Deep Learning-based Dimensionality Reduction for Anomaly Detection in Smart Grids

Anila Kousar
Dept. of Electrical Engineering
Mirpur Univ. of Science and Technology (MUST)
Mirpur AJK-10250, Pakistan
(e-mail: anila.pe@must.edu.pk)

Saeed Ahmed
Dept. of Electrical Engineering
Mirpur Univ. of Science and Technology (MUST)
Mirpur AJK-10250, Pakistan
(e-mail: saeed.ahmed@must.edu.pk)

Abdullah Altamimi
Dept. of Electrical Engineering
College of Engineering, Majmaah University
Al-Majmaah-11952, Saudi Arabia
(e-mail: a.altamimi@mu.edu.sa)

Su Min Kim*
Dept. of Electronics Engineering
Tech Univ. of Korea (TU Korea)
Siheung, Gyeonggi 15073, South Korea
(e-mail: suminkim@tukorea.ac.kr)
*Corresponding author

Zafar Ali Khan
Dept. of Electrical Engineering
Mirpur Univ. of Science and Technology (MUST)
Mirpur AJK-10250, Pakistan
(e-mail: zafar.pe@must.edu.pk)

***Abstract*—Smart Grids (SG) have emerged as one of the complex cyber-physical systems integrating information and communication technologies to existing power system infrastructure for reliable power delivery. However, this integration makes the system highly susceptible to cyber-attacks, and brings forth the challenge of dealing with big data by the expanded size of SG. To this end, an autoencoder-based feature extraction technique is employed for obtaining discriminant features while preserving the primordial properties of the system. The reconstructed features are then provided as input to binary support vector machine classifier to detect cyber-intrusions, outliers and attacks in SG. Various IEEE test cases are used in the simulation. The performance comparison shows that our proposed scheme outperforms the existing schemes in capturing prominent features leading to improved detection accuracy of the classifier.**

***Index Terms*—cyber-physical systems, smart grids, cyber-attacks, autoencoder, support vector machine**

I. Introduction

Cyber-physical systems (CPS) are complex structures incorporating sophisticated computing, controlling and communication elements to ensure fast, efficient and reliable operation in cyber-space. The integration of 3Cs-computing, controlling and communication-besides increasing the complexity of system has brought forth the challenge of overcoming the system vulnerability to cyber-attacks [1], [2]. Cyber-physical smart grids (SG) have shown exponential growth over the past decade, and have become lucrative target for the hackers while performing communication in cyber-space. Multitudinous studies are carried out to investigate the security compromises in SG using machine-learning (ML) schemes [3]–[8]. Qi et al. [5] used semi-supervised ML approach to detect stealthy attacks in SG. To assess the performance of SG under cyber assaults, Sengan et al. [6] used deep learning to detect malicious events and activities in the system. Introducing different cyber-attacks in smart grid network, Takkidin et al. [7] proposed autoencoder-based ML technique for anomaly detection. Various ML schemes were compared by Saddam et al. [8] to identify and detect false data injection attack in smart grids. However, it becomes complicated and inefficient to build machine learning model with growing dimensions of CPS as the expanded system size increases computational time leading to reduced efficiency of an anomaly detection model, requiring special attention to combat the curse of high dimensionality. Normally, dimensionality reduction (DR) issue is resolved using either feature selection (FS) or feature extraction (FE) techniques. Feature selection technique is based on selecting the optimal features expected to generate the desired outcome while ignoring the less significant features. The search space obtained by FS is basically a subset of the original space, thereby may result in loss of valuable information. The issue of data loss is addressed by FE as it reduces the dimensions by projecting high dimension space to a low dimension space considering all the features. However, FE results in loss of data interpretability.

Lately, researchers have shown great interest on machine learning-based anomaly detection in CPS while considering the curse of high dimensionality [9]–[11]. Authors in [12] employed genetic algorithm-based feature selection to select the discriminant features for DR and detected cyber assaults in SG by inputting the reconstructed features to euclidean distance-based machine learning scheme. In another study [13], authors detected the assailed data in SG utilizing isolation forest machine learning scheme, resolving the issue of high dimensionality by using principal component analysis (PCA)-based feature extraction method. Recent studies reveal that linear discriminant analysis and PCA-based FE methods are outperformed by autoencoder (AE)-based FE in mining the characteristics of non-linear datasets as in SG.

Contrary to previous studies, this study proposes a deep-denoising autoencoder (DAE)-based FE technique for DR leading to detection of stealthy data integrity attack (SDIA) in SG measurements by employing support vector machine (SVM) classifier. The comparison with existing models shows

979-8-3503-1327-7/23/$31.00 ©2023 IEEE

that the proposed DR scheme learns more robust features that exhibit non-linear properties, resulting in improved detection accuracy of the classifier.

The rest of the paper is organized as follows. Section II explains theoretical modelling of stealthy data integrity assault and DAE. The proposed DAE-based DR and SVM-based attack detection schemes are described in section III. The simulation results are given in section IV. Lastly, the paper is concluded in section V.

II. STEALTHY DATA INTEGRITY ASSAULT AND DAE MODEL

A. SDIA Model

In launching SDIA, an intelligent hacker attempts to impart fictitious values in CPS measurements collected by sensors on wireless communication links. We assume that adversary is fully aware of SG network and becomes successful in dodging the Bad Data Detector (BDD) and bypass the operator at Power Control Centre (PCC). State Estimation (SE) is an online monitoring of system states and approximation of power system state variable Π=$[\Pi_1, \Pi_2, \Pi_3, ..., \Pi_n]^{\mathrm{T}}$, based on RTU measurements X=$[x_1, x_2, x_3, ..., x_m]^{\mathrm{T}}$, where n and m are positive integers, and $\Pi_i, x_j \in \mathbb{R}$ for $i = 1, 2, ..., n$ and $j = 1, 2, ..., m$. In AC power flow, the state variables and measurements are related as:

$$X = h(\Pi) + \upsilon, \tag{1}$$

Where h(Π) shows non-linearity between X and Π, υ is a Gaussian matrix υ=$[\upsilon_1, \upsilon_2, \upsilon_3, ..., \upsilon_n]^{\mathrm{T}}$. For DC power flow model, (1) can be further simplified to:

$$X = H\Pi + \upsilon. \tag{2}$$

Where H is Jacobian Matrix composed of impedance data and topology only.

The difference between the observed measurements X and estimated measurements $\hat{X}$ gives residual R, which forms basis for BDD in current power systems, given as:

$$R = X - \hat{X} = X - H\hat{\Pi}. \tag{3}$$

The intrusion detection in BDD is done using L-test proposed by Monticelli [14] which uses largest normalized residual with predefined threshold value Λ. Therefore, the assumption of data being attacked is based on following condition to be false:

$$\max_i |R_i| < \Lambda, \tag{4}$$

Where R_i is the component of vector R.

During the attack, the assailant cracks the state variables to inject false values by changing real power flows and real power injections. Let the foe makes attack vector $X_{assault}$ = X + ϕ, where ϕ = Hc, is the false value infused in the measured data X. Such an attack could not be detected by BDD and bypasses the operator at PCC leading to successful launch of attack in the system. For example, if the malicious user intends to change the variable x_1 by falsifying the measurements by 5%, crafts a vector c considering (5) as given:

$$c = [-0.05x_1, 0, ..., 0]. \tag{5}$$

The compromised measurements are then calculated as given in (6) by employing power flow equations and state vector $\Pi\epsilon = \hat{X} + c$:

$$X_\epsilon = H\Pi_\epsilon + \upsilon. \tag{6}$$

B. Working Principle of DAE Model

Deep-denoising autoencoder is a variant of autoencoder used to tackle the curse of high dimensionality suitable for non-linear data as in SG. The number of input and output features remains same, however, DR is achieved by projection of high dimension space to low dimension space. As an extension of AE, DAE is more intelligent to learn the prominent features, map more informative latent space for DR leading to the robust handling of assailed or corrupted data. Zero-masking noise (ZDAE) and additive Gaussian noise (GDAE) are the two basic preferences of DAE for corruption addition in the data. DAE model consists of three basic parts: encoder, latent-space and decoder, given in Fig. 1. Using non-linear mapping given in (7), encoder translates the corrupted data , into latent space ν, instead of original input data y.

$$\nu = f\left(w_1 y + b\right), \tag{7}$$

where f(.) is non-linear activation function, w_1 is the weight matrix and b is the optimized tor. Utilizing non-linaer transformation at the output layer, the decoder cracks the latent space into reconstructed vector $\hat{y}$ as follows:

$$\hat{y} = g\left(w_1 y + c\right), \tag{8}$$

where g(.) is the non-linear activation function.

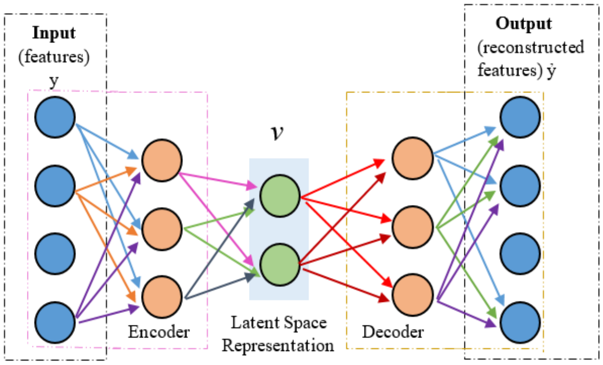

Fig. 1. Basic deep-denoising autoencoder model

III. PROPOSED DAE-BASED DIMENSIONALTIY REDUCTION AND SVM-BASED DETECTION SCHEMES

Power system network transmits the power generated at the grid to the consumer end through transmission network. The sensors installed at various points in electric power network collect and send the measured data over wireless channel to PCC. The adversary may infuse biased values in the transmitted data compromising the integrity of the data. After the data is received at PCC, the DAE attempts to obtain latent space representation which is then inputted to SVM-based model to identify SDIA. The layout of the proposed scheme is given in Fig. 2.

Two familiar approaches of inducing attack to the DAE model are ZDAE and GDAE. In this study, we introduce a new approach, Eclectic corruption addition scheme (EDAE), for addition of corruption to the measured data where noise is induced considering Gaussian distribution with mean and variance fetched by evaluating SG data.

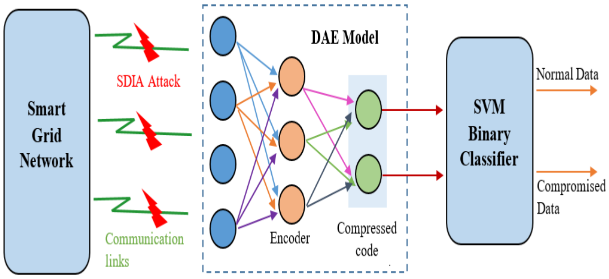

Fig. 2. DAE-based DR for SDIA detection in SG

A. Proposed EDAE Corruption Addition Scheme

In the proposed EDAE scheme, noise or corruption is added during model training to the dataset $Z = \{z_1, z_2, z_3, ..., z_m\}$, where m is the number of data samples. The z_i is a data sample consisting of n features, $z_i = \{fe_{i1}, fe_{i2}, ..., fe_{in}\}$. The infused corruption Π is normal distribution $\mathcal{N}(\varrho, \sigma)$, where ϱ is a vector of mean values $\varrho = \{\varrho_1, \varrho_2, \varrho_3, ..., \varrho_n\}$ and σ is the variance vector $\sigma = \{\sigma_1, \sigma_2, \sigma_3, ..., \sigma_n\}$. Subsequently, the training data is obtained as $Z_0 = Z + \Pi$.

B. Proposed SVM-based Attack Detection

Once the DAE model is trained and compressed latent space code is obtained, it is given as an input to support vector machine binary classifier as the detection of SDIA in SG in binary classification problem. SVM distinguishes the two classes in the dataset employing largest margin hyperplane segregation technique in the feature space of the training dataset. To identify the test data samples as normal or compromised, sign of hyperplane is employed. Gaussian radial basis function is used as kernel function in this study. The classification function for SVM is as follows:

$$F(Z) = \operatorname{sgn}\{f(z)\} \tag{9}$$

The function f(z) ranges from -∞ to +∞ representing signed distance of unspecified sample from decision boundary. A positive sign indicates sample belongs to normal data whereas negative sign suggests otherwise.

IV. SIMULATION RESULTS

Various IEEE standard test cases as IEEE-14, -39 and -57 bus systems are employed to validate the effectiveness of the proposed scheme, using Matpower 7.0 toolbox. The measurement features in the data set include active power flows in the transmission lines and active power injections into the buses. The attack design includes static approach assuming that assailant is dormant with limited access to the sensors; and nomadic approach where the adversary is mobile and can randomly access various sensros or meters. For training of DAE model, 50% of the measured data was used while SVM took 75% data samples for training employing 4-fold cross validation. The proposed model is constructed using standard IEEE 14-, 39-, 57-bus systems, however, to limit the length of paper we show results for the standard IEEE 14-bus system only.

Simulation results of the proposed EDAE scheme and its comparison with existing ZDAE and GDAE are given in Tables I-VI. It can be seen that the EDAE outperforms the existing schemes and reconstructs the features with minimum error. This gives an understanding that a fine robust latent space is apprehended that can be fed to SVM-model for identification of SDIA. Note that dataset contains 200,000 samples, and all are well-reconstructed close to the original feature value, however, constrained to the paper length and for clear understanding, three finest features in each case are presented here.

TABLE I
THREE FINEST RECONSTRUCTED FEATURES (STATIC ATTACK)

EDAE-based Reconstruction Scheme			
Feature Number	***18***	***38***	***49***
Feature Value	50.4	-42	-6.8649
Reconstructed Value	5.04 x 10^1	-4.20 x 10^1	-6.86 x 10^0
Error	0.00010462	5.18 x 10^{-5}	6.47 x 10^{-5}
Error Ratio	6.09 x 10^{-6}	6.75 x 10^{-7}	1.12 x 10^{-6}

TABLE II
THREE FINEST RECONSTRUCTED FEATURES (STATIC ATTACK)

ZDAE-based Reconstruction Scheme			
Feature Number	***09***	***15***	***19***
Feature Value	-10.608	73.443	-5.7354
Reconstructed Value	-10.59362174	73.46094698	-5.729508174
Error	0.01437826	0.017946978	0.005891826
Error Ratio	0.001355417	0.000244366	0.001027274

Performance evaluation of the classification model is basaed on analyzing standard error metrics such as accuracy, F1-score and ROC curve. Fig. 3 and Fig. 4 show accuracy score of the proposed scheme DAE + SVM in comparison with various existing schemes such as Genetic Algorithm (GA)+SVM and Principal Component Analysis (PCA)+SVM, for static and nomadic attacks. It is clearly visible that the proposed scheme performed well and achieved higher accuracy in comparison with other schemes. The accuracy score of

TABLE III
THREE FINEST RECONSTRUCTED FEATURES (STATIC ATTACK)

GDAE-based Reconstruction Scheme			
Feature Number	***25***	***27***	***47***
Feature Value	17.551	28.447	-28.447
Reconstructed Value	1.76 10^{1}	2.84 10^{1}	-2.84 10^{1}
Error	0.001103067	0.0006857	0.0002416
Error Ratio	6.28 10^{-5}	2.41 10^{-5}	8.50 10^{-6}

TABLE IV
THREE FINEST RECONSTRUCTED FEATURES (NOMADIC ATTACK)

EDAE-based Reconstruction Scheme			
Feature Number	***14***	***29***	***43***
Feature Value	73.554	5.6798	-6.7198
Reconstructed Value	7.36 10^{1}	5.68 10^{0}	-6.72 10^{0}
Error	0.000534421	6.26 10^{-5}	0.000406535
Error Ratio	7.27 10^{-6}	1.10 10^{-5}	6.05 10^{-5}

the proposed scheme, DAE+SVM is 4.3 % and 0.6% higher as compared to PCA+SVM and GA+SVM respectively, for static attack. In case of nomadic attack, the accuracy score of the proposed scheme performed is 91.662 whereas GA+SVM and PCA+SVM achieved 90.545 and 89.656 accuracy, respectively.

Receiver operating curve (ROC) is another well-known performance evaluation metric in classification problems. It is a plot between false positive rate and false negative rate defining sensitivity and specificity of model. As close the value of area under ROC is to unity, as effective the model is. Fig. 5 and Fig. 6 show the ROC curves obtained for the proposed and existing schemes. It could be clearly seen that the area under curve of the proposed scheme is almost equal to one, confirming its efficiency over other schemes.

TABLE V
THREE FINEST RECONSTRUCTED FEATURES (NOMADIC ATTACK)

ZDAE-based Reconstruction Scheme			
Feature Number	***18***	***24***	***38***
Feature Value	-24.313	7.734	24.313
Reconstructed Value	-24.31091173	7.739212442	24.31673587
Error	0.002088271	0.005212442	0.003735869
Error Ratio	8.589 10^{-5}	0.000673965	0.000153657

TABLE VI
THREE FINEST RECONSTRUCTED FEATURES (NOMADIC ATTACK)

GDAE-based Reconstruction Scheme			
Feature Number	***13***	***22***	***42***
Feature Value	153.53	43.836	-43.836
Reconstructed Value	153.5325412	43.83835556	-43.83366463
Error	0.00254121	0.002355563	0.002335371
Error Ratio	1.655 10^{-5}	5.373 10^{-5}	5.327 10^{-5}

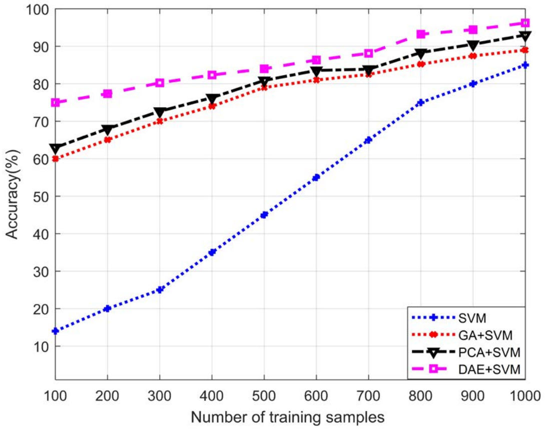

Fig. 3. Accuracy comparison of the proposed and existing schemes (static attack)

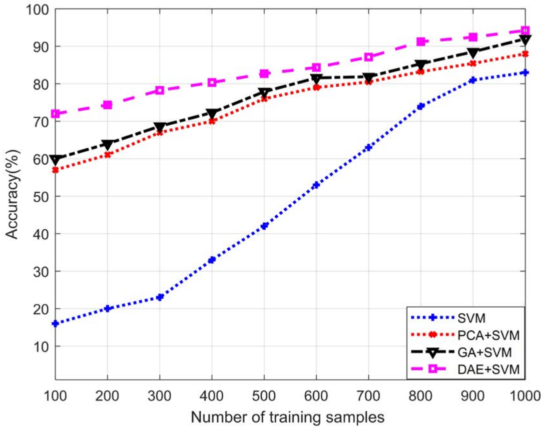

Fig. 4. Accuracy comparison of the proposed and existing schemes (nomadic attack)

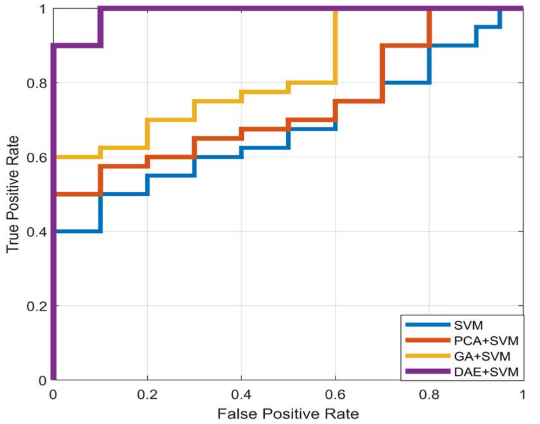

Fig. 5. ROC comparison of the proposed and existing schemes (static attack)

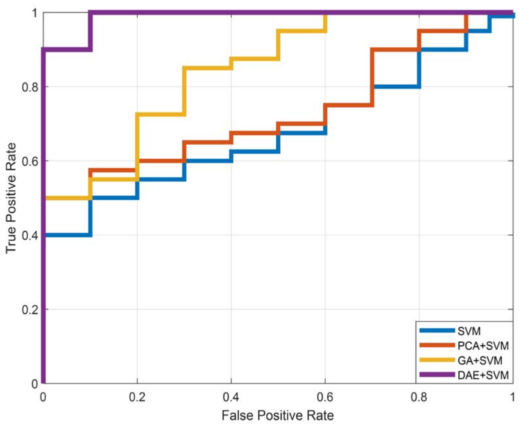

Fig. 6. ROC comparison of the proposed and existing schemes (nomadic attack)

V. Conclusion

This paper presents a deep DAE-based scheme to address the curse of dimensionality that grows with the expanding size of the power system and pulls a robust latent space code from the SG data set. Next, the latent space representation code is fed an SVM model to identify SDIA. We compared the proposed scheme with existing approaches such as GA+SVM, PCA+SVM, and simple SVM. The results reveal that the suggested DAE+SVM-based approach exhibits promising identification efficiency schemes under intermittent functioning conditions. Subsequently, the DAE+SVM performs better for SDIA detection in SG-based cyber-physical system communications networks.

Acknowledgment

This work was supported in part by the MSIT, Korea, under the ICAN program (IITP-2023-RS-2022-00156326) supervised by the IITP and in part by the NRF funded by the Korea government (MSIT) (No. 2022R1F1A1074556).

References

[1] A. A. Jamal, A. A. M. Majid, A. Konev, T. Kosachenko, and A. Shelupanov, “A review on security analysis of cyber physical systems using machine learning,” *Materials Today: Proceedings*, vol. 80, pp. 2302-2306, Apr. 2023.

[2] A. Rostami, M. Mohammadi, and H. Karimipour, “Reliability assessment of cyber-physical power systems considering the impact of predicted cyber vulnerabilities,” *International Journal of Electrical Power & Energy Systems*, vol. 147, p. 108892, May 2023.

[3] A. Raza, Memon, S., M. A. Nizamani, and M. H. Shah, “Machine Learning-Based Security Solutions for Critical Cyber-Physical Systems,” In Proc. *International Symposium on Digital Forensics and Security (ISDFS)*, June 2022.

[4] M.E. Sahin, and F. Muheidat, “The security concerns on cyber-physical systems and potential risks analysis using machine learning,” *Procedia Computer Science*, vol. 201, pp. 527-534, 2022.

[5] R. Qi, C. Rasband, J. Zheng, and R. Longoria, “Detecting cyber attacks in smart grids using semi-supervised anomaly detection and deep representation learning,” *Information*, vol. 12, no. 8, p. 328, Aug. 2021.

[6] S. Sengan, S. V, I. V, and L. Ravi, “Detection of false data cyber-attacks for the assessment of security in smart grid using deep learning,” *Computers and Electrical Engineering*, vol. 93, p. 107211, July 2021.

[7] A. Takiddin, M. Ismail, U. Zafar, and E. Serpedin, “Deep autoencoder-based anomaly detection of electricity theft cyberattacks in smart grids,” *IEEE Systems Journal*, vol. 16, no. 3, pp. 41064117, Jan. 2022.

[8] A. Saddam, M. Irshad, S. A. Haider, J. Wu, D. N. Deng, and S. Ahmad, “Protection of a smart grid with the detection of cyber-malware attacks using efficient and novel machine learning models,” *Frontiers in Energy Research*, vol. 10, p. 1102, Aug. 2022.

[9] S. Ahmed, Y. Lee, S.-H. Hyun, and I. Koo, “Mitigating the impacts of covert cyber attacks in smart grids via reconstruction of measurement data utilizing deep denoising autoencoders,” *Energies*, vol. 12, no. 16, p. 3091, Aug. 2019.

[10] M. Hariharan, K. Polat, and R. Sindhu, “A new hybrid intelligent system for accurate detection of parkinsons disease,” *Computer Methods and Programs in Biomedicine*, vol. 113, no. 3, pp. 904913, Mar. 2014.

[11] Y. Liu, Y. Li, X. Tan, P. Wang, and Y. Zhang, “Local discriminant preservation projection embedded ensemble learning based dimensionality reduction of speech data of parkinsons disease,” *Biomedical Signal Processing and Control*, vol. 63, p. 102165, Jan. 2021.

[12] S. Ahmed, Y. Lee, S.-H. Hyun, and I. Koo, “Covert cyber assault detection in smart grid networks utilizing feature selection and euclidean distance-based machine learning,” *Applied Sciences*, vol. 8, no. 5, p. 772, May 2018.

[13] 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 Transactions on Information Forensics and Security*, vol. 14, no. 10, pp. 27652777, Mar. 2019.

[14] A. Monticelli, *State Estimation in Electric Power Systems: A Generalized Approach,* Norwell, MA, USA: Springer, 1999.