A Deep Learning Approach to Anomaly Sequence Detection for High-Resolution Monitoring of Power Systems

Kursat Rasim Mestav, Xinyi Wang, *Student Member, IEEE,* and Lang Tong, *Fellow, IEEE*

***Abstract*—A deep learning approach is proposed to detect data and system anomalies using high-resolution continuous point-on-wave (CPOW) or phasor measurements. Both the anomaly and anomaly-free measurement models are assumed to have unknown temporal dependencies and probability distributions. Historical training samples are assumed for the anomaly-free model, while no training samples are available for the anomaly measurements. By transforming the anomaly-free observations into uniform independent and identically distributed sequences via a generative adversarial network, the proposed approach deploys a uniformity test for anomaly detection at the sensor level. A distributed detection scheme that combines sensor level detections at the control center is also proposed which combines local detections to form more reliable detections. Numerical results demonstrate significant improvement over the state-of-the-art solutions for various bad-data cases using real and synthetic CPOW and PMU data sets.**

***Index Terms*—System event detection, continuous point-on-wave (CPOW) measurements, bad-data detection, distributed anomaly detection, generative adversary networks (GAN).**

I. Introduction

We consider the problem of detecting data and system anomalies using possibly unsynchronized high-resolution power system measurements. Besides conventional synchrophasor measurements, we consider continuous point-on-wave (CPOW) measurements sampled at up to 100 kHz. At these sampling rates, power system measurements exhibit strong temporal dependencies.

High-resolution measurements currently exist in the field in various monitoring devices [1], mostly used for local protection purposes and also for post-event analysis. Rarely they are streamed to the control center for real-time monitoring. However, with the increasing penetration of inverter-based resources, there are cogent needs for high-resolution monitoring that goes beyond using low-resolution SCADA and PMU based measurements [1]–[3]. To this end, we aim to fill a theoretical and practical gap in using high-resolution measurements to detect anomalies at the sensor level and combine sensor-level detections at the control center to form more accurate anomaly detections.

The challenge of anomaly detection with high-resolution measurements is threefold. First, temporal dependencies in high-resolution measurements are difficult to model. Conventional techniques based on sample-by-sample detection or assuming independent data samples tend to perform poorly. In this work, we stress the significance of *anomaly sequence detection* where anomaly detection is made based on a measurement sequence rather than individual samples.

Second, anomalies are rare events, and there are uncountably many possibilities that anomalies can occur; no single model nor sufficient historical data are available to characterize and validate anomaly data. Therefore, anomaly detection should be derived from the anomaly-free model only, independent of the types of anomalies that may occur.

Finally, defining "normalcy" is nontrivial. While a power system has well-defined nominal operating conditions, it has frequent transients as generators are dispatched in real-time. There is no standard data model that leads to well-defined statistical tests for anomaly-free data.

We consider three types of anomalies. One is the conventional bad data caused by malfunctioning sensors and communication errors that generate outliers. The second is data anomaly from data attacks, where an adversary manipulates sensor data to affect the operator's decision process. The third is system anomalies such as faults and operation contingencies that manifest themselves in data. We do not distinguish among different data anomalies.

A. Related work

The classic anomaly detection in wide-area power system monitoring is bad-data detection in the context of power system state estimation [4]–[6]. A standard approach is *post-estimation bad-data detections* where state estimation is performed first as if there were no bad data. Data anomaly is declared when the (normalized) residue error (computed using the estimated state) is greater than a certain threshold. As a result, inaccuracy of state estimation caused by anomaly data circulates back to affect bad-data detection.

An alternative is the *pre-estimation bad-data detection* techniques that detect data anomaly before state estimation, thus breaking the path of estimation error propagation. The key idea is to replace the estimated state in the post-estimation scheme with a *predicted state* using the past mea-

Kursat Rasim Mestav (krm264@cornell.edu), Xinyi Wang (xw555@cornell.edu), and Lang Tong (lt35@cornell.edu) are with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853, USA.

This work is supported in part by the National Science Foundation under Award 1809830, Award 1816397, Award 1932501, and, Power Systems and Engineering Research Center (PSERC)

This work is supported in part by NVIDIA GPU grant program. We thank NVIDIA for giving us Titan XP GPU as a grant to carry out our work in deep learning.

surements and apply residue test on the predicted measurement [7], [8]. Such techniques exploit temporal dependencies in the data for prediction, thus more relevant to the anomaly sequence detection problem considered in this paper. These techniques, however, assume specific temporal dependency models that are difficult to obtain.

A more direct pre-estimation approach is to detect anomalies based on features of anomaly-free data. One of the earliest such techniques is using a neural network classifier trained with anomaly-free data [9]. A separate line of approaches is to extract features from the anomaly-free data and classify data in the feature space. Examples include the use of principal component analysis to characterize the signal subspace of the anomaly-free data [10] and the formulation of the problem as the detection of a change in measurement probability distribution [11].

There is a growing literature on the use of machine learning for bad-data detection in power systems since the mid-1990s [9]. These techniques can be categorized based on how data are used in learning. Supervised learning requires labeled training data in both anomaly-free and anomaly cases [12], semi-supervised learning requires training samples for the anomaly-free data [13], [14], and unsupervised learning requires no training data [11]. An ensemble learning technique is proposed in [15] that combines a collection of bad-data detectors.

Because it is difficult to obtain labeled anomaly data for training, the semi-supervised and unsupervised learning paradigms are of particular significance. Although not designed for power system state estimation, two types of semi-supervised anomaly detectors that use only training samples under the anomaly-free model can be applied for bad-data detection in power systems. One is the one-class support vector machine OC-SVM [16] that separates anomaly and anomaly-free data deterministically. The other is based on the idea of auto-encoder in deep neural network [17]. An implicit assumption is that anomaly and anomaly-free data do not share a common data domain for these methods, which rarely holds in power system measurement models.

Statistical learning approaches to anomaly detection start from the premise that anomaly and anomaly-free data come from different probability distributions. To this end, a recent work of particular relevance is [13] that focuses on dynamic data attacks of power system state estimation. Although the attack models in [13] suggests an anomaly sequence detection problem, the proposed mitigation strategy is a sample-by-sample detection scheme based on anomaly-free probability distributions from historical samples. The idea of universal bad-data detection methods developed in [14] is a semi-supervised learning technique that learns the inverse generative model of the anomaly-free data using a generative adversarial network (GAN) approach using Wasserstein distance [18], followed by a coincidence test. The approach developed in [14] relies on that the observations are i.i.d., which is unreasonable for high-resolution data.

There is significant literature on detecting the so-called *data injection attacks* by an adversary who can inject, remove, and substitute data to affect system and market operations [19]–[24]. In particular, an attacker may create a fake sequence of system states such that the manipulated measurements and the fake state sequence satisfy the underlying power flow equation, which makes the manipulated data *unobservable.* There is no effective anomaly detection solution for such attacks in the literature. The technique proposed here gives a viable solution.

B. Summary of approach and contributions

We develop a data-driven machine learning technique to detect anomalies from high-resolution CPOW and PMU measurements. By stressing the significance of *anomaly sequence detection*, the proposed approach is a notable departure from the conventional sample-by-sample detection solutions, and it is perhaps the first anomaly sequence detection method for power system monitoring.

The main technical contribution of this work is twofold. First, we develop a sensor-level non-parametric anomaly sequence detection method in which no assumptions are made for the anomaly data model. The anomaly-free model is also assumed to be unknown, except that historical training samples are available, making the proposed technique a data-driven solution. By not assuming any anomaly model, the proposed sensor-level detection applies to bad-data anomalies, data injection attacks, and system anomalies that manifest themselves in anomaly data patterns. To our best knowledge, there is no existing alternative in the power system monitoring literature.

A significant challenge of anomaly sequence detection is the unknown temporal dependencies in measurements. To this end, we propose a GAN-based independent component analysis, referred to as ICA-GAN, that transforms anomaly-free measurements with unknown statistical dependencies to uniform independent and identically distributed (*i.i.d.*) samples. We then apply a uniformity test that distinguishes uniform *i.i.d.* samples from the anomaly-free hypothesis from non *i.i.d.* and/or non-uniform anomaly samples. While ICA [25], [26] and uniformity tests [27]–[29] have been developed separately in the past, a combination of them for anomaly-detection is a novel contribution.

Second, we propose a distributed detection framework that combines sensor-level detections for system-level anomaly detection. Such techniques are essential because individual sensors have access to local measurements only, and their detections are likely to be unreliable. Distributed detection plays crucial roles in various surveillance applications and has been studied extensively [30]. Classic techniques require known anomaly and anomaly-free probability models, and measurement samples are assumed to be conditionally *i.i.d.* These assumptions do not apply to the anomaly detection

model considered here. To our best knowledge, there are no effective non-parametric decentralized techniques in the literature. The proposed technique exploits a key feature of ICA-GAN that, under the anomaly-free hypothesis, maps dependent measurement sequences to uniform *i.i.d.* sequences, making it possible that local detectors with the same false positive rate (FPR) are used. To this end, we derive a fusion rule that combines individual sensor decisions.

Finally, we test the proposed technique under three anomaly scenarios, using real data set from the EPFL network [31], [32] and a larger synthetic Northern Texas network with PMU measurements [33]. These illustrations cover a natural data anomaly, an unobservable data-injection attack, and a system anomaly. They serve as demonstrations of the versatility of the proposed detection method.

II. System and Anomaly Detection Model

A. Measurement and anomaly models

The proposed anomaly detection solution applies to measurements involve a single sensor at a remote terminal or a group of possibly unsynchronized sensors. Let the measurement sequence* at sensor i be (z_{it}), which we model as a random process generated from power system state sequence (x_t), additive noise (w_{it}), and anomaly sequence (a_{it}):

$$z_{it} = h_i(x_t) + w_{it} + a_{it}, \tag{1}$$

where the measurement function $h_i(\cdot)$ at sensor i encodes system parameters and topology information is assumed unknown. Herein, we make the assumption that noise processes (w_{it}) at different sensors are statistically independent whereas the anomaly sequences (a_{it}) may be dependent.

For natural data anomalies, (a_{it}) are assumed to be independent of $(h_i(x_{it}))$, ambient noise (w_{it}) and measurements elsewhere (z_{jt}). For adversarial data anomalies, very little can be assumed about a_{it}. In particular, a_{it} may be a function of past measurements and statistically dependent on the system state in some arbitrary fashion. An extreme type of unobservable attack can be constructed in the form of $a_{it} = h_i(x'_t) - h_i(x_t)$ where the adversary substitute the actual system measurement $h_i(x_t)$ by a fictitious measurement corresponding to a fictitious state x'. For system anomalies with post-contingency measurement function $h'_i(\cdot)$, the anomaly sequence can simply be $a_{it} = h'_i(x_t) - h_i(x_t)$.

B. Sensor level anomaly sequence detection

At the sensor level, we formulate the anomaly sequence detection problem as a non-parametric hypothesis testing of a time series. We assume that at time t, we have a block of M of current and past measurements $Z_{it} = (z_{it}, z_{i(t-1)}, \cdots, z_{i(t-M+1)})$. Let the null hypothesis $\mathcal{H}_0$ model the anomaly-free data and the alternative $\mathcal{H}_1$ for the anomaly data. In particular,

$$\begin{aligned} &\mathcal{H}_{i0} : Z_{it} \sim f_{i0} \quad \text{vs.} \quad \mathcal{H}_{i1} : Z_{it} \sim f_{i1} \in \mathscr{F}_{i,\epsilon} \\ &\mathscr{F}_{i,\epsilon} := \{f, ||f - f_{i0}|| \geq \epsilon\} \end{aligned} \tag{2}$$

where f_{i0} and f_{i1} are the underlying joint probability distributions of Z_{it} under $\mathcal{H}_{i0}$ and $\mathcal{H}_{i1}$, respectively. Note that $\mathcal{H}_{i0}$ is a simple hypothesis with a single probability distribution f_{i0} and $\mathcal{H}_{i1}$ a *composite hypothesis* with a set $\mathscr{F}_\epsilon$ of distributions some ϵ distant away†. The requirement of ϵ separation of the null and the alternative hypothesis is to ensure consistency of the detector.

Under (2), each sensor makes an individual binary decision $u_{it} = \mathcal{D}_i(Z_{it}) \in \{0, 1\}$ on anomaly based on Z_{it}: $u_{it} = 1$ means that the detector rejects the null (anomaly-free) hypothesis $\mathcal{H}_{i0}$, and $u_{it} = 0$ means that the null hypothesis $\mathcal{H}_{i0}$ is accepted. In practice, a sensor produces a detection every M samples when non over-lapping blocks are used. The size of M has both theoretical and practical implications. A larger M means better detection reliability with considerably higher complexity in learning and implementation.

C. Distributed detection and data fusion model

We now consider a power system with K local PMU/CPOW sensors as shown in Fig. 1, where sensor i produces a local binary sensor detection u_{it} at time t. We assume that local decisions $\{u_{it}\}$ are communicated synchronously to the control center (fusion center) where a global decision ν_t on anomaly is made. Let $u_t = (u_{1t}, \cdots, u_{Kt})$ be the local decision vector at the control center.

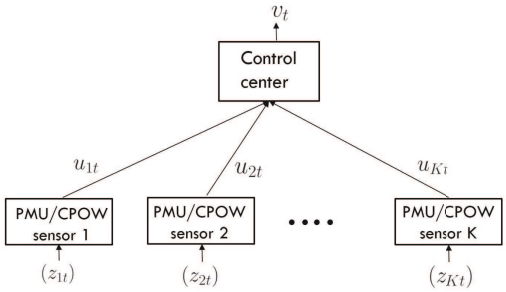

Fig. 1: A schematic of a distributed anomaly detection system.

Assume that the local detector at sensor i has false positive rate (FPR) of α_i and true positive rates of β_i, the control center faces the following binary hypothesis testing problem $\mathcal{H}_0$ vs. $\mathcal{H}_1$ where $\mathcal{H}_0$ corresponds to the anomaly-free hypothesis and $\mathcal{H}_1$ the anomaly:

$$\mathcal{H}_0 : u_t \sim P_0 \text{ vs. } \mathcal{H}_1 : u_t \sim P_1, \tag{3}$$

*We adopt the standard notation that (x_t) denotes a sequence of measurements.

†The distance measure of probability distributions can be arbitrary. Examples include the total variation and Jensen-Shannon distances.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPWRS.2022.3168529, IEEE Transactions on Power Systems

where $u_{it} \sim \mathcal{B}(\alpha_i)$ is Bernoulii[‡] with parameter α_i under $\mathcal{H}_0$ and $u_{it} \sim \mathcal{B}(\beta_i)$ under $\mathcal{H}_1$. In (3), P_0, P_1 are joint probability mass functions of u_t under $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. Note that, while the FPR α_i of is a design parameter that can be controlled, the TPR β_i is unknown, and it varies with the realized anomaly.

III. Sensor-Level Anomaly Sequence Detection

We now focus on the anomaly detection problem at a particular sensor. For brevity, we drop the sensor index i in the subscripts of relevant variables.

Fig. 2 shows a schematic of the proposed technique, which includes an independent component analysis (ICA) preprocessing $\mathcal{G}_\theta$ and a uniformity test. At time t, a vector consisting of M measurements $Z_t = (z_t, z_{t-1}, \cdots, z_{t-(M-1)})$ is passed through a neural network trained to extract a block of uniformly distributed independent components $V_t = (v_{t,1}, \cdots, v_{t,N})$ under the anomaly-free model (2). The training of ICA-GAN is discussed in Sec III-B, where either an offline or online training using past anomaly-free measurements can be used.

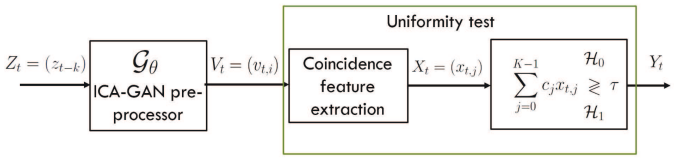

Fig. 2: A schematics of ICA-GAN for anomaly sequence detection.

The uniformity test takes the output of the ICA preprocessor $\mathcal{G}_\theta$ and produces a N-dimensional coincidence feature $X_t = (x_{t,0}, \cdots, x_{t,N})$ followed by a linear classifier whose output Y_t lables the input sequence Z_t as anomaly ($Y_t = 1$) or anomaly-free ($Y_t = 0$).

The implementations of the ICA preprocessing and uniformity test are described next.

A. Anomaly sequence detection via uniformity test

We begin with the uniformity test for anomaly sequence detection, assuming that the preprocessing step has generated V_t that, under the anomaly-free hypothesis $\mathcal{H}_0$, contains *i.i.d.* uniformly distributed random variables $(v_{t,i})$ in $[0, 1]$.

To derive the detection feature vector X_t, we first quantize $v_{t,i}$ uniformly into a discrete random variables $\bar{v}_{t,i}$ of alphabet size L, *i.e.,*

$$\bar{v}_{t,i} = l \quad \text{if } v_{t,i} \in \mathscr{B}_l = [l/L, (l+1)/L),$$

[‡]The Bernoulli random variable $X \sim \mathcal{B}(p)$ is defined here by $\Pr(X = 1) = p$ and $\Pr(X = 0) = 1 - p$.

where we refer $\mathscr{B}_k$ to as the *k-th quantization bin.* Such a quantization transforms the original anomaly detection problem (2) to the classical uniformity test defined as

$$\begin{array}{ll} \mathcal{H}_0': & (\bar{v}_{t,i}) \overset{\text{i.i.d.}}{\sim} P_0' = (\frac{1}{L}, \cdots, \frac{1}{L}), \\ \mathcal{H}_1': & (\bar{v}_{t,i}) \sim P_1' \in \mathscr{F}_{\epsilon'}, \end{array} \tag{4}$$

where $\mathscr{F}_{\epsilon'} = \{p = (p_1, \cdots, p_L) |\ ||p - P_0'|| > \epsilon\}$. Note that the probability distribution under $\mathcal{H}_0$ above is unknown whereas P_0' in (4) is known.

Following the classic work of David [27] and Viktorova and Chistyakov [28], we define a N-dimensional detection feature vector $X_t = (x_{t,0}, \cdots, x_{t,N})$ where $x_{t,k}$ is the number of quantization bins that have exactly k samples of $(\bar{v}_{t,1}, \cdots, \bar{v}_{t,N})$. In particular, $x_{t,0}$ is the number of quantization bins that contains no samples of $(\bar{v}_{t,i})$ and $x_{t,1}$ the number of quantization bins containing one sample.

With the feature vector X_t, a linear anomaly detector for hypothesis testing (2) is given by

$$\sum_{k=0}^{N} c_k x_{t,k} \underset{\mathcal{H}_1'}{\overset{\mathcal{H}_0'}{\gtrless}} T_\alpha, \tag{5}$$

where T_α is the threshold that controls the level of false positive detection rate.

In [29], Paninski shows that the above detector is consistent when only $x_{t,0}$ is used ($c_k = 0, k > 1$) so long as N grows faster than $\sqrt{L}$ as $N = o(\frac{1}{\epsilon^4}\sqrt{L})$. Remarkably, the sample complexity can be significantly less than the size of the alphabet. When the coefficients of the linear detector is carefully chosen as in [28], the detector in (5) is asymptotically most powerful.

The threshold of the test statistics affects the true and false-positive probabilities of the detection. The threshold T_α of the $x_{t,1}$ coincidence test with the constraint on the false-positive probability to no greater than α is given by

$$T_\alpha = \min\{k : \Pr(x_{t,1} \le k; \mathcal{H}_0) \le \alpha\}. \tag{6}$$

The computation of T_α amounts to evaluating $P_l := \Pr(x_{t,1} = l; \mathcal{H}_0)$, which was given by Von Mises in [34]:

$$P_l = \sum_{j=l}^{L} (-1)^{j+l} \binom{j}{l} \binom{L}{j} \frac{N!}{(N-j)!} \frac{(L-j)^{N-j}}{L^N}.$$

B. Extracting Independent Components via ICA-GAN

Independent Component Analysis (ICA), a generalization of Principle Component Analysis (PCA), extracts a set of *independent components* from a block of measurements. Originally proposed by Jutten and Herault [35] and Comon [25], ICA has found a wide range of applications when statistical independence is essential in learning and inference tasks. ICA typically requires nonlinear processing, and neural network techniques have been proposed [36]. More recently, Brackel and Bangio introduced a deep learning solution

based on GAN [26]. However, their technique that enforces independence through resampling does not perform well in time series. Here we propose an alternative solution, referred to as ICA-GAN, based on direct minimization of Wasserstein distance [18] between the distribution of ICA estimates and that of uniform *i.i.d.* random variables.

Assume that the measurement vector has a nonlinear ICA representation,

$$Z_t = f(\tilde{V}_t), \tag{7}$$

where $\tilde{V}_t = (\tilde{v}_{t,1}, \cdots, \tilde{v}_{t,N})$ has uniform *i.i.d.* components $\tilde{v}_{t,i} \overset{\text{i.i.d.}}{\sim} \mathcal{U}(0,1)$. The proposed ICA-GAN produces a minimum Wasserstein-distance estimate V_t of $\tilde{V}_t$.

The learning structure of ICA-GAN, shown in Fig 3, is an *inverse GAN*[§], where the ICA-GAN neural network $\mathcal{G}_\theta$ with weights vector θ, once properly trained, maps a sequence of arbitrary distributed random variables to a uniform *i.i.d.* sequence. A discriminator neural network $\mathcal{D}_\eta$ with weights vector η, through a dual optimization, computes the estimated gradient of Wasserstein distance between the distribution of the estimated ICA V_t and that of $\tilde{V}_t$ with uniform *i.i.d.* components. The stochastic gradient of the Wasserstein distance is used to update generator neural network coefficient θ and discriminator neural network coefficients η. See an implementation of ICA-GAN in Algorithm 1.

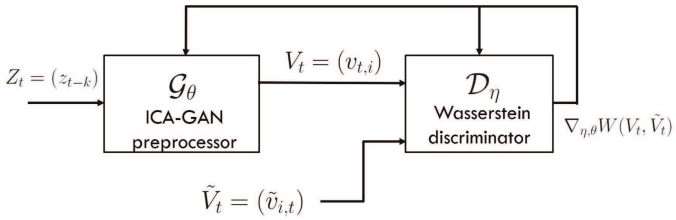

Fig. 3: Learning structure of ICA-GAN.

Ideally, if an ICA representation of the measurement exists, and the training of ICA-GAN preprocessor converges, $\mathcal{G}_\theta$ transforms the unknown measurement distribution under $\mathcal{H}_{i0}$ in (2) to the uniform *i.i.d.* distribution in V_t. In practice, however, ICA-GAN is trained with historical data samples from the anomaly-free model. With a sufficiently high-dimensional implementation and adequate training, we expect approximately uniform *i.i.d.* entries of vector V_t. See discussions on implementations in Sec V.

IV. System-level Anomaly Detection

We now consider the distributed detection problem at the system level where the control center receives binary decisions $\{u_{it}\}$ from individual sensors. From Sec. IV, we know that ICA-GAN at each sensor transforms different sensor measurements Z_{it} to the same uniform *i.i.d.* samples under $\mathcal{H}_0$. Because the uniformity detector at all sensors are identical, they all have same false positive rate $\alpha = \alpha_i$. Furthermore, because noise process (w_{it}) are statistically independent across sensors, we have, under $\mathcal{H}_0'$, u_{it} are *i.i.d.* Bernoulii $\mathrm{B}(\alpha)$, and $u_t \sum_i u_{it}$ a binomial random variable $\mathrm{Bin}(K, \alpha)$.

We derive next a Neyman-Pearson detection rule at the control center given detection vector $u_t = (u_{1t}, \cdots, u_{Kt})$ under the standard conditional independent assumption, *i.e.,* conditional on $\mathcal{H}_k'$, u_t have independent entries. Because anomaly model is arbitrary, we further assume that the true positive rates β_i for all detectors are the same. Let $\beta = \beta_i$.

[§]The standard GAN trains a generative network that transforms a uniform distribution to an underlying distribution of a data set.

Algorithm 1 ICA-GAN. The experiments in the paper used the values $\alpha = 0.0001$, $\lambda = 0.1$, $b = 100$, $c = 10$, $M = 80$, $N = 50$.

Require: : α, the learning rate. λ, the gradient penalty coefficient. b, the batch size. c, the number of iterations of the discriminator per generator iteration. M, the block size for the data sequence. N, the number of the independent components for ICA.

for Number of training iterations **do**
  for $k = 1, \cdots, c$ **do**
    for $i = 1, \cdots, b$ **do**
      Sample $U = (U_1, \cdots, U_N) \overset{\text{i.i.d}}{\sim} \mathcal{U}^{(N)}(0,1)$ from uniform distribution.
      Sample a random time t for the start of the time sequence. Get $Z_t = (z_t, z_{t-1}, \cdots, z_{t-(M-1)})$ measurements sequence from data sequence.
      Sample a random number $\epsilon \sim \mathcal{U}(0,1)$.
      $\tilde{U} \leftarrow g_\theta(Z_t)$
      $\hat{U} \leftarrow \epsilon U + (1-\epsilon)\tilde{U}$
      $L_i \leftarrow f_\omega(\tilde{U}) - f_\omega(U) + \lambda(\|\nabla_{\hat{U}} f_{\omega(\hat{U})}\|_2 - 1)^2$
    end for
    Update the discriminator parameter ω by descending its stochastic gradient:
    $$\omega \leftarrow Adam(\nabla_\omega \big[\tfrac{1}{b} \sum_{i=1}^{b} L_i\big])$$
  end for
  for $i = 1, \cdots, b$ **do**
    Sample a random time t for the start of the time sequence. Get $Z_t = (z_t, z_{t-1}, \cdots, z_{t-(M-1)})$ measurements sequence from real data sequence.
    $L_i \leftarrow -f_\omega(g_\theta(Z_{\{t,\cdots,t+(M-1)\}}))$
  end for
  Update the ICA-GAN generator parameter θ by descending its stochastic gradient:
  $$\theta \leftarrow Adam(\nabla_\theta \big[\tfrac{1}{b} \sum_{i=1}^{b} L_i\big]$$
end for

Following [30], the log-likelihood ratio is given by

$$\begin{aligned}\mathcal{L}(u_t) &= \log \frac{\Pr(u_t|\mathcal{H}_1')}{\Pr(u_t|\mathcal{H}_0')} \\ &= \log \frac{\beta(1-\alpha)}{\alpha(1-\beta)} \sum_i u_{it} + K \log \frac{1-\beta}{1-\alpha}.\end{aligned}$$

Noting that $\beta > \alpha$ for any reasonable local detector, the Neyman-Pearson test is given by a threshold on the sum of sensor decision variables $\sum_i u_{it}$:

$$\sum_{i=1}^{K} u_{it} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \tau, \tag{8}$$

where τ is chosen to satisfy the false positive rate constraint. Given the desired upper bound α_0 on the false positive rate of the central detector, we set

$$\tau = \min\left\{k : \alpha_0 \le \sum_{j=k}^{K} \binom{K}{j} \alpha^j (1-\alpha)^{(K-j)}\right\},$$

where we ignore possible randomizations to make the false positive exactly α_0.

Note that the detector defined in (8) is uniformly most powerful (UMP) under the assumptions that sensor-level detectors produce (conditionally) independent decisions with identical TPR. Note also that, although we assume that sensors synchronously communicate their local decisions $\{u_{it}\}$, the above derivation shows that the central detector can just as well operate *asynchronously.* The structure of the detector and the decision rule (8) remain the same. Indeed, the above idea also applies to local sensor decisions where the sensor combines multiple detections from smaller blocks to produce more reliable detections. The advantage is that training ICA-GAN with a low-dimensional input vector is considerably simpler than training a higher dimensional one.

V. Numerical Case Studies

We present three case studies that cover the three types of anomalies considered in this paper. Wherever possible, publicly available real data sets were used.

ICA-GAN implementations in the three case studies shared the same structure, although parameters used are tuned differently depending on the training data. The specific data sets used in the case studies are described in their respective subsections. In all three case studies, we trained the generator with three hidden layers and 100 neurons at each hidden layer. Hyperpolic tangent in Case I and Rectified Linear Units (ReLU) in Case II-III activation function at the final layer were used as the activation functions. For the discriminative network, we also used three hidden layers with 100 neurons. A modification of a standard implementation of Wasserstein distance[¶] was used in the ICA-GAN training with Adam optimization algorithm using mini-batches of 100 data samples.

[¶]https://keras.io/examples/generative/wgan_gp/

In performance evaluation, we obtained the *receiver operating characteristic* (ROC) curves over Monte Carlo simulations. ROC curves plot TPR (probability of detection) against FPR (probability of false alarm), which shows the detection power across the entire range of FPR constraints. We paid specific attention to FPR=0.05 as in standard power system applications [6].

A. Benchmark techniques and implementations

While there are few comparable techniques in the literature for detecting general sequence anomalies in CPOW and high-resolution PMU measurements, we compared three benchmarks that have similar characteristics with ICA-GAN and are potentially applicable in the applications considered in the case studies presented here.

The normalized residue test (NRT) [6] is the classic technique for bad-data detection for power system state estimation. NRT collects measurements from the local sensors and form a centralized anomaly detection. When multiple anomalies occur simultaneously, a standard approach is to remove bad data recursively. In our implementation, we apply the NRT-test to isolate the measurement with the largest total-residue-error calculated over the sequence. If it failed the NRT-test, the data would be declared bad and removed from the system until either the measurement data pass the test or the system becomes unobservable.

The one-class support vector machine (OC-SVM) [16] is a semisupervised machine learning method trained with anomaly-free historical data. It operates under a similar set of model assumptions, except that it does not deal with temporal dependencies in data. We used the radial basis function as a nonlinear kernel. We evaluated the results on the test sequences using the anomaly score function we achieved. We varied the threshold parameter of SVM to get different points on the ROC curve. We used the scikit-learn library for the implementation.

The fast unsupervised anomaly detection (F-AnoGAN) [17] is an auto-encoder technique trained on anomaly-free data. We used a generator and a discriminator with three hidden layers in Wasserstein GAN and a deep neural network with two hidden layers and 100 neurons in each layer in the auto-encoder. The input took 80 consecutive measurement samples and encoded them into latent variables of dimension to 50. We evaluated the results considering the reconstruction error of the auto-encoder. The training was done on a GPU using the Tensorflow-GPU library.

B. Case I: System anomalies in CPOW measurements

We used the EPFL data set involving a battery energy storage system connected at a bus [32], as shown in Fig. 4 (top left). The battery system produced injections that emulated different levels of anomaly events. We used CPOW measurements on the bus voltage and current measurements

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPWRS.2022.3168529, IEEE Transactions on Power Systems

A and B. Direct measurements on anomaly current at C were not used. The CPOW measurements were direct samples of the voltage/current waveforms at 50kHz, and the anomaly power injection varied from 0 to 500kW. The EPFL data set contained anomaly and anomaly-free data, each with 100,000 samples within 2 seconds.

Fig. 4 (top-right and the bottom panel) shows the anomaly and anomaly-free waveforms of the bus voltage and current measurements. There is little difference between the anomaly and anomaly-free voltage CPOW measurements, while noticeable differences are shown in the current measurements. It was expected that the two current detectors would be more reliable than the voltage detector. However, the control center would not know which detector would be reliable a priori.

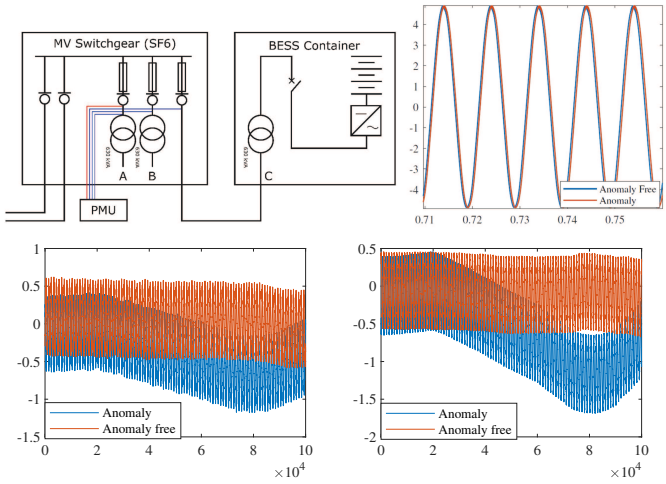

Fig. 4: Learning structure of ICA-GAN.

Three sensor-level detectors were implemented using the bus voltage and current measurements at A and B. The anomaly-free training data were separated into training and testing sets of the ratio 6:4. The training set contained 120 batches of 500 consecutive samples and the test set 80 batches. For each batch of samples, ICA-GAN generated a batch of preprocessed samples on which uniformity tests were made. A single decision was made by each sensor every 0.01s. The thresholds for the sensor-level detection were chosen such that their FPRs were all equal to 0.2.

The detector at the control center combined two consecutive blocks of the three local decisions. Fig. 5 shows the ROC curves of the local and central detectors[‖]. We observed that the central detector significantly improved the performance of local detectors, even when combining the less reliable bus voltage sensor. In particular, at FPR = 0.05, the TPR is above 0.7 whereas local detectors' TPR were below 0.31.

[‖]We did not include OC-SVM and F-AnoGAN in the central detector performance because the local detectors for these algorithms did not have ROC curves above the 45^o diagonal to be useful.

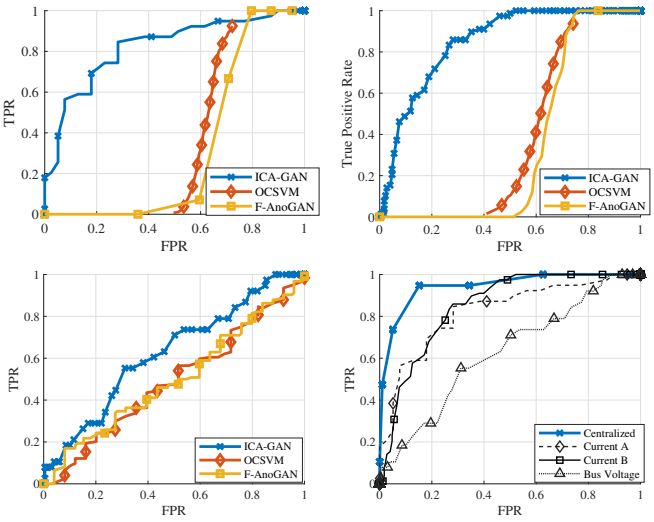

Fig. 5: ROC curves (TPR at FRP=0.05). Top left: Detector at A (TPR=0.2564). Top right: Detector at B (TPR=0.3077). Bottom left: Detector at the bus(TPR=0.1053). Bottom right: Detector at the control center (TPR=0.7368.)

C. Case II: Natural anomalies in PMU measurements

Here we considered natural anomalies (bad data) involving multiple non-interacting anomalies. Two sets of simulations were performed. One is a small four-bus system used in the EPFL Smart Grid Project [31]. We used the 50 Hz PMU measurements collected on April 1st, 2016 from 5 PM to 6 PM. The second is based on the 133-bus synthetic North Texas transmission system [33] where one hour of 30 Hz PMU measurements are used in the simulation. We simulated non-interacting bad sequences and unobservable attack sequences for each system. For both systems, we used the one-phase equivalent of the three-phase systems.

The anomaly-free data were real-data measurements on the EPFL and North Texas Synthetic Systems. Gaussian mixture anomaly sequences were added to the anomaly-free measurements. Four of the ten measurements in the EPFL system and 6 out of 266 measurements in the North Texas system contained anomalies.

We separated the available data into training and test sets. Using EPFL data sequence we created a training set that has 1000 batches of 80 consecutive anomaly-free samples and a test set that has 500 batches of 80 anomaly sequences and 500 batches of 80 anomaly-free sequences for each measurement. Each test sequence consisted of 1.6 seconds of PMU measurement. Similarly, using the North Texas data, we created a training set with 900 batches of 80 consecutive anomaly-free samples from the historical samples, and a test set with 225 batches of 80 anomaly sequences and 225 batches of 80 anomaly-free sequences for each measurement. Each test sequence consisted of 2.6 seconds of PMU measurement.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPWRS.2022.3168529, IEEE Transactions on Power Systems

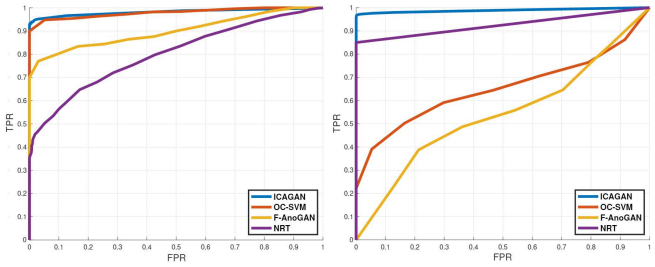

Fig. 6: ROC curves for anomaly case 1. Left: The EPFL System. Right:The Synthetic North Texas system.

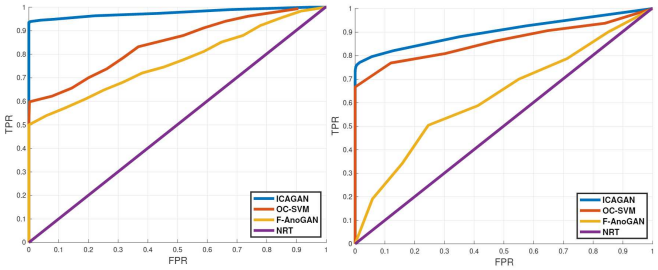

Fig. 7: ROC curves for anomaly case 2. Left: The EPFL System. Right:The Synthetic North Texas system.

The Wasserstein ICA-GAN was trained to obtain the transformation function from the measurements to the independent components. $b = 80$ consecutive measurements were used as inputs for the generator and the 50-dimensional output of the generator was transferred to the discriminator. We fed another 50-dimensional i.i.d. uniform samples to the discriminator.

As a preprocessing step before applying ICA-GAN, we used a linear least-squares prediction to decorrelate the measurement samples. The input layer of the ICA-GAN neural network was a linear least-squares predictor that whitens the input sequence.

After the ICA-GAN generator was used to convert the samples to i.i.d. sequence samples, we used an additional step to convert the distribution of the ICA sequence to uniform distribution. We used the empirical CDF of anomaly-free samples to achieve this transformation. After these steps, with the trained ICA-GAN we constructed the uniformity test algorithm. We used the samples to apply the K_1-coincidence test as defined in 5. We used this approach for each measurement sequence individually. If at least one anomaly measurement sequence is detected, we assumed it is a successful detection.

The ROC curves of ICA-GAN and benchmark techniques are in Fig. 6. We observed that ICA-GAN achieved the best TPR across all FPRs and Table I (second column) shows TPR at FPR=0.05. ICA-GAN had a significantly higher true positive rate than the tested benchmarks.

The conventional NRT did not work well on the EPFL system simulation compared to the Texas system simulation possibly because a larger ratio of measurements had a bad sequence. OC-SVM performed similarly to ICA-GAN's on the EPFL data set but was less successful on the Texas system simulation. F-AnoGAN had worse performance than ICA-GAN and OC-SVM in both cases. ICA-GAN had the best performance on both systems and had higher than 90% TPR even for small FPRs.

D. Case III: unobservable attacks on PMU state estimation

We considered the extreme case of unobservable attack as an example to demonstrate the potential and importance of exploiting the inherent statistical properties and temporal dependencies of the data. An attack is "unobservable" when the data are manipulated so that the altered measurements and a fake state satisfy the underlying measurement equation. Therefore, no algebraic technique is capable of detecting such an attack. However, such attacks inevitably alter the underlying probability distribution and inter-temporal dependencies of the measurements. It is through these properties that our approach can make consistent detection.

We simulated the unobservable attacks, which is an extreme case of attack. The purpose was to illustrate that attacks that could not be detected based on the system model alone could be detected by ICA-GAN that exploited distribution properties. We constructed an unobservable data attack on the EPFL system and the North Texas system. We used the method in [19] to obtain an attack vector a_t. We assumed that the measurements in the system were such that the system was marginally observable. Then it was possible to design an unobservable data attack by manipulating 4 out of 10 measurements on EPFL system data and on 6 out of 266 measurements on North Texas synthetic system data. The attack vector can then be added to the measurements as an unobservable attack:

$$z_t' = z_t + w_t a_t,$$

where we constructed w_t from independent samples from a Gaussian Mixture Model.

Fig. 7 showed the ROC curves where ICA-GAN had better performance than all compared methods with varying significance levels. The TPRs at FRP= 0.05 for the tested benchmarks were shown in the third column of Table I.

As expected, NRT performed as if it were a random selection without using measurements in both cases. OC-SVM performance was the closest to ICA-GAN, but there was still a significant difference. F-AnoGAN had a worse performance than OC-SVM on both systems. ICA-GAN had the best performance in both simulations because; i) the independent component analysis approach transformed the consecutive measurements to an independent sequence

of samples ii) it detected the changes in the probability distribution rather than trying to find outlier samples.

Algorithms	bad data (EPFL)	bad data (Texas)	data attack (EPFL)	data attack (Texas)
ICA-GAN	97%	95%	94%	80%
OC-SVM	94%	35%	61%	71%
F-AnoGAN	78%	5%	53%	18%
NRT	50%	88%	5%	5%

TABLE I: TPR values of different algorithms at FPR=0.05 constraint under bad data and data attack anomalies.

Next, we experimented with the application of the anomaly detection scheme as a data cleansing step for state estimation. When the bad data was detected, we deleted the bad measurements from the measurement vector. A linear Bayesian estimator was used to replace anomaly sensor data with pseudo-measurements from clean measurements, followed by the standard weighted-least-squares state estimator. Table II showed the average least-squares of the tested benchmarks along with the performance when there were no anomalies and the performance when anomalies were undetected. We observed that ICA-GAN had the potential as an effective data cleansing technique.

Algorithms	bad data (EPFL)	bad data (Texas)	data attack (EPFL)	data attack (Texas)
Anomaly-free Meas.	6.6e-07	2.5 e-06	6.7e-07	4.3 e-06
Anomaly Meas.	1.1e-02	1.4 e-02	2.3e-02	2.3 e-02
Cleaned by ICA-GAN	1.7e-03	3.2 e-03	3.7e-03	6.4 e-03
Cleaned by OC-SVM	3.2e-03	6.2 e-03	1.5e-02	1.2 e-02
Cleaned by F-AnoGAN	6.1e-03	8.2 e-03	1.3e-02	2.1 e-02
Cleaned by NRT	8.7e-03	4.0 e-03	2.1e-02	2.2 e-02

TABLE II: Average squared error of sate estimation.

VI. Discussions

We discuss in this section some of the limitations of the proposed approach, practical implementation issues, and unresolved problems outside the scope of this work that requires further investigation.

1) Some limitations of the proposed techniques.: We have taken a minimalist approach in modeling anomalies, which covers a wide range of anomalies under the statistical hypothesis testing framework. However, certain anomalies may not alter the underlying probability distribution, therefore undetectable by the proposed technique. One such case is the timing attack considered in Barreto *et al.* [37] where algebraic techniques that exploit the deterministic rank-one property are used in detection.

The proposed machine learning approach is based on a GAN approach to characterize the anomaly-free distribution implicitly. We assume that historical data used to train the deep learning network are certifiably anomaly-free. To this end, we assume that cross-validation techniques are used in selecting historical data to be used in training, which minimizes but does not eliminate contamination of training data by natural anomalies.

The more difficult challenge is the adversarial learning problem, where an adversary may manipulate training data. Currently, there is no fault-proof technique applicable to real-time anomaly detection problems considered in this paper to our best knowledge.

2) Offline vs. online training.: A critical component of ICA-GAN is the GAN training of a neural network that extracts independent components. In principle, such training can be performed either offline using historical data or online using recent measurements. Effective online training, in particular, allows the monitoring system to track system variations dynamically, provided that training converges quickly.

In our experiments, 1.2 seconds of CPOW measurements appeared to be sufficient in the system anomaly detection in the EPFL battery energy system data set in Case Study I (Sec. V.B). For the relatively slower PMU measurements in Case Study II-III, 40 minutes of data were used in training. These empirical results suggest that online training may potentially be viable.

3) Anomalies vs. system dynamics.: We make a practical (rather than mathematical) distinction between normal operations such as topology/load/generation changes from system or data anomalies. Because the proposed technique can be applied at the sensor level that may not have global information about network conditions (such as topology changes) and dispatch points, the detection algorithm may generate false alarms by mistakenly treating normal operations as anomalies.

In practice, the sensor-level detection should be synchronized with the five-minute real-time dispatch period in real-time market operations so that the detection algorithm discounts measurements during the normal transient periods and known topology changes. How to coordinate system operations with anomaly detection in the monitoring system is of practical significance and deserves future investigation.

VII. Conclusion

We developed a data-driven deep learning approach to anomaly detection consists of sensor-level detectors that assume no prior models on anomaly and anomaly-free data; only anomaly-free training samples are used. Sensor-level decisions are combined at the control center to produce more reliable global decisions. To our best knowledge, the proposed technique is the first designed specifically for high-resolution measurements for CPOW and PMU streaming and can deal with both data and system anomalies.

References

[1] A. Silverstein and J. Follum, "High-resolution, time-synchronized grid monitoring devices," North American Synchrophasor Initiative,

Tech. Rep. NAPSI-2020-TR-004, March 2020, [ONLINE], available (2020/8/20) at https://www.naspi.org/node/819.

[2] Y. Liu, "Beyond today's synchrophasor," in *NSF Workshop on Forging Connections between Machine Learning, Data Science, and Power Systems Research*, March 2020, [ONLINE], available (2020/8/20) at https://sites.google.com/umn.edu/ml-ds4pes/presentations.

[3] X. Wang, Y. Liu, and L. Tong, "Adaptive subband compression for streaming of continuous point-on-wave and pmu data," *IEEE Transactions on Power Systems*, pp. 1–1, 2021.

[4] F. C. Schweppe, J. Wildes, and D. P. Rom, "Power system static state estimation, Parts I, II, III," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-89, pp. 120–135, 1970.

[5] H. M. Merrill and F. C. Schweppe, "Bad data suppression in power system static state estimation," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-90, no. 6, pp. 2718–2725, 1971.

[6] A. Abur and A. G. Expósito, *Power System State Estimation: Theory and Implementation*. CRC Press, 2004.

[7] D. M. Falcao, P. A. Cooke, and A. Brameller, "Power system tracking state estimation and bad data processing," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-101, no. 2, pp. 325–333, 1982.

[8] A. Abur, A. Keyhani, and H. Bakhtiari, "Autoregressive filters for the identification and replacement of bad data in power system state estimation," *IEEE Transactions on Power Systems*, vol. 2, no. 3, pp. 552–558, Aug 1987.

[9] H. Salehfar and R. Zhao, "A neural network preestimation filter for bad-data detection and identification in power system state estimation," *Electric Power Systems Research*, vol. 34, no. 2, pp. 127 – 134, 1995.

[10] K. Mahapatra, N. R. Chaudhuri, R. G. Kavasseri, and S. M. Brahma, "Online analytical characterization of outliers in synchrophasor measurements: A singular value perturbation viewpoint," *IEEE Transactions on Power Systems*, vol. 33, no. 4, pp. 3863–3874, 2018.

[11] M. N. Kurt, O. Ogundijo, C. Li, and X. Wang, "Online cyber-attack detection in smart grid: A reinforcement learning approach," *IEEE Transactions on Smart Grid*, vol. 10, no. 5, pp. 5174–5185, 2019.

[12] M. Ozay, I. Esnaola, F. T. Yarman Vural, S. R. Kulkarni, and H. Vincent Poor, "Smarter security in the smart grid," in *2012 IEEE Third International Conference on Smart Grid Communications (SmartGridComm)*, 2012, pp. 312–317.

[13] Y. Chen, S. Huang, F. Liu, Z. Wang, and X. Sun, "Evaluation of reinforcement learning-based false data injection attack to automatic voltage control," *IEEE Transactions on Smart Grid*, vol. 10, no. 2, pp. 2158–2169, 2019.

[14] K. R. Mestav and L. Tong, "Universal data anomaly detection via inverse generative adversary network," *IEEE Signal Processing Letters*, vol. 27, pp. 511–515, 2020.

[15] M. Zhou, Y. Wang, A. K. Srivastava, Y. Wu, and P. Banerjee, "Ensemble-based algorithm for synchrophasor data anomaly detection," *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 2979–2988, 2019.

[16] B. Schölkopf, R. Williamson, A. Smola, J. Shawe-Taylor, and J. Platt, "Support vector method for novelty detection," *Proceedings of the 12th International Conference on Neural Information Processing Systems*, pp. 582–588, 1999.

[17] T. Schlegl, P. Seeböck, S. M. Waldstein, G. Langs, and U. Schmidt-Erfurth, "f-anogan: Fast unsupervised anomaly detection with generative adversarial networks," *Medical Image Analysis*, vol. 54, pp. 30 – 44, 2019.

[18] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," 2017.

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

[20] J. Kim, L. Tong, and R. J. Thomas, "Dynamic attacks on power systems economic dispatch," *48th Asilomar Conference on Signals, Systems and Computers*, November 2014.

[21] L. Jia, J. Kim, R. J. Thomas, and L. Tong, "Impact of data quality on real-time locational marginal price," *IEEE Transactions on Power Systems*, vol. 29, no. 2, pp. 627–636, 2014.

[22] J. Kim, L. Tong, and R. J. Thomas, "Subspace methods for data attack on state estimation: A data driven approach," *IEEE Transactions on Signal Processing*, vol. 63, no. 5, March 2015.

[23] J. Liang, L. Sankar, and O. Kosut, "Vulnerability analysis and consequences of false data injection attack on power system state estimation," *IEEE Transactions on Power Systems*, vol. 31, no. 5, pp. 3864–3872, 2016.

[24] 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 Transactions on Power Systems*, vol. 33, no. 5, pp. 4775–4786, 2018.

[25] P. Comon, "Independent Component Analysis, a new concept?" *Signal Processing*, vol. 36, pp. 287–314, Apr. 1994.

[26] P. Brakel and Y. Bengio, "Learning Independent Features with Adversarial Nets for Non-linear ICA," Oct. 2017, arXiv:1710.05050.

[27] F. N. David, "Two combinatorial test of whether a sample has come from a given population," *Biometrika*, vol. 37, no. 1/2, pp. 97–110, 1950.

[28] I. I. Viktorova and V. P. Chistyakov, "On the calculation of the power of the test of empty boxes," *Theory of Probability & Its Applications*, vol. 9, no. 4, pp. 648–653, 1964.

[29] L. Paninski, "A coincidence-based test for uniformity given very sparsely sampled discrete data," *IEEE Transactions on Information Theory*, vol. 54, no. 10, pp. 4750–4755, Oct 2008.

[30] P. K. Varshney, *Distributed Detection and Data Fusion*. Springer, 1997.

[31] M. Pignati *et al.*, "Real-time state estimation of the EPFL-campus medium-voltage grid by using pmus," in *2015 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, Feb 2015, pp. 1–5.

[32] F. Sossan, E. Namor, R. Cherkaoui, and M. Paolone, "Achieving the dispatchability of distribution feeders through prosumers data driven forecasting and model predictive control of electrochemical storage," *IEEE Transactions on Sustainable Energy*, vol. 7, no. 4, p. 1762–1777, Oct 2016.

[33] A. B. Birchfield, T. Xu, K. M. Gegner, K. S. Shetye, and T. J. Overbye, "Grid structural characteristics as validation criteria for synthetic networks," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 3258–3265, 2017.

[34] R. Von Mises, "Über aufteilungs-und besetzungswahrscheinlichkeiten," *Revue de la Faculté des Sciences de l'Université d'Istanbul*, vol. 4, p. 145–163, 1939.

[35] C. Jutten and J. Herault, "Blind separation of sources, part i: An adaptive algorithm based on neuromimetic architecture," *Signal Processing*, vol. 24, no. 1, pp. 1–10, 1991.

[36] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural Networks*, vol. 13, no. 4, pp. 411 – 430, 2000.

[37] S. Barreto, M. Pignati, G. Dán, J.-Y. Le Boudec, and M. Paolone, "Undetectable timing-attack on linear state-estimation by using rank-1 approximation," *IEEE Transactions on Smart Grid*, vol. 9, no. 4, pp. 3530–3542, 2018.

Kursat Mestav Kursat R. Mestav received the B.S. degree in electrical and electronics engineering from Bilkent University, Ankara, Turkey in 2016, and the M.S. and Ph.D. degrees in electrical and computer engineering from Cornell University in 2019 and 2020, respectively. In 2021, he joined Qualcomm as a senior systems engineer for AI/ML-based 5G NR modem design. His research interests include statistical inference and machine learning with application to energy systems and the smart grid. He was the recipient of the 2018 best paper award at IEEE PES General Meeting.

Lang Tong Lang Tong (S'87,M'91,SM'01,F'05) is the Irwin and Joan Jacobs Professor in Engineering and the Cornell site-director of the Power Systems Engineering Research Center (PSERC). His current research focuses on energy and power systems, smart grids, and the electrification of transportation systems.

Xinyi Wang Xinyi Wang received B.Sc degree in electrical engineering from Tianjin University, Tianjin, China, in 2019. She is currently working toward the Ph.D. with Cornell University, Ithaca, NY, USA. Her current research interest lies in statistical inferences, machine learning and algorithms designed for time series and their application to power and energy systems.