
5: **z** = A(c) % Select only corresponding **z** measurements
6: **IF** length(z) $< A_{min}$
7: **END** % End if not enough data for Blind ICA
ELSE % continue with calc
8: **HA** = FastICA(**z**) % Run fastica for **HA**
9: $\mathbf{z}_a = \mathbf{z} + \mathbf{HA}\delta\mathbf{y}$ % Apply attack vector

Output: false data $\mathbf{z}_a$

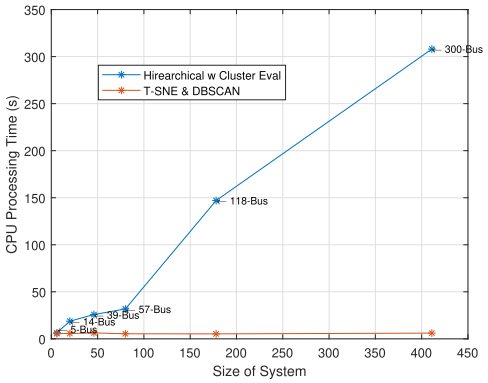

Fig. 3. CPU processing time for the combined T-SNE/DBSCAN algorithm with an equivalent hierarchical method with embedded cluster selection performed on systems of increasing size. Performed for 1000 observations.

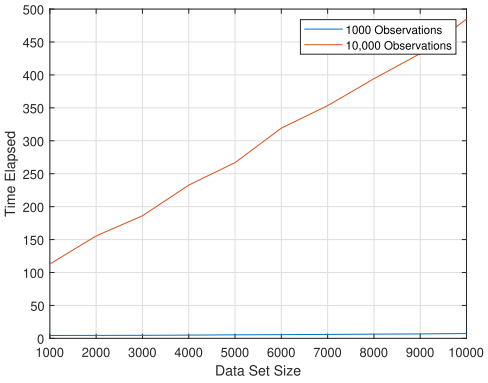

Fig. 4. CPU processing time for the combined T-SNE and DBSCAN algorithm for increasing size of random data array up to 10,000 data points. Performed for 1000 and 10,0000 observations.

C. Performance Analysis

To demonstrate the performance of the proposed algorithm, a case study is carried out on a system with 14 lines equipped with D-FACTS for MTD. As shown in Figure 2, the proposed algorithm successfully identifies and clusters the potential topology sets, even only minor changes on topology (1% of base admittance) are identified. The computational performance of T-SNE and DBSCAN for different IEEE standard systems is shown in Figure 3 and compared with hierarchical clustering with embedded cluster evaluation. The case studies are performed for 1000 sets of observations. We note that, for small scale systems (such as the 5-bus case), the computational performance are similar, but, as the system becomes larger, the time to completion grows quickly for hierarchical clustering.

1) Real-Time Attacks in Large Systems: For real time operation, the bottle-neck for attacking with this technique comes in the identification and classification of the last meter measurement set within the wider pool, i.e. the ability to identify which model the attack should be based upon. The proposed technique can also be practical for large systems which may contain a high number of measurements. In Figure 4 we show the time to completion for the T-SNE and DBSCAN portions of the algorithm in the presence of very large random arrays (up to 10k points). It demonstrates the (expected) linear relationship from the T-SNE time complexity. Even when considering large data arrays with a large number of observations, the time to perform the T-SNE/DBSCAN flow is relatively short. For example, using 10,000 observations for a 10,000 point system it took around 8.06 minutes using only an Intel Core i7-7820X CPU with 64GB of ram. We would expect that a highly motivated supranational attacker would have access to much more sophisticated hardware and would be able to execute such attacks even quicker.

D. Load Profile Bucketing

As seen in [21], large load variations can make distinguishing changes from MTD in the system hard. Fundamentally, load variation can be used itself to hide MTD. Therefore we consider an extension to reduce a highly variate load system back to a steady loads style assumption under which the attacker can have more success. The attacker will use a combination of T-SNE and discrete bucketing to group load sets by their full system profile. The load variation within these individual buckets will be small and equivalent to a steady loads style assumption. The attacker can then run the attack over one of these buckets and it works as if the steady load assumption were in place.

IV. Physical Gaussian Watermarking With CUSUM

While physical watermarking has not been applied in the power system space, the concept has been proposed in control systems such as in [22] where a watermark is added into LQG-based control signals to drive detection. However, the papers in these areas aren't true "physical" watermarks as they only change signal parameter dependencies and not the underlying physical plant itself. At the same time, it should be noted that while MTD in the form of D-FACTS control to change system topology has been explored, the use of watermarks in combination with MTD has not been investigated and there is an opportunity to incorporate a true physical watermark into the system plant to enhance the system security.

Previously, topology perturbation and transmission switching have been proposed as methods to drive detection of FDI attacks [17], [21], These methods implement significant changes to line admittance as required by the change needed in residual (typically around 10-20% for D-FACTS based changes), which may not only lead to interruption on system operation, but also provide opportunities for the data-driven attack to spot the existence of MTD and counter it. It is, hence, crucial that the deployment of MTD can remain hidden to the attacker.

In this work, it is assumed that the SO will incorporate the capability of D-FACTS devices into the OPF model to optimize and select the lowest cost scenario, as shown in [33]. MTD will then be applied around this point. As outlined in [34], there is a non-trivial cost incurred when applying conventional MTD. This cost comes in the form of non-optimal usage of power system assets. Where previously D-FACTs were applied to minimise losses from reactive power, they are now being used for MTD purposes away from this optimal point. As a result, the defender will wish to reduce the overall application of MTD.

In this context, this section proposes a novel method to achieve this by combining MTD with physical watermarking, which makes the MTD itself indistinguishable from the noise profile of the system, and monitoring sequential errors for long-run trends by using cumulative summed monitoring (CUSUM). CUSUM is a sequential analysis technique which monitors for change detection over a number of measurements. Samples taken from the process are assigned a weighting and summed to monitor change detection. In this case, we will monitor the measured residual r under MTD defined by

$$CEM_t = \sum_{j=1}^{t} r_j - T. \tag{32}$$

CEM is the Cumulative Error Monitor (CEM) decision statistic, T is the target value of residual dictated by monitoring the statistic under normal conditions and t is the number of periods in a measurement set, with upper and lower control limits CEM_t^+ and CEM_t^-. As r is an absolute value, the lower bound CEM_t^- will be 0. CEM_t^+ can be selected based on engineering judgement from prior observations. Usually the upper bound can be defined in terms of the residual variance and mean value under no attack:

$$CEM_t^+ = \bar{r} + B\sigma_r. \tag{33}$$

where B is defined by the user based on previous observations and minimising type 2 error.

The proposed defense strategy introduces these minor errors by using D-FACTS devices to alter the line admittance by a vector $\mathbf{w}$. The size of admittance changes applied to each line is based on the output from a pseudo random number generator (PRNG), the seed value of which is only known by the network operator. This can be achieved with existing technology via a Unified Power Flow Controller (UPFC) in combination with a processing unit. The watermark may be applied selectively such that

$$w_m \in \{0, \mathcal{N}(0,\ p)\}. \tag{34}$$

where p is the max change applied to the branch admittance.

The resulting power flow profile under physical watermarking will be equal to

$$\mathbf{z}_w = (\mathbf{H} + \mathbf{w})\mathbf{x} + \mathbf{e}. \tag{35}$$

where $\mathbf{w}$ represents the vector of admittance changes applied to branches and is known to the SO.

The impact of applying a Gaussian style watermark in physical system parameters is shown in Figure 5. Compared with direct binary perturbation, the proposed MTD show similar profile as underline noise and make it extremely hard for clustering algorithm to identify the existence of MTD or to counter it.

The key advantages of the proposed defense mechanism can be summarised as below, which will be validated in the next session:

1) As the proposed MTD is on magnitude with the noise levels, the change in power flow observations resulting from the MTD becomes difficult to be identified. Therefore, MTD stays stealthy to the attacker.
2) Due to the stealthiness of the proposed MTD, it significantly increases the chance of the detection of FDI

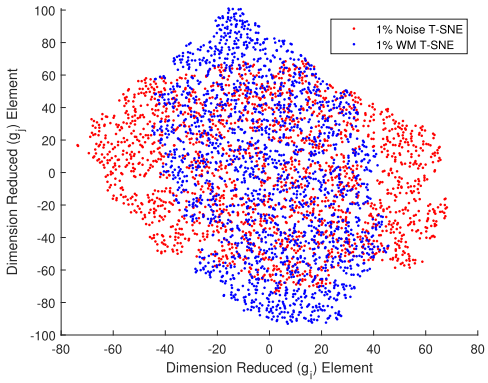

Fig. 5. Power flow profile observations of 1% Gaussian watermark MTD applied to 14 lines intermittently under T-SNE dimensionality reduction. The X and Y axis values are non-dimensional probabilistic reductions (g_i & g_j from equation 30). The system is reduced from a 34 dimension meter IEEE 14-bus system.

attack and is specifically resilient to intelligent attack types such as the proposed DBSCAN blind-ICA attack.

3) The significantly-reduced magnitude of topology changes leads to fewer interruptions on the system stability and economic operation.

A CEM violation indicates there might be an attacker present but a false positive is also a possibility. We believe the best way to use the CEM approach would be as a form of an event trigger for more substantial MTD i.e. if a CEM violation is witnessed the SO implements some larger scale MTD to probe for attacks. This would help the system operator minimise the use of costly MTD whilst also offering additional protection of the cumulative error approach.

V. Results and Analysis

This section assesses the performance of the proposed intelligent blind FDI in the presence of different forms of MTD on the standard IEEE 14-Bus and IEEE 118-bus test systems [35]. All simulations were implemented using the MATPOWER toolbox in MATLAB [36] and performed using Intel Core i7-7820X CPU with 64GB of ram running on a Windows 10 system. In the graph legends, TP refers to topology perturbation (MTD via D-FACTs perturbation) and RS refers to switching MTD via circuit breaker control.

A. Model Assumptions

The priority of this section is to capture the change in detection between the blind FDI technique and the proposed intelligent attack under different types of MTD. Some assumptions have been made across all simulations:

- Uncoloured Gaussian noise error of 1% noise-to-signal was added to meter values as error **e** as seen previously in [37].
- A steady load assumption is made with load variation of around 0.1% for initial simulations as seen in [21]. Additional case studies were performed with multiple load profiles.
- A minimum number of observations of 250 is assumed initially which rises to 1000 sequentially over the course of the simulation.

B. Line Applications of MTD

In this paper, MTD is applied at the branch level in a fixed order shown in Table I to inductance values, based on a % of the branch inductance. This order is consistent between MTD type to ensure a fair comparison between the MTD performance. Number of lines perturbed NLP refers to the number of adjusted lines within a given scenario. Line adjustments are not applied simultaneously and therefore a simulation will have $NLP+1$ potential underlying topologies that a successful attack will need to model for. The NLP perturbation list is additive and the topologies within them randomly selected from within this list.

C. Transmission Switching

The first form of MTD we trial is direct use of system circuit breakers to create new topologies (transmission switching).

TABLE I
Order of Number of Lines Perturbed (NLP) Applied

Bus 1	Bus 2	R	X	NLP
1	2	0.01938	0.05917	1
1	5	0.05403	0.22304	2
2	3	0.04699	0.19797	3
2	4	0.05811	0.17632	4
2	5	0.05695	0.17388	5
3	4	0.06701	0.17103	6
4	5	0.01335	0.04211	7
4	7	0	0.20912	8
4	9	0	0.55618	9
5	6	0	0.25202	10
6	11	0.09498	0.1989	11
6	12	0.12291	0.25581	12
6	13	0.06615	0.13027	13
9	10	0.03181	0.0845	14
9	14	0.12711	0.27038	15
10	11	0.08205	0.19207	16

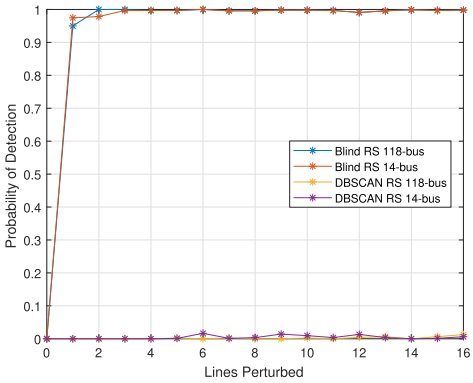

Fig. 6. Probability of detection of blind FDI attack and the new attack under transmission switching for IEEE 14-bus and 118-bus systems under 99% confidence interval. Lines are not perturbed simultaneously.

In this case, lines are switched into and out of operation to change the underlying topology incidence matrix. This creates significant changes in the overall power measurement matrix. Figure 6 shows the impact of transmission line switching on the blind FDI attack and DBSCAN attack for the 14 bus and 118 bus cases. For the standard blind FDI attack, transmission switching is highly effective at introducing residual errors and driving alarms. With a single line switching the detection is 100% for the standard blind FDI attack. However, these large changes in the system flows make it easy for an attacker to identify the MTD. Compared with the standard attack, the DBSCAN attack out-performs the standard blind FDI whenever MTD is used. Detection remained low (less than 1%) with up to 15 lines being switched in/out across the network at different times. Even with 16 possible topologies in use the detection remained under 3%. Transmission switching is unlikely to be used for the sole purpose of attack detection due to the significant impact on the system operability.

D. Admittance Perturbation

Admittance perturbation is the most commonly proposed method of MTD for power systems in the current literature. This sub-section implements an admittance perturbation

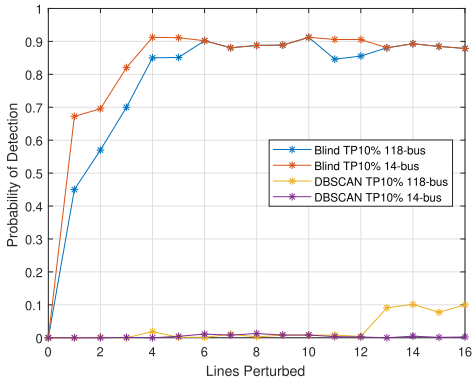

Fig. 7. Probability of detection of blind FDI attack and the new attack under admittance perturbation for IEEE 14-bus and 118-bus systems under 99% confidence interval. Lines are not perturbed simultaneously.

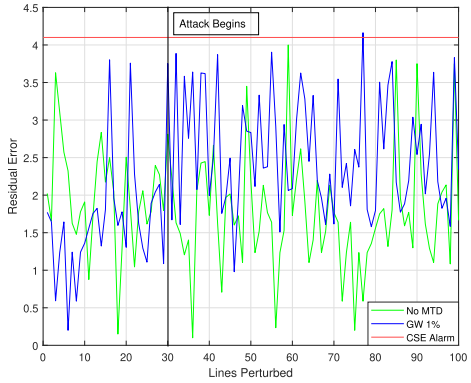

Fig. 8. Conventional CSE Residual error for run numbers on 14-bus system with the Gaussian Watermark applied to 14 lines. Bus angle change of 20 degrees attempted across the system by the FDI attack.

defense against the typical blind FDI attack and the proposed DBSCAN version. A quantity equal to 10% branch admittance is injected into the lines given by the order in I. As discussed, branch admittance is applied independently and the number of underlying topologies will be equal to $NLP + 1$. When the inductance is injected, the system operator is expecting to see the change in admittance reflected in the resulting power flows. If the attacker is unaware and does not reflect the new admittance in their attacking vector, the residual will increase significantly and BDD will be triggered. The results of admittance perturbation on detection of the standard and DBSCAN blind FDI attack are shown in Figure 7. System models with branch admittance perturbations of 10% were implemented. The standard blind FDI attack performs poorly against this form of MTD. For a single line at 10% perturbation a detection level of over 95% is achieved. The detection rates for the DBSCAN informed attack were consistently low. This is due to the distinctive clusters of power flows emerging under the steady loads assumption. There is a small spike which appears around 12 lines perturbed. This is probably due to the increasing number of lines perturbed in the system likely causing a misclustering in the underlying data-set or depriving a cluster of enough data points for a decent model.

E. Physical Gaussian Watermarking With Cumulative Errors

A novel MTD implementation is trialled here. In the same order and manner as the transmission switching and admittance perturbation sections we apply a Gaussian style physical watermark as defence. Inductance change of 1% to the system varied over a random distribution. Only one line change is applied at a time to keep it consistent with the other forms of MTD. The admittance profile is varied using a PRNG with a profile equivalent to the underlying noise of the system. As we have used a 1% noise for our simulations the p value is set equal to 1% to ensure that this profile is not visible to the attacker. This is combined with cumulative summed error monitoring watching for sustained increased errors over 10 measurements with a cumulative limit based on

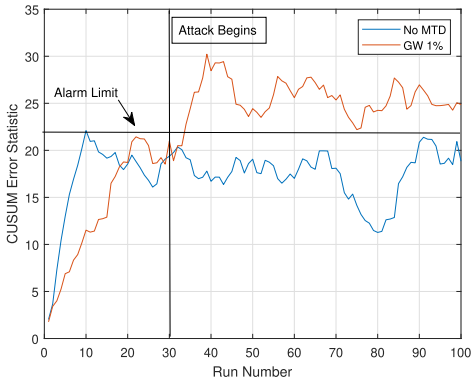

Fig. 9. CUSUM rolling summations for run 14-bus system with the Gaussian Watermark applied to 14 lines. Bus angle change of 20 degrees attempted across the system by the FDI attacker.

2 standard deviations above the average CUSUM measurement error summation under normal conditions.

Figures 8 and 9 illustrate the implementation of the Gaussian Watermark with assumption that FDI attack starts from time instance 30. Figure 8 shows the traditional CSE residual error resulting from an FDI attack in presence of the Gaussian Watermark. It is clear that the small system changes can not directly drive the detection of FDI attack in conventional residual-based BDD. Monitoring for the average of last 10 measurements allows the system operator to identify long term trends in the data, which in this case are caused by small but sustained gross errors introduced from the FDI attack. In Figure 9 the CUSUM method of detection is applied based on the last 10 measurements. Initially, we do not implement any attack for the first 30 runs of the system and we note residual CUSUM averages in line the normal value. At run 30 we introduce the attack vector. From inspection, it is much clearer that the system is under attack and a alarm is raised after 4 consecutive measurements.

As shown in Figure 10, under the DBSCAN blind FDI attack, the CUSUM Gaussian watermark shows significant improvements. As additional lines are added, these detection rates are close to 100% compared with under 10% for standard admittance perturbation. The is due to the difficulty DBSCAN algorithm has in identifying clusters for MTD on magnitude

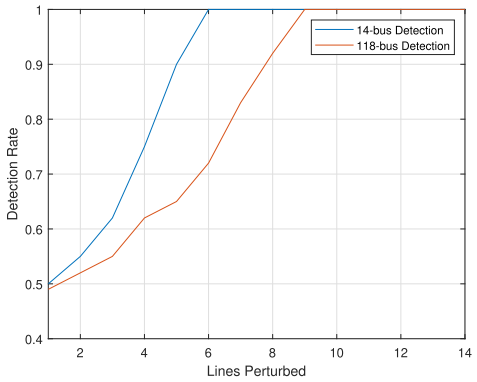

Fig. 10. DBSCAN detection results under proposed Gaussian watermark with cumulative errors over 10 measurements. 14-bus and 118-bus systems simulated with baseline 10 measurement average as detection trigger.

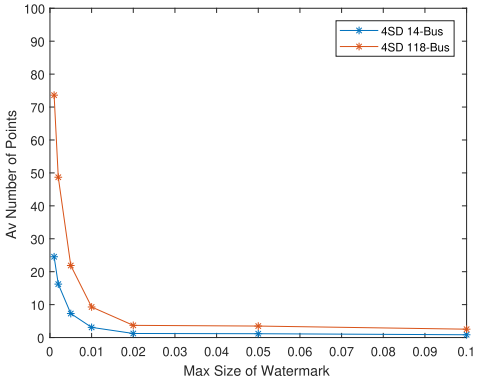

Fig. 11. Average number of points required to break a four standard deviation upper limit for increasing size of watermark applied to a single line.

and identical to noise profile of the system. Type-II error based on two standard deviation moves from the CUSUM average appears to give around 3% type-II error for this kind of measurement approach across 1000 measurements. As seen in 10, this cumulative approach also requires multiple measurements which potentially could lead to the attacker having additional time to attack before being caught. Therefore, there is a trade-off between the speed to spot attacks and the magnitude of the added watermark. Figure 11 illustrates this for the 118-bus and 14-bus systems where, for a lower level of added watermark, a larger number of measurement points are needed to break the threshold.

F. Load Variance Impact

In previous case studies, the simulations have been performed under steady load assumptions [21]. This session investigates the impact of large load variation on the performance of the DBSCAN attack. As shown in Figure 12, large load variations reduce the effectiveness of the DBSCAN under pressure from topology perturbation due to the increasing challenge to cluster the topology changes under high varying load. To circumvent this challenge, we separate differing load values into buckets based on the system profiles reduced via T-SNE. Load values are observed directly, dimension reducing them via T-SNE and assigning them to bins of similar values.

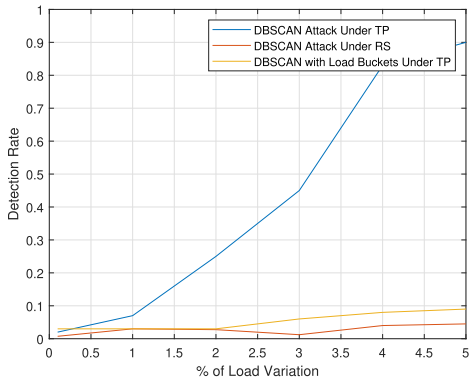

Fig. 12. Detection of DBSCAN method with 10 lines perturbed with increasing load variance. Also featured is the DBSCAN with load profile reduction analysis with load variation effectively reduced using 10 load buckets.

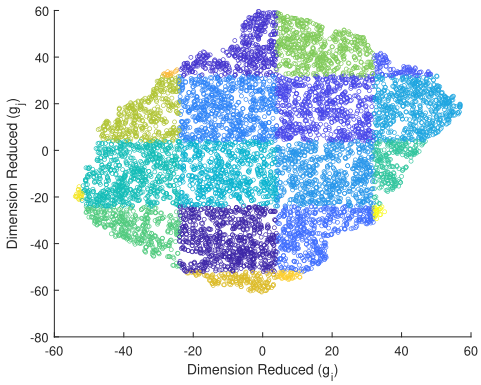

Fig. 13. Real power load profile values reduced by T-SNE. Variation of 10% shown. Different colours represent different proposed load buckets. 10k measurements.

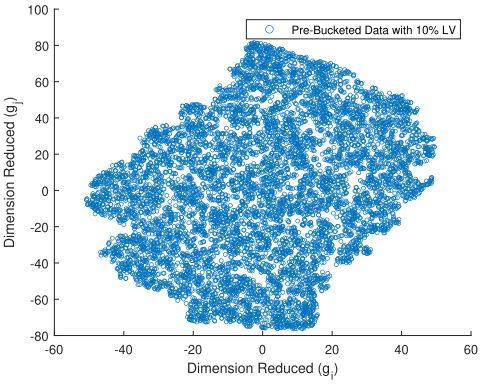

Fig. 14. Power Observations for a 10 lines perturbed system using D-FACTs of 10% under load variance of 10% shown. This is prior to bucketing of data by load profile.

In Figure 13 the profile of the loads themselves are observed, dimension reduced and bucketed. Within each load groups, measurement observations can be used to obtain the clear MTD groups to develop the attack model. Load bucketing reduces the effective load variation back down to a steady loads style scenario. Figure 14 and Figure 15 show the results of high load variation on the system under MTD with and

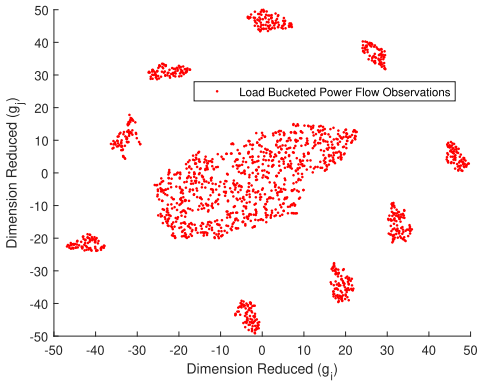

Fig. 15. Post-bucketed power flow data with T-SNE applied for a 10 lines perturbed system using D-FACTs of 10%. Original load variance of 10% was used.

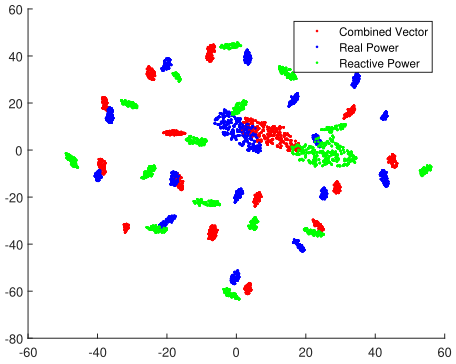

Fig. 16. Observations of 1% MTD applied to AC system up to 16 lines intermittently. Data cuts of real power, reactive power and a combined vector incorporating both are compared. 1% Gaussian noise assumed.

without load bucketing applied. It is clear, that under load bucketing, the distinct groups of measurement observations become clearer as a result of MTD. Applying this bucketing reduces the effective variance at the power flow significantly from 10% to under 1% with around 25 buckets for a 14-bus system. This effectively replicates the steady loads assumption even in the case of a more variate system. As the system variance becomes larger, additional bucket can be added to accommodate the larger variance of the system. The effect of this can be seen in Figure 12 with lowered detection for the DBSCAN attack when implemented against topology perturbation style defence. Bucketing in this manner will require a large amount of data, however based on the frequency of measurement at around 2-5s [38] and the lengthy attack development phase [1], such data requirement can be easily satisfied. Blind attacks will always require larger past data requirement than full knowledge attacks as they need to build a model, unlike the full knowledge attacks which already possess the model.

G. Blind AC Replay Attack

We have implemented our clustering approach with a blind replay style attack against an AC state estimation. Under this attack, the attacker attempts to inject a previously observed vector. The attacker is competing with MTD and

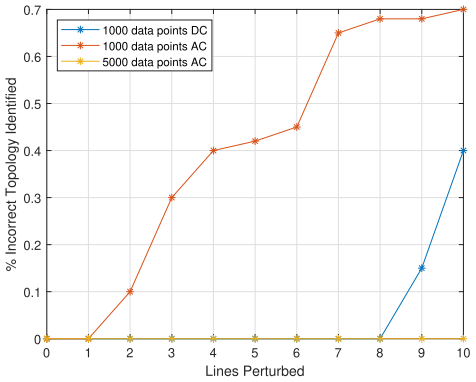

Fig. 17. AC System probability of wrong cluster identified for in presence of D-FACTs MTD with increasing lines perturbed to 14-bus system.

wants to select the replay vector from a pool of values only containing those using the same topology configuration. In Figure 16 we can see that the distinctive cluster relationship exists within the AC model as shown previously for DC. Figure 17 demonstrates that the proposed pre-clustering algorithm performs well in AC state estimation provided a large number of samples are received. The non-linearity in the AC model significantly reduces the correction rate of clustering but increasing the number of observations allows good performance for the AC model.

VI. Conclusions & Further Work

This paper, for the first time, investigates how unsupervised learning and dimensionality reduction can be applied in blind FDI attacks to exploit the vulnerability of current forms of MTD. By incorporating a combination of T-SNE dimensionality reduction and the DBSCAN clustering algorithm, power flow observations can be clustered into their relative topology profiles and the mixing matrix for the blind FDI attack can be calculated using only data under the same network topology. This technique is shown to be effective against admittance perturbation and transmission switching techniques. A novel defense strategy against this new type attack is proposed through combining MTD with physical watermarking to add an indistinguishable Gaussian style physical watermark into the plant topology and monitoring the sequential errors for long-run trends by using CUSUM. This technique is demonstrated to be effective at both inhibiting the attacker's ability to predict topological changes from visible power flows and reducing the overall impact on system operation by reducing the level of topology changes.

Further work on this topic entails enhancing the blind FDI model to model for other scenarios i.e. subset attacks, optimal design of physical watermarking scheme and analysing the effects of MTD on topological discovery techniques.

References

[1] G. Liang, S. R. Weller, J. Zhao, F. Luo, and Z. Y. Dong, "The 2015 ukraine blackout: Implications for false data injection attacks," *IEEE Trans. Power Syst.*, vol. 32, no. 4, pp. 3317–3318, Jul. 2017.

[2] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Trans. Inf. Syst. Secur.*, vol. 14, no. 1, pp. 1–33, May 2011.

[3] X. Liu, Z. Li, X. Liu, and Z. Li, "Masking transmission line outages via false data injection attacks," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 7, pp. 1592–1602, Jul. 2016.

[4] R. Deng, G. Xiao, R. Lu, H. Liang, and A. V. Vasilakos, "False data injection on state estimation in power Systems—Attacks, impacts, and defense: A survey," *IEEE Trans. Ind. Informat.*, vol. 13, no. 2, pp. 411–423, Apr. 2017.

[5] G. Liang, J. Zhao, F. Luo, S. R. Weller, and Z. Yang Dong, "A review of false data injection attacks against modern power systems," *IEEE Trans. Smart Grid*, vol. 8, no. 4, pp. 1630–1638, Jul. 2017.

[6] M. A. Rahman and H. Mohsenian-Rad, "False data injection attacks with incomplete information against smart power grids," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2012, pp. 3153–3158.

[7] M. Esmalifalak, H. Nguyen, R. Zheng, and Z. Han, "Stealth false data injection using independent component analysis in smart grid," in *Proc. IEEE Int. Conf. Smart Grid Commun. (SmartGridComm)*, Oct. 2011, pp. 244–248.

[8] R. Deng and H. Liang, "False data injection attacks with limited susceptance information and new countermeasures in smart grid," *IEEE Trans. Ind. Informat.*, vol. 15, no. 3, pp. 1619–1628, Mar. 2019.

[9] M. N. Kurt, Y. Yilmaz, and X. Wang, "Real-time detection of hybrid and stealthy cyber-attacks in smart grid," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 2, pp. 498–513, Feb. 2019.

[10] Y. Wang, M. M. Amin, J. Fu, and H. B. Moussa, "A novel data analytical approach for false data injection cyber-physical attack mitigation in smart grids," *IEEE Access*, vol. 5, pp. 26022–26033, 2017.

[11] S. Ahmed, Y. Lee, S.-H. Hyun, and I. Koo, "Unsupervised machine learning-based detection of covert data integrity assault in smart grid networks utilizing isolation forest," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 10, pp. 2765–2777, Oct. 2019.

[12] F. Wen and W. Liu, "An efficient data-driven false data injection attack in smart grids," in *Proc. IEEE 23rd Int. Conf. Digit. Signal Process. (DSP)*, Nov. 2018, pp. 1–5.

[13] 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, Mar. 2015.

[14] J. Hao, R. J. Piechocki, D. Kaleshi, W. H. Chin, and Z. Fan, "Sparse malicious false data injection attacks and defense mechanisms in smart grids," *IEEE Trans. Ind. Informat.*, vol. 11, no. 5, pp. 1–12, Oct. 2015.

[15] J. Zhang, Z. Chu, L. Sankar, and O. Kosut, "Can attackers with limited information exploit historical data to mount successful false data injection attacks on power systems?" *IEEE Trans. Power Syst.*, vol. 33, no. 5, pp. 4775–4786, Sep. 2018.

[16] S. Wang, W. Ren, and U. M. Al-Saggaf, "Effects of switching network topologies on stealthy false data injection attacks against state estimation in power networks," *IEEE Syst. J.*, vol. 11, no. 4, pp. 2640–2651, Dec. 2017.

[17] K. L. Morrow, E. Heine, K. M. Rogers, R. B. Bobba, and T. J. Overbye, "Topology perturbation for detecting malicious data injection," in *Proc. 45th Hawaii Int. Conf. Syst. Sci.*, Jan. 2012, pp. 2104–2113.

[18] C. Liu *et al.*, "Reactance perturbation for enhancing detection of FDI attacks in power system state estimation," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Nov. 2017, pp. 523–527.

[19] 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 Trans. Inf. Forensics Security*, vol. 15, pp. 2320–2335, 2020.

[20] B. Li, G. Xiao, R. Lu, R. Deng, and H. Bao, "On feasibility and limitations of detecting false data injection attacks on power grid state estimation using D-FACTS devices," *IEEE Trans. Ind. Informat.*, vol. 16, no. 2, pp. 854–864, Feb. 2020.

[21] J. Tian, R. Tan, X. Guan, and T. Liu, "Enhanced hidden moving target defense in smart grids," *IEEE Trans. Smart Grid*, vol. 10, no. 2, pp. 2208–2223, Mar. 2019.

[22] Y. Mo, S. Weerakkody, and B. Sinopoli, "Physical authentication of control systems: Designing watermarked control inputs to detect counterfeit sensor outputs," *IEEE Control Syst.*, vol. 35, no. 1, pp. 93–109, Feb. 2015.

[23] A. Monticelli, *State Estimation Electric Power Systems: A Generalized Approach*. Boston, MA, USA: Kluwer, 1999.

[24] Z.-H. Yu and W.-L. Chin, "Blind false data injection attack using PCA approximation method in smart grid," *IEEE Trans. Smart Grid*, vol. 6, no. 3, pp. 1219–1226, May 2015.

[25] W.-L. Chin, C.-H. Lee, and T. Jiang, "Blind false data attacks against AC state estimation based on geometric approach in smart grid communications," *IEEE Trans. Smart Grid*, vol. 9, no. 6, pp. 6298–6306, Nov. 2018.

[26] L. J. P. Van der Maaten and G. E. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.

[27] D. G. E. Silva, M. Jino, and B. T. D. Abreu, "Machine learning methods and asymmetric cost function to estimate execution effort of software testing," in *Proc. 3rd Int. Conf. Softw. Test., Verification Validation*, 2010, pp. 275–284.

[28] N. Pezzotti, B. P. F. Lelieveldt, L. V. D. Maaten, T. Hollt, E. Eisemann, and A. Vilanova, "Approximated and user steerable tSNE for progressive visual analytics," *IEEE Trans. Vis. Comput. Graphics*, vol. 23, no. 7, pp. 1739–1752, Jul. 2017.

[29] L. McInnes, J. Healy, and S. Astels. *Benchmarking Performance and Scaling of Python Clustering Algorithms*. Accessed: Sep. 29, 2020. [Online]. Available: https://hdbscan.readthedocs.io/en/latest/

[30] M. Ester, H. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. 2nd Int. Conf. Knowl. Discovery*, 1996, pp. 226–231.

[31] C. Konstantinou and M. Maniatakos, "A data-based detection method against false data injection attacks," *IEEE Des. Test*, early access, Nov. 8, 2019, doi: 10.1109/MDAT.2019.2952357.

[32] M. Alshawabkeh, B. Jang, and D. Kaeli, "Accelerating the local outlier factor algorithm on a GPU for intrusion detection systems," in *Proc. 3rd Workshop General-Purpose Comput. Graph. Process. Units (GPGPU)*, 2010, p. 124.

[33] J. Aghaei, M. Gitizadeh, and M. Kaji, "Placement and operation strategy of FACTS devices using optimal continuous power flow," *Scientia Iranica*, vol. 19, no. 6, pp. 1683–1690, Dec. 2012.

[34] S. Lakshminarayana and D. K. Y. Yau, "Cost-benefit analysis of moving-target defense in power grids," *IEEE Trans. Power Syst.*, early access, Jul. 20, 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9144465/, doi: 10.1109/TPWRS.2020.3010365.

[35] R. Christie. *Power Systems Test Case Archive*. Accessed: Sep. 29, 2020. [Online]. Available: http://labs.ece.uw.edu/pstca/

[36] 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.

[37] K. Khanna, B. K. Panigrahi, and A. Joshi, "AI-based approach to identify compromised meters in data integrity attacks on smart grid," *IET Gener., Transmiss. Distrib.*, vol. 12, no. 5, pp. 1052–1066, 2018.

[38] K. C. Budka, J. G. Deshpande, and M. Thottan, *Communication Networks for Smart Grids*. London, U.K.: Springer, 2014, pp. 148–169.

Martin Higgins (Graduate Student Member, IEEE) received the B.Sc. degree in physics from the Queen Mary, University of London, in 2011, and the M.Sc. degree from Imperial College London, U.K., in 2012, where he is currently pursuing the Ph.D. degree in electrical engineering as part of the CDT in Smart Grids collaboration integrated MRES and PHD with the University of Strathclyde. His research interests lie in power systems cyber-security, false data injection attacks, and moving target defense.

Fei Teng (Member, IEEE) received the B.Eng. degree from Beihang University, China, in 2009, and the Ph.D. degree from Imperial College London, in 2015. He is currently a Lecturer with the Department of Electrical and Electronic Engineering, Imperial College London, U.K. His research focus on the efficient and resilient operation of future cyber-physical power systems.

Thomas Parisini (Fellow, IEEE) received the Ph.D. degree in electronic engineering and computer science in 1993 from the University of Genoa. He was with Politecnico di Milano and since 2010 he holds the Chair of Industrial Control and is Director of Research at Imperial College London. He is a Deputy Director of the KIOS Research and Innovation Centre of Excellence, University of Cyprus. Since 2001 he is also Danieli Endowed Chair of Automation Engineering with University of Trieste. In 2009–2012, he was Deputy Rector of University of Trieste. In 2018, he received an *Honorary Doctorate* from University of Aalborg, Denmark. He authored or coauthored more than 320 research papers in archival journals, book chapters, and international conference proceedings. His research interests include neural-network approximations for optimal control problems, distributed methods for cyber-attack detection and cyber-secure control of large-scale systems, fault diagnosis for nonlinear and distributed systems, nonlinear model predictive control systems and nonlinear estimation. He is a co-recipient of the IFAC Best Application Paper Prize of the *Journal of Process Control*, Elsevier, for the three-year period 2011–2013 and of the 2004 Outstanding Paper Award of the IEEE TRANSACTIONS ON NEURAL NETWORKS. He is also a recipient of the 2007 IEEE Distinguished Member Award. In 2016, he was awarded as Principal Investigator at Imperial of the H2020 European Union flagship Teaming Project KIOS Research and Innovation Centre of Excellence led by University of Cyprus. In 2012, he was awarded an ABB Research Grant dealing with energy-autonomous sensor networks for self-monitoring industrial environments. He currently serves as 2020 President-Elect of the IEEE Control Systems Society and will serve as President during 2021–2022. He has served as Vice-President for Publications Activities and during 2009–2016 he was the Editor-in-Chief of the IEEE TRANSACTIONS ON CONTROL SYSTEMS TECHNOLOGY. Since 2017, he is Editor for Control Applications of *Automatica* and since 2018 he is the Editor-in-Chief of the *European Journal of Control*. He is also the Chair of the IFAC Technical Committee on Fault Detection, Supervision & Safety of Technical Processes-SAFEPROCESS. He was the Chair of the IEEE Control Systems Society Conference Editorial Board and a Distinguished Lecturer of the IEEE Control Systems Society. He was an elected member of the Board of Governors of the IEEE Control Systems Society and of the European Control Association (EUCA) and a member of the board of evaluators of the 7th Framework ICT Research Program of the European Union. He is currently serving as an Associate Editor of the *International Journal of Control* and served as Associate Editor of the IEEE TRANSACTIONS ON AUTOMATIC CONTROL, of the IEEE TRANSACTIONS ON NEURAL NETWORKS, of *Automatica*, and of the *International Journal of Robust and Nonlinear Control*. Among other activities, he was the Program Chair of the 2008 IEEE Conference on Decision and Control and General Co-Chair of the 2013 IEEE Conference on Decision and Control. Prof. Parisini is a fellow of the IFAC.