Data-driven Physics-based Solution for False Data Injection Diagnosis in Smart Grids

Rodrigo D. Trevizan, Cody Ruben, Keerthiraj Nagaraj, Layiwola L. Ibukun, Allen C. Starke
Arturo S. Bretas, Janise McNair, Alina Zare
Dept. of Electrical & Computer Engineering
University of Florida; Gainesville, FL 32611
rodtrevizan@ufl.edu, cruben31@ufl.edu, k.nagaraj@ufl.edu, llibukun@ufl.edu, allen1.starke@ufl.edu
arturo@ece.ufl.edu, mcnair@ece.ufl.edu, azare@ece.ufl.edu

***Abstract*—This paper presents a data-driven and physics-based method for detection of false data injection (FDI) in Smart Grids (SG). As the power grid transitions to the use of SG technology, it becomes more vulnerable to cyber-attacks like FDI. Current strategies for the detection of bad data in the grid rely on the physics based State Estimation (SE) process and statistical tests. This strategy is naturally vulnerable to undetected bad data as well as false positive scenarios, which means it can be exploited by an intelligent FDI attack. In order to enhance the robustness of bad data detection, the paper proposes the use of data-driven Machine Intelligence (MI) working together with current bad data detection via a combined Chi-squared test. Since MI learns over time and uses past data, it provides a different perspective on the data than the SE, which analyzes only the current data and relies on the physics based model of the system. This combined bad data detection strategy is tested on the IEEE 118 bus system.**

***Index Terms*—anomaly detection, false data injection, gross error analysis, machine intelligence, power system state estimation, Reed-Xaoli.**

I. Introduction

The next-generation power grid, named the Smart Grid (SG), has drawn the attention of academia, industry and government due to the great impact of such systems on society, economics and the environment. These next generation systems integrate control, communication and computation aiming to achieve stability, efficiency and robustness of the physical processes. While a great amount of research has been done towards these objectives, science and technology related to the cyber-physical security of SGs are still immature. Additionally, many critical infrastructures are currently transitioning towards the paradigm of SGs by increasing the dependency of control of physical processes on communication networks, thus becoming exposed to cyber-threats [1].

Power system monitoring is critical for guaranteeing reliable operation of power grids. Currently, real-time monitoring is done through Power System State Estimation (PSSE) [2]. PSSE provides relevant information on the condition of a power grid based on the readings of sensors that measure electrical quantities. These meter readings are commonly transmitted to a Supervisory Control and Data Acquisition (SCADA) system, which implements centralized monitoring and control for the electrical grid, where PSSE is performed. One very important feature of PSSE is its error processing capability. Measurements that are clearly inconsistent are discarded in the pre-filtering step, which precedes state estimation itself. Following state estimation using pre-filtered data, a post-processing step called bad data analysis is performed. This step aims at detecting bad data or Gross Errors (GE), which correspond to statistically large errors.

The increasing dependence on power system monitoring and control raises concerns with respect to cyber-threats. One type of cyber-attack that has drawn the attention of the academic community is the false data injection (FDI) attack, whereby a subset of measurements values are modified by an adversarial attacker aiming to disrupt the power grid. While Bad Data Analysis is capable of detecting many instances of Gross Errors via tests such as $J(\hat{x})$, largest normalized residual [3] or innovation-based [4] approaches, cyber-attacks might be engineered to be very hard to detect [5]. Methods devised to treat FDI attacks include Generalized Likelihood Ratio Detector with L-1 Norm Regularization [6], a scheme for protecting a selected set of measurements and verifying the values of a set of state variables independently [7] and the estimation of the normalized composed measurement error for detection of malicious data attacks [8]. However, some of these methods present a few drawbacks, such as large computational costs, including search routines, and the assumption that a few measurements can be protected from cyber-attacks. Additionally, they do not use information from past data to improve its robustness.

In this paper we propose a data-driven, physics-based method for the detection of FDI-type attacks in SG. This method combines state estimation-based Bad Data Analytics with purely data-driven anomaly analysis to detect malicious data attacks.

The remainder of the paper is organized as follows. In Section II, a review of PSSE is presented. The details of the data-driven method are shown in Section III. The method that combines phisics-based and data-driven methods is presented in Section IV. The results of numerical tests used to evaluate the performance of this combined method are shown in Section V. Finally, Section VI presents the conclusions of this work.

978-1-7281-1981-6/19/$31.00 ©2019 IEEE

II. Power System State Estimation

In modern Energy Management Systems (EMS), the State Estimation (SE) process is the core process for situational awareness of a power system and is used in many EMS applications, including the detection of bad data. The common approach to SE is using the classical Weighted Least Squares (WLS) method described in [2]. In this approach, the system is modeled as a set of non-linear equations based on the physics of the system:

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

where $z \in \mathbb{R}^m$ is the measurement vector, $x \in \mathbb{R}^N$ is the vector of state variables, $h : \mathbb{R}^N \to \mathbb{R}^m$ is a continuously non-linear differentiable function, and $e \in \mathbb{R}^m$ is the measurement error vector. Each measurement error, e_i is assumed to have zero mean, standard deviation σ_i and Gaussian probability distribution. m is the number of measurements and N is the number of states.

In the classical WLS approach, the best estimate of the state vector in (1) is found by minimizing the cost function $J(x)$:

$$J(x) = \|\mathrm{z} - h(x)\|^2_{R^{-1}} = [\mathrm{z} - h(x)]^T R^{-1} [z - h(x)] \tag{2}$$

where R is the covariance matrix of the measurements. In this paper, we consider the standard deviation of each measurement to be 1% of the measurement magnitude, which has been shown to improve the detection of bad data [9]. In order to solve this problem, (1) is linearized at a certain point x^* in (3) and the optimal states are found through an iterative process.

$$\Delta z = H \Delta x + e \tag{3}$$

where $H = \frac{\delta h}{\delta x}$ is the Jacobian matrix of h at the current state estimate x^*, $\Delta z = z - h(x^*) = z - z^*$ is the correction of the measurement vector and $\Delta x = x - x^*$ is the correction of the state vector. The WLS solution is the projection of Δz onto the Jacobian space by a linear projection matrix P, i.e. $\Delta z = P \Delta \hat{z}$. Letting $r = \Delta z - \Delta \hat{z}$ be the residual vector, the P matrix that minimizes $J(x)$ will be orthogonal to the Jacobian range space and to r; $\Delta \hat{z} = H \Delta \hat{x}$. This is in the form:

$$\langle \Delta \hat{z}, r \rangle = (H \Delta \hat{x})^T R^{-1} (\Delta z - H \Delta \hat{x}) = 0. \tag{4}$$

Solving (4) for $\Delta \hat{x}$:

$$\Delta \hat{x} = (H^T R^{-1} H)^{-1} H^T R^{-1} \Delta z. \tag{5}$$

At each iteration, a new incumbent solution x^*_{new} is found and updated following $x^*_{new} = x^* + \Delta \hat{x}$. (5) is solved each iteration until $\Delta \hat{x}$ is sufficiently small to claim convergence of the solution. Once the SE converges, the final residual values in r are used for the detection of bad data in the measurement vector z. The measurements are considered to be i.i.d, so the statistical chi-squared test is used. The value of $J(x)$ is compared to the chi-squared value, $\chi^2_{(m-N),p}$, for $(m - N)$ degrees of freedom and probability p. If $J(x)$ is larger than the chi-squared value, an error is detected in the measurements, as shown in (6). This error is identified by finding the largest normalized residual value, (7).

$$J(\hat{x}) = \sum_{i=1}^{m} \left[\frac{z_i - h_i(\hat{x})}{\sigma_i} \right]^2 > \chi^2_{(m-N),p} \tag{6}$$

$$r_i^N = \left| \frac{r_i}{\sqrt{\Omega_{i,i}}} \right| \tag{7}$$

where $\Omega = R - H(H^T R^{-1} H)^{-1} H^T$ is the covariance of residuals.

III. Data-Driven Machine Intelligence

The Machine Intelligence layer of the smart power grid uses the knowledge of already verified data to learn the normal state of a properly functioning grid. It is then able to detect any anomalies introduced into the system at any point forward and alerts the Network layer to identify the anomaly, isolate it from the remainder of the system and take appropriate action to prevent contamination of the system, with regards to both power distribution in other subsystems, and data assimilation by the Machine Learning system itself.

The Machine Learning layer is implemented using the famous Reed-Xaoli (RX) Anomaly Detection [10] algorithm described in (8), where z is the new incoming data, μ is the mean and $\sum^{-1}$ is the inverse covariance matrix. (8) calculates the Mahalanobis distance squared, $\delta^{RX}(z)$, of a given data z, from the mean, μ of the distribution.

$$\delta^{RX}(z) = (z - \mu)^T \sum\nolimits^{-1} (z - \mu) \tag{8}$$

The anomaly detector is trained with the first k number of incoming samples that has not been flagged by the state estimator to generate an initial μ and $\sum^{-1}$ as a starting point. It then accepts new data and uses (8) to determine its Mahalanobis distance and compares it to a threshold value. If the result is below the threshold, the new data is considered to be normal data but if the result is above the threshold, the new data is flagged as an anomaly.

Because data is dynamic and it changes gradually over time, the anomaly detector must be able to adapt with changing trends and so the mean, μ and inverse covariance matrix, $\sum^{-1}$ are updated using the Woodbury Matrix Identity [11] in equations (9) and (10) respectively. Note that this update is done only if the incoming data is considered normal data.

$$\mu_{new} = (1 - \alpha)\mu + \alpha(z - \mu) \tag{9}$$

$$\sum\nolimits_{new}^{-1} = \frac{1}{1 - \alpha} \left[\sum\nolimits^{-1} - \frac{(z - \mu)(z - \mu)^T}{\frac{1-\alpha}{\alpha} + (z - \mu)^T (z - \mu)} \right] \tag{10}$$

where z is the new data, μ is old mean, $\sum^{-1}$ is the old inverse covariance matrix and α is a hyper-parameter value between zero and one that determines how much importance is given to the new data sample versus the old mean. We determine the value of k and α through experimentation.

IV. Data-driven Physics Model for FDI Diagnosis

In the proposed data driven physics based solution, we make use of results from both the physics-based SE and the data-based RX algorithm to find a combined distance measure. This hybrid anomaly score will be compared with a threshold to detect whether a given measurement is anomalous or normal.

For the SE part of the distance measure, each sample is analyzed individually as described in Section II. The current states of the system are estimated based on the measurement set, and minimizing the objective function (6). The $J(\hat{x})$ is used as the SE portion of the combined distance measure.

The purely data-driven analysis is divided in two parts. Initially for first k samples, we obtain predictions from SE and identify the normal samples to form the initial mean and covariance matrix for RX algorithm. We regard this phase as training phase for the data driven RX algorithm, where the trained model tries to understand the distribution of normal samples. After obtaining the initial mean and covariance matrices, we start the testing phase of RX algorithm where each incoming sample is used to calculate the Mahalanobis distance squared as shown in equation (8). We decide whether to update the mean and covariance matrix based on a threshold value for the samples in the testing phase. This threshold is based on the distribution of Mahalanobis distance squared values for the normal samples collected during training phase and it is calculated as the sum of their mean and a constant times their standard deviation. We observed that this distribution was roughly following a bell curve, so a threshold value corresponding to the constant value of 1.5 would cover nearly 94% of the all the Mahalanobis distance values for normal samples. Hence, we decided the constant value to be 1.5.

The distance measures obtained from SE and RX algorithm are then added and compared with a threshold value based on the confidence level to detect whether a given sample is normal or anomalous as shown in (11).

$$J_C = \delta^{RX}(z) + J(\hat{x}) > \chi^2_{(2m-N),p} \tag{11}$$

The distance measures from the SE and the RX algorithm are added together and considered independent due to the independence of the two estimation strategies. The SE uses the physics based model of the power system and only considers the current measurement set. The RX algorithm considers past samples and is based purely on data, so these distance measures are determined independently. The degrees of freedom chosen for this combined Chi-squared test is a summation of the $m-N$ degrees of freedom in the SE Chi-squared test and the m features used in the RX algorithm.

V. Case Study

The proposed strategy for data-driven and physics-based FDI diagnosis was validated using the IEEE 118-bus system. The measurement set included all real and reactive power flows and injections and all voltage magnitudes, resulting in 1070 measurements. Using the MATLAB package MATPOWER, 10,000 samples, or measurement sets, were generated with Gaussian noise. 530 of these samples were chosen at random to insert a GE into a single measurement within the sample. These errors were of random size between 20 and 30 standard deviations away from the true measurement value. Out of these 10,000 samples, 2,000 (k) were used for training the RX algorithm (with an α value of 0.8) while the remaining 8,000 were used for testing. The implementation of data driven physics based model and evaluation of results was conducted using Python libraries such as Pandas, SciPy [12], and Scikit-learn [13] in Anaconda environment. The statistical tests (6) and (11) are performed with a 95% confidence level.

A. Performance Analysis

To properly evaluate the performance of the 2 estimations strategies included in this paper, we make use of classification metrics [14] described below:

A **Confusion Matrix** is a table that describes the performance of a classification model on a set of test data whose ground truth values are known.

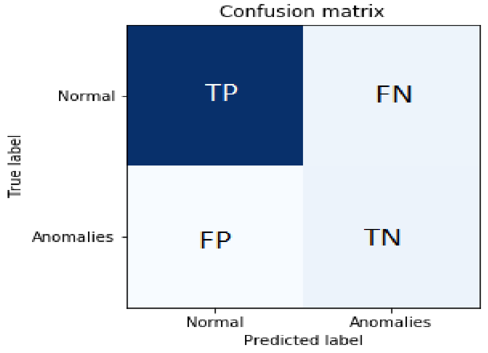

Fig. 1. Confusion Matrix Model

In Fig 1, **True Positives (TP)** refer to normal data that is predicted as normal data. **True Negatives (TN)** refer to anomalous data predicted as anomalous. **False Positives (FP)** refer to anomalous data predicted to be normal. **False Negatives (FN)** refer to normal data predicted to be anomalous.

Accuracy is the ratio of correctly predicted samples to the total number of samples. Accuracy is a good performance metric when the class sizes are balanced in the dataset. Normally in the real world, when working with anomaly detection problem, the number of anomalous samples is usually a lot lesser than the number of normal samples. This means the class sizes are skewed in nature and accuracy would not serve as a good performance metric for an anomaly detection problem. We are including it in our analysis as it is one of the most commonly used classification performance metrics. We will also be including metrics such as Precision, Recall and F1-score, which would provide a better measure of performance for a given anomaly detection strategy. (12) shows the formula to calculate overall accuracy of the model.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{12}$$

Precision is the ratio of number of correctly predicted normal samples to the overall predicted normal samples. Precision is an important metric when we want to minimize False Positives. (13) shows the formula for calculating Precision performance metric.

$$Precision = \frac{TP}{TP + FP} \tag{13}$$

Recall (also called as Sensitivity) is the ratio of number of correctly predicted normal samples to number of true normal samples. This performance metric gives us an idea of how good our model is at identifying normal samples. If we want to minimize the False Negatives, we want to have a very good value of Recall without precision being too low. Recall can be given high preference if there is a need to update the system parameters or store the data for future analysis when the measurement is normal. (14) shows the formula for calculating Recall performance metric.

$$Recall = \frac{TP}{TP + FN} \tag{14}$$

F1-score is the harmonic mean of Precision and Recall. It would be better to have a single performance metric that would consider both Precision and Recall, and which strikes a balance between them. A simple arithmetic mean would result in high value even if the model is terrible, as if one of the values between Precision and Recall is high, this would increase arithmetic mean. When we consider harmonic mean, it would result in a value which is more closer to the lower value among Precision and Recall than arithmetic mean, hence resulting in a more appropriate performance metric. This metric is more useful than accuracy since we usually have uneven class distribution for anomaly detection problems. (15) shows the formula for calculating F1-score performance metric.

$$\text{F1-score} = \frac{2 * Recall * Precision}{Recall + Precision} \tag{15}$$

B. Numerical Results

We present the numerical values of performance metrics such as Confusion matrix, Accuracy, Precision, Recall and F1-score for both State estimator method and the proposed Data-driven Physics based method.

Figure 2 shows the TP, TN, FP and FN values in a confusion matrix for the state estimator method for identifying normal and anomalous samples. The predicted values are compared with the ground truth to arrive at these values.

In table I, we have presented the numerical values of Accuracy, Precision, Recall and F1-score for both normal and anomalous measurements from the anomaly detection model obtained by state estimator method. As we discussed earlier, when the class size is not balanced, just looking at metrics for overall data might mislead the readers about the performance of a model, hence we present the values of the performance metrics for each class. From table I, we can see that some of the values for State estimator based method are low, notably the F1 Score for the anomalous data which is very important for an anomaly detection problem.

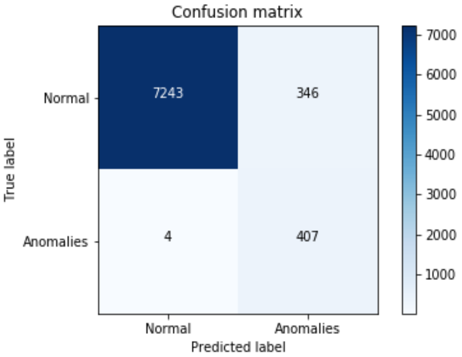

Fig. 2. Confusion Matrix for State Estimator model

TABLE I
ANOMALY DETECTION METRICS FOR STATE ESTIMATOR

Class	Accuracy	Precision	Recall	f1-score
Normal	95.44	99.99	95.00	98.00
Anomalies	99.00	54.00	99.00	70.00

Figure 3 shows the TP, TN, FP and FN values in a confusion matrix for the Data-driven Physics based method for identifying normal and anomalous samples. Comparing Figure 3 with Figure 2 reveals that the combined model does much better job with TP and FN values, and nearly same level of performance as State estimator when it comes to FP and TN.

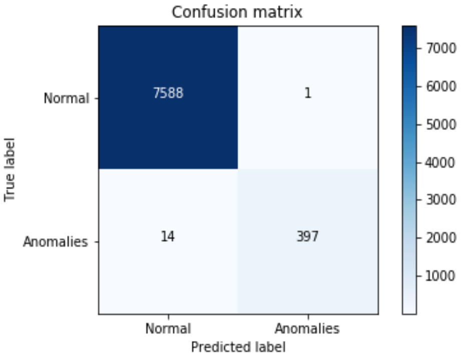

Fig. 3. Confusion Matrix for Data-driven Physics based model

In table II, we have presented the numerical values of Accuracy, Precision, Recall and F1 Score for both normal and anomalous measurements from the anomaly detection model obtained by Data-driven Physics based method. The value of F1 score has improved drastically for the combined model when compared with the State estimator model for the anomalous data.

In tables I and II, we have observed the values of performance metrics for class-wise data. In table III, we can see the

TABLE II
ANOMALY DETECTION METRICS FOR COMBINED MODEL

Class	Accuracy	Precision	Recall	f1-score
Normal	99.98	99.99	99.99	99.99
Anomalies	96.59	99.99	97.00	98.00

values of performance metrics for the overall dataset for both state estimator and combined models. We have included this table to make an easy comparison between the performance of the 2 anomaly detection models obtained from state estimator and data driven physics based methods. From table III, it is evident that the proposed Data driven physics based model performs better than the state estimator model in detecting whether a given measurement is anomalous or normal.

TABLE III
PERFORMANCE RESULTS: COMPARISON OF DIFFERENT METHODOLOGIES FOR OVERALL TEST DATA

Method	Accuracy	Precision	Recall	f1-score
State Estimator	95.62	98.00	96.00	96.00
Combined Model	99.81	99.99	99.99	99.99

VI. CONCLUSION

This paper presents a new Chi-squared test for the detection of gross errors in power system measurements. Gross errors come from many sources in power systems, including FDI attacks as the grid becomes more digital as a SG. The new Chi-squared test combines information from the classical WLS SE residual vales and the Machine Learning RX Anomaly Detection Mahalanobis distance values. The error detection test results show that this Data-driven and Physics-based combined solution improves the performance of the detection of errors in the measurement set for multiple metrics, especially the f1-score, which is an important metric in anomaly detection problems. The combination of data-driven and physics-based solutions in power systems is critical to the future success of the SG and this paper presents a successful application of both solutions that will lead to other advances in the security and reliability of the SG.

In our previous work [15], we proposed an hybrid distributed and decentralized software-defined networking architecture for monitoring SG's network communications and physical data measurements for anomalous behavior. For future work we plan to integrate the produced data-driven physics model with the described SDN architecture to formulate a 3-layered cyber-security system for SGs.

ACKNOWLEDGMENT

This material is based upon work supported by the National Science Foundation under Grant No. 1809739.

REFERENCES

[1] M. Farag, M. Azab, and B. Mokhtar, "Cross-layer security framework for smart grid: Physical security layer," in *IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe)*. IEEE, 2014, pp. 1–7.

[2] A. Monticelli, *State estimation in electric power systems: a generalized approach*. Springer Science & Business Media, 1999, vol. 507.

[3] E. Handschin, F. C. Schweppe, J. Kohlas, and A. Fiechter, "Bad data analysis for power system state estimation," *IEEE Transactions on Power Apparatus and Systems*, vol. 94, no. 2, pp. 329–337, March 1975.

[4] N. G. Bretas, A. S. Bretas, and S. A. Piereti, "Innovation concept for measurement gross error detection and identification in power system state estimation," *IET Generation, Transmission Distribution*, vol. 5, no. 6, pp. 603–608, June 2011.

[5] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," in *Proceedings of the 16th ACM Conference on Computer and Communications Security*, ser. CCS '09. New York, NY, USA: ACM, 2009, pp. 21–32. [Online]. Available: http://doi.acm.org/10.1145/1653662.1653666

[6] O. Kosut, L. Jia, R. J. Thomas, and L. Tong, "Malicious data attacks on smart grid state estimation: Attack strategies and countermeasures," in *2010 First IEEE International Conference on Smart Grid Communications*, Oct 2010, pp. 220–225.

[7] R. B. Bobba, K. M. Rogers, Q. Wang, H. Khurana, K. Nahrstedt, and T. J. Overbye, "Detecting false data injection attacks on dc state estimation," in *Preprints of the First Workshop on Secure Control Systems, CPSWEEK*, vol. 2010, 2010.

[8] A. S. Bretas, N. G. Bretas, B. Carvalho, E. Baeyens, and P. P. Khargonekar, "Smart grids cyber-physical security as a malicious data attack: An innovation approach," *Electric Power Systems Research*, vol. 149, pp. 210 – 219, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0378779617301657

[9] N. G. Bretas and A. S. Bretas, "A two steps procedure in state estimation gross error detection, identification, and correction," *International Journal of Electrical Power & Energy Systems*, vol. 73, pp. 484 – 490, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0142061515002495

[10] H. Kwon and N. M. Nasrabadi, "Kernel rx-algorithm: a nonlinear anomaly detector for hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 2, pp. 388–397, Feb 2005.

[11] M. C. D. K. C. H. Brendan Alvey, Alina Zare, "Adaptive coherence estimator (ace) for explosive hazard detection using wideband electromagnetic induction (wemi)," *Proc.SPIE*, vol. 9823, pp. 9823 – 9823 – 7, 2016. [Online]. Available: https://doi.org/10.1117/12.2223347

[12] E. Jones, T. Oliphant, P. Peterson *et al.*, "SciPy: Open source scientific tools for Python," 2001, [Online; accessed 2018-11-07]. [Online]. Available: http://www.scipy.org/

[13] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[14] S. Godbole and S. Sarawagi, "Discriminative methods for multi-labeled classification," in *Advances in Knowledge Discovery and Data Mining*, H. Dai, R. Srikant, and C. Zhang, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 22–30.

[15] A. Starke, J. McNair, R. Trevizan, A. Bretas, J. Peeples, and A. Zare, "Toward resilient smart grid communications using distributed sdn with ml-based anomaly detection," in *IFIP Conference on Wired/Wireless Internet Communications*, vol. 1. IFIP, 2018, pp. 1–12.