Cyber Attacks on Smart Energy Grids Using Generative Adverserial Networks

Saeed Ahmadian[♭], Heidar Malki[*] and Zhu Han[♭]
[♭]Department of Electrical and Computer Engineering, University of Houston, TX USA
[*]Department of Engineering Technology, University of Houston, TX USA

***Abstract*—Recently, cyber-attacks to smart energy grid has become a critical subject for Energy System Operators (ESOs). To keep the energy grid cyber-secured, the attacker's behavior, resources and goals must be modeled properly. Then, the counter-measurement actions can be designed based on the attacker's model. In this paper, a new zero-sum game based on the Generative Adversarial Networks (GANs) is presented. The attacker to energy smart grid pursues two objects. The first goal is to be undetected by the system defender, and the second goal is to make profit via its False Data Injection (FDI) into the system. On the defender hand, the ESO needs to detect FDI using fast and reliable models. Thus, each party tries to defeat the other part. GANs are deep layer networks which consist of two rivals networks: the generator and the discriminator. The Generator Network (GN) aims to create the data similar to real data (plays the attacker role) and the Discriminator Network (DN) wants to properly detect whether the data is real or faked by the GN (plays the system defender role). In this paper, a new algorithm is presented to model both the attacker and ESO in the GAN frame work. Finally, A five-bus smart grid case is considered to show the effectiveness of the presented algorithm.**

***Keywords*—Cyber-security, smart grid, false data injection, Generative Adverserial Networks (GANs).**

I. Introduction

Securing the cyber-physical systems against the attackers is one of the freshly and critically concerns of the System Operators (SOs) and users [1]–[3]. The most common attack to smart grid is to compromise the measurement systems such that the attacker takes advantage of the electricity market interactions and electricity price fluctuations [4]. Considering the current literature works, the successful attack happens when the FDI causes the transmission line congestion [5], [6]. Thus, based on this objective, different FDI detection methods are introduced. Using Supported Vector Machine (SVM) in [7], the attack signals are linearly classified. Using power system voltage angles Markov graph, a new statistical detection method is offered in [8]. AC state estimation of the energy systems and the corresponding cyber-attacks are also presented in [9].

Recently, Deep Neural Networks (DNNs) and data feature mapping via deep layers are well-attended in the research papers [10]–[12]. Using DNNs, the data is transformed into new feature space (latent space) with high dimension. In fact, new representation of given data in a high dimension space is obtained via DNNs. DNNs are mainly used in different applications. In [13]–[17], deep representation of input data is obtained to classify images into specific groups. Indeed, using Convolutional Neural Networks (CNNs) and convolving different Kernel filters in each layer, the non-linear space of data is turned into a linear one.

Recently, the Generative Adversarial Networks (GANs) have attracted researchers' attentions due to their broad application. Initially the GANs were introduced in [18] as unsupervised DNNs. Basically the GANs are consists of two networks that compete against each other. The Generative Network (GN), is fed by Gaussian noise and tries to generate the data very similar to real data. The GN's objective function is to maximize its rival's error in detection of faked data (generated by the GN) from real data. In other word, the GN wants to deceive its rival the Discriminative Network (DN) which is very similar to the attacker's goal in the energy smart grids. On the other side, the DN wants to minimize its discrimination error between fake and real data. The DN is given both generated data by the GN and real data, and is supposed to label them correctly. This is very close to ESO' role in the energy grids.

In this paper, using GAN, a new structure to model the attacker and the ESO is presented. The attacker desires to make profit via the electricity market transactions. Thus, if the attacker knows what types of the data is required to inject into the measurement system, the rest of the process is the matter of the encryption and decryption techniques in the a Man-In-The-Middle attack (MITM) [19]. In the Real-Time (RT) electricity market, The ESO receives the RT data from different measurement devices in the grid. Then, using the power system state estimator, the state variables to calculate the RT electricity prices are obtained [5]. Traditionally, if the residual value of the state estimator is sufficiently small enough, the data is considered as real data. Thus, if the attacker uses the real data features to generate the new one, it can easily fool the traditional detector. To prevent the attacker to use the GN and fake the ESO, a real-time DN model is presented in this paper that is trained with GN simultaneously and learn the attacker's behaviour. In fact, the DN is fed with provided information from the attacker and real data from the measurement system to help ESO find FDI attacks.

The rest of the paper is organized as follows. In Section II, the electricity market principles and the process of market clearing price is discussed. Then, using electricity market information, the desired information which attacker wants to fake is explained. In Section III, The GANs and presented structure for cyber-attack modelling is introduced. In Section IV, the presented model is implemented in a five bus system and the results are analyzed. Finally, Section V draws the conclusion for offered cyber-security model.

II. Electricity Markets and The Attacker's Desired Information

The FDI attack process is depicted in Fig. 1. The attacker injects FDI into the Remote Terminal Units (RTUs) or generally any measurements devices in the grid. Then these data are passed into the supervisory control and Data acquisition (SCADA) system. The SCADA sends the data to

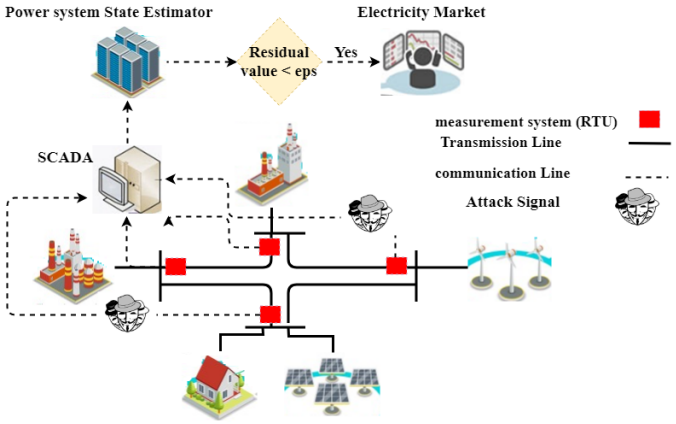


Fig. 1: The Attack process in power system

the power system state estimator. If the residual value of the state estimator is less than a specified threshold, the data is used to calculate electricity prices. Therefore, if the attacker can simulate the process of the power system state estimator and electricity market clearing process, it can easily obtain the desired input which causes high electricity prices. Thereafter, using the GN the desired input can be produced by the attacker.

In this section, the electricity market and the power system state estimation process is first introduced. Then considering these two processes (subsections II-A and II-B), the attackers' objective function is presented in subsection II-C.

A. Power System Electricity market

ESOs are responsible to calculate electricity prices for the electricity markets. The attacker desires to manipulate the RTUs' output so that the possibility of the transmission lines congestion increases. Because due to the lines' congestion, the electricity price difference between two connected buses via congested line will increase. Thus, the attacker can buy the electricity in the Day-Ahead market [20], [21] and by injecting malicious data into the RTUs, increase the electricity price in the RT market and consequently sell its own electricity. Therefore, the optimization problem in (1) is implemented by the ESO to obtain the RT prices, i.e.,

$$\min_{\lambda'_{nt},\Delta p_{it},\Delta d_{jt}} C_{RT} = \left[\sum_{i=1}^{i=N_g}\sum_{t=1}^{t=T} C'_{it}\Delta p_{it}\right] \tag{1a}$$

s.t.

$$\sum_i \Delta p_{it} - \sum_j \Delta d_{jt} = 0, \tag{1b}$$

$$\Delta p_{it}^{min} \le \Delta p_{it} \le \Delta p_{it}^{max}; \forall i \in N_g, \tag{1c}$$

$$\Delta d_{jt}^{min} \le \Delta d_{jt} \le \Delta d_{jt}^{max}; \forall j \in N_d, \tag{1d}$$

$$\sum_{n=1}^{N_{bus}} GSF_{n-k}\left(\Delta p_{nt} - \Delta d_{nt}\right) \le 0; \forall k \in N_{line}. \tag{1e}$$

In (1a), C'_{it} is the marginal cost of the i^{th} generator which produce Δp_{it} to compensate mismatch demand Δ_{jt} in the RT market. Eq. (1b) shows the incremental balance between the generation and demand. Eqs. (1c) and (1d) depict the incremental generation and dispatch-able loads upper and lower bounds, respectively. Eq. (1e) indicates that the line flow increment must not increase to make the congestion worse. The RT electricity price for different nodes (buses) in the grid, depends on the Lagrangian multipliers of (1b) and (1e). The RT price for the n^{th} bus at time t is given by (2).

$$\lambda_{nt}^{RT} = \pi + \sum_{k=1}^{N_{line}} GSF_{n-k}\Delta\zeta_{nk} \tag{2}$$

In (2), π is the Lagrangian multiplier correspond to Eq. 1b. In fact, π shows the shadow price of total load and generation balance in the RT market. Moreover, $\Delta\zeta_{nk}$ is the dual variable corresponding to Eq. (1e). It shows the shadow price of mismatch between load and generation in each bus.

B. Power System State Estimator

Considering the measurement vector as $\mathbf{Z}$, the power system Jacobian matrix as $\mathbf{H}$, the power system state variables as $\mathbf{X}$, the measurement error $\mathbf{e}$ with a zero mean and covariance matrix $\mathbf{C}$ in a Gaussian distribution, we have

$$\mathbf{Z} = \mathbf{HX} + \mathbf{e}. \tag{3}$$

The state estimator in the electricity grids finds the optimal state vectors using $\hat{\mathbf{X}} = \mathbf{argmin}_{\bar{\mathbf{X}}}\mathbf{E}\left\|\mathbf{X} - \hat{\mathbf{X}}\right\|_2^2$. The optimal value for $\hat{\mathbf{X}}$ is $\hat{\mathbf{X}} = \left(\mathbf{H'C^{-1}H}\right)'\mathbf{H'C^{-1}Z} = [\mathbf{M}]\,\mathbf{Z}$. Thus, the residual value of state estimator is given by

$$\|\mathbf{r}\|_2 = \left\|\mathbf{Z} - \mathbf{H}\hat{\mathbf{X}}\right\|_2 = \|\mathbf{Z} - \mathbf{HMZ}\|_2 \le \epsilon. \tag{4}$$

C. Attacker's objective in the market

The attacker's desired data are the RT measurements in (3), which cause the most difference between the RT price in (2) and the DA price (the DA prices are assumed to be constant in this paper). Moreover, the attacker needs to use the measurements that meet the condition in (4). Therefore, given set of possible measurements $\{z_1, z_2, ..., z_n\} \in \mathbf{Z}$, the attacker's optimization problem is presented as follows

$$\max_{z^{att}\in\mathbf{Z}} \mathbb{E}\left\|\lambda_{nt}^{RT}\right\|_2 \tag{5a}$$

$$\|\mathbf{r}\|_2 \le \epsilon. \tag{5b}$$

III. Cyber Attacks and Counter-measurements

In this section, first the GANs are introduced. Then, using the GAN structure, a new cyber-security model is presented in which the attacker by implementing the optimization problem in (5) can obtain the required information to compromise the RTUs.

A. Generative Adversarial Networks (GANs)

GANs are consists of two networks: the one (GN) aims to generate data close to reality and the other (DN) which detects faked from original data and updates the GN parameters. Considering x as real data, y as noise given to GN, $p_D(x)$ as the DN' distribution over x and $p_g(y)$ as GN's distribution over y, the zero sum game between these two network can

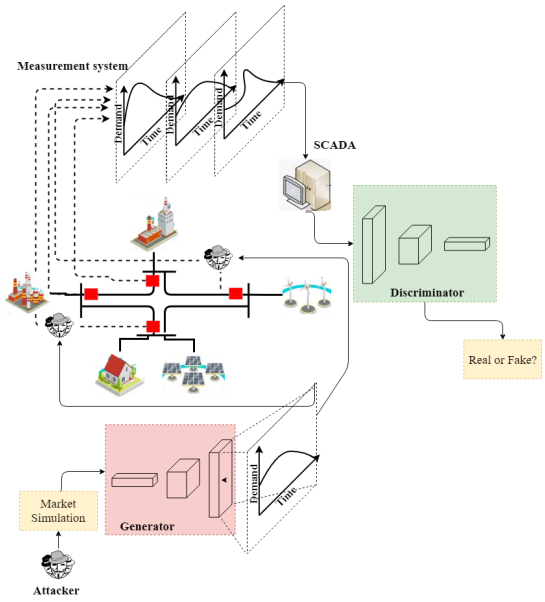


Fig. 2: Presented Cyber-security model using GAN.

be formulated as the following min-max optimization problem [18]

$$\min_{G}\max_{D}\mathbb{E}_{x\sim p_{data}(x)}\left[\log D(x)\right]+\mathbb{E}_{y\sim p_y(y)}\left[\log(1-D(G(y)))\right]. \tag{6}$$

Here, $G(y,\Theta_y)$ is the mapping function from y space into data space of $G(y)$ and Θ_y represents GN' parameters. In fact, $G(y,\Theta_y)$ generates the attack signal from noise y. Therefore, when the attacker solves Eq. (6) and finds the generator parameters (Θ_y), then it can easily generates new attack signal and injects that into energy measurement system by just inputting noise to G. Meanwhile, $D(x,\Theta_x)$ maps input data x into $D(x)$ space which is a scalar value (probability of being real or fake) with Θ_x as the DN's parameters. Based on the optimization problem in (6), the DN wants to maximize the probability of correct data labeling.

B. Presented Cyber-Security Model

Fig. 2 shows the presented cyber-security model in the power system based on the GAN structure. The attacker obtains the desired type of FDIs based on the results from (5). Then using the GN, the desired FDI's features are extracted and a new attack signal is generated that can easily bypass the power system state estimator. These malicious data along with the real data (non-compromised RTUs) are gathered by the SCADA system and fed to the DN. The DN is in the loop with the GN and is trained simultaneously to discriminate the real data (z_{real}) from the attacker's fake data ($G(z_{att})$).

Considering Fig. 2 and the measurement vector in (3) as the set of non-compromised and the attacked RTUs ($\{z_{real}, G(z_{att})\}$), the problem in (6), is reformulated

$$\min_{G}\max_{D}\mathbb{E}_{z_{real}\sim p_{data}(z_{real})}\left[\log D(z_{real})\right]+ \mathbb{E}_{y\sim p_y(y)}\left[\log(1-D(G(z_{att})))\right] \tag{7a}$$

s.t.

$$z_{att}\in\left\{\begin{array}{c} argmax\mathbb{E}\left\|\lambda_{nt}^{RT}\right\|_2 \\ \text{s.t} \\ \left\|\mathbf{r}\right\|_{\mathbf{2}}\leq\epsilon \end{array}\right\}. \tag{7b}$$

In (7a), z_{real} are the real data and $G(z_{att})$ are the manipulated data generated from z_{att}. The ESO (DN) aims to maximize labeling process of real measurements as one ($\log D(z_{real})$) and the attacked measurements as zero ($\log(1-D(G(z_{att})))$). While the attacker wants to manipulate the measurement system to maximize ($\log D(z_{att})$). The advantage of the presented model over the existing detection method is described in (7b). It shows that the generated signal by the attacker already can bypass the traditional detection method in (4) and the residual value of the injected fake signal to the measurement system, is low enough to pass through the system. Thus, the discriminator doesn't play the role of traditional detection method. In particular, the discriminator is a secondary evaluation detecting system which acts over bypassed signal from traditional residual detector. The step by step procedure to implement the presented cyber-security model in (7) is summarized as follows.

1) Use the online information of energy consumption provided by the ESO to obtain the real data (z_{real}) distribution.

2) Use z_{real} in the optimization problem in (5) to obtain the z_{att}. This step requires some estimation of the network's parameters. In fact, if the attacker has some knowledge about energy network parameters can easily run (5). In other word, Eqs (1a) to (1e) provide initial point for the attacker to start with. If the attacker doesn't have the estimation about the electricity grid's specifications, still can use the optimization in (5). It should use the online information provided by the ESO on its website. In this case, the attacker only needs to match the demand data with the RT prices and sort them by using the objective function in (5). Therefor the best real data (z_{real}) which maximize (5a) are considered as (z_{att}).

3) Repeat steps 1 and 2 for N iterations of required batches for the GAN training.

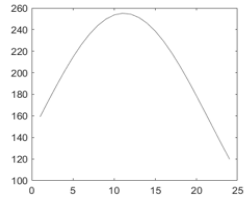
4) Start the GAN training to update the generator and the Discriminator together. After the training process the output of the generators create fake demand signal which can fool the discriminator and the discriminator must try to label demand data correctly.

IV. Simulation and Results

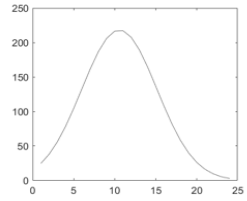
In this section, the simulation results for a five-bus test case system [5] are presented. The possible compromised measurements are considered to be the demand at any of the buses. Since there are 3 demands at buses 2, 3 and 4, the attackers needs to know the feature of all three of them. Considering the normal distribution for the real data, the result for the first set of the attack data (z_{att}) using the optimization problem in (5) is depicted in Fig. 3. Since, the measurements data (real data) are published by the ESO in public access (online websites), therefor they can be easily obtained and get processed for presented structure. In following the results of simulation for the attacker and the DN (ESO) is presented.

A. Attacker Results

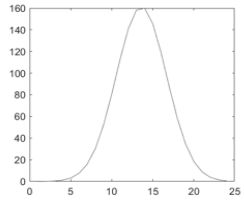
The attacker's objective function in (7) is depicted in Fig. 4. The data set for this simulation has 2000 samples in batch of 16. Due to lack of data the model has trained twice with shuffling the samples. As it is depicted, the loss function is decreased in total, and the attacker has produced better attack



(a) Energy demand at bus 2.



(b) Energy demand at bus 3.



(c) Energy demand at bus 4.

Fig. 3: Desired data (z_{data}) using a normal distribution for real data (z_{real}).

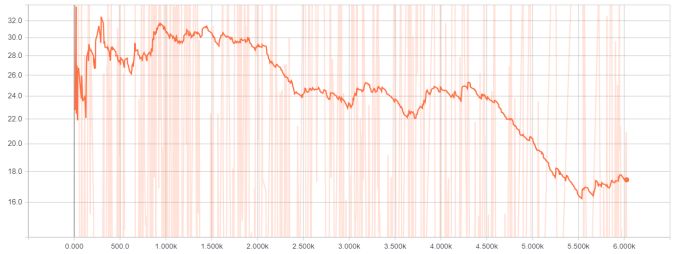


Fig. 4: Attacker's objective function in (7).

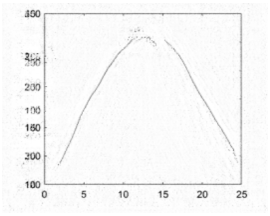
signals after each iteration. However, using the bigger data set will cause more accurate attack signals. In the figure, the loss function at the beginning is about 32 and at the end of the training is about 16. The generated attack signal by the attacker (GN) are also depicted in Fig. (5). Since the GN loss function is not small enough, it is clear that the GN still needs more training steps with more diverse data samples.

B. ESO (DN) Results

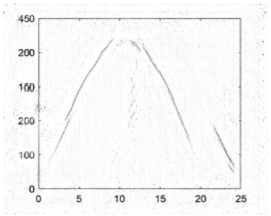
The ESO desire to detects FDI accurately and very fast. Using the proposed method, The DN's losses ($\mathbf{E}\left[\log(D(z_{real}))\right]$ and $\mathbf{E}\left[\log(1-D(G(z_{att})))\right]$) when the DN is fed with the real and fake data are depicted in figs. (6a) and (6b) respectively. The loss value for the DN fed by the real data is about 0.504 while the DN's loss value when it is fed by the fake data is 0.093. It means that the DN detects FDI very well and with high accuracy. Surely, the major reason for that is due to weak faked signal by the attacker. Meanwhile, it must be mentioned that the DN might detects the real data as faked data due to its loss value when it is fed by the real data.

V. Conclusion

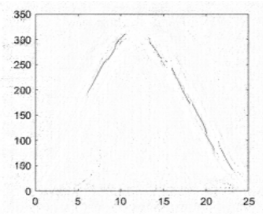
In this paper a new Cyber-security model based on the GAN structure is presented. The attacker is considered to play the generative network role and the ESO is the discriminative network. The attacker, considering the proposed optimization



(a) Energy demand at bus 2.

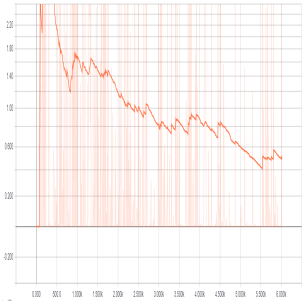


(b) Energy demand at bus 3.

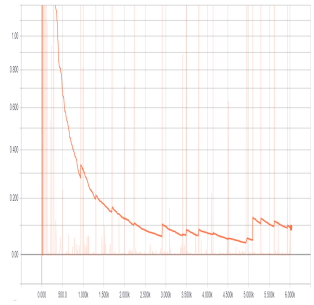


(c) Energy demand at bus 4.

Fig. 5: Generated data by the attacker ($G(z_{att})$).



(a) The DN's loss function when it is fed with real data ($\mathbf{E}\left[\log(D(z_{real}))\right]$).



(b) DN's loss function,fed with fake data from the attacker ($\mathbf{E}\left[\log(1-D(G(z_{att})))\right]$).

Fig. 6: DN's loss functions.

problem in this paper, produces the fake data that not only bypass the power system state estimator but also make profit in the electricity market. On the other hand, the DN by extracting the attacker's data features tries to discriminate the FDI attacks. The proposed model tested on a 5-bus energy smart grid and the results shown the small amount of loss function for the DN that proves the effectiveness of proposed model.

References

[1] M. M. Pour, A. Anzalchi, and A. Sarwat, "A review on cyber security issues and mitigation methods in smart grid systems," in *SoutheastCon 2017*, Charlotte, NC, Mar. 2017.

[2] J. Hao, R. J. Piechocki, D. Kaleshi, W. H. Chin, and Z. Fan, "Sparse Malicious False Data Injection Attacks and Defense Mechanisms in Smart Grids," *IEEE Transactions on Industrial Informatics*, vol. 11, no. 5, pp. 1–12, Oct. 2015.

[3] Q. Yang, J. Yang, W. Yu, D. An, N. Zhang, and W. Zhao, "On False Data-Injection Attacks against Power System State Estimation: Modeling and Countermeasures," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 3, pp. 717–729, Mar. 2014.

[4] L. Xie, Y. Mo, and B. Sinopoli, "Integrity Data Attacks in Power Market Operations," *IEEE Transactions on Smart Grid*, vol. 2, no. 4, pp. 659–666, Dec. 2011.

[5] M. Esmalifalak, H. Nguyen, R. Zheng, L. Xie, L. Song, and Z. Han, "A Stealthy Attack Against Electricity Market Using Independent Component Analysis," *IEEE Systems Journal*, pp. 1–11, 2017.

[6] G. Liang, S. R. Weller, F. Luo, J. Zhao, and Z. Y. Dong, "Generalized FDIA-Based Cyber Topology Attack with Application to the Australian Electricity Market Trading Mechanism," *IEEE Transactions on Smart Grid*, Early Access 2017.

[7] M. Esmalifalak, L. Liu, N. Nguyen, R. Zheng, and Z. Han, "Detecting Stealthy False Data Injection Using Machine Learning in Smart Grid," *IEEE Systems Journal*, vol. 11, no. 3, pp. 1644–1652, Sep. 2017.

[8] H. Sedghi and E. Jonckheere, "Statistical Structure Learning to Ensure Data Integrity in Smart Grid," *IEEE Transactions on Smart Grid*, vol. 6, no. 4, pp. 1924–1933, Jul. 2015.

[9] 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, Sep. 2016.

[10] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, p. 436, 2015.

[11] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*. MIT press Cambridge, 2016, vol. 1.

[12] A. Mobiny and H. Van Nguyen, "Fast capsnet for lung cancer screening," *arXiv preprint arXiv:1806.07416*, 2018.

[13] A. Mobiny, S. Moulik, and H. Van Nguyen, "Lung cancer screening using adaptive memory-augmented recurrent networks," *arXiv preprint arXiv:1710.05719*, 2017.

[14] T.-H. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, and Y. Ma, "Pcanet: A simple deep learning baseline for image classification?" *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5017–5032, 2015.

[15] H. Wu and S. Prasad, "Semi-supervised deep learning using pseudo labels for hyperspectral image classification," *IEEE Transactions on Image Processing*, vol. 27, no. 3, pp. 1259–1270, 2018.

[16] G. Cheng, C. Yang, X. Yao, L. Guo, and J. Han, "When deep learning meets metric learning: Remote sensing image scene classification via learning discriminative cnns," *IEEE Transactions on Geoscience and Remote Sensing*, 2018.

[17] A. Mobiny and M. Najarian, "Text-independent speaker verification using long short-term memory networks," *arXiv preprint arXiv:1805.00604*, 2018.

[18] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[19] F. Callegati, W. Cerroni, and M. Ramilli, "Man-in-the-middle attack to the https protocol," *IEEE Security & Privacy*, vol. 7, no. 1, pp. 78–81, 2009.

[20] S. Ahmadian, B. Vahidi, J. Jahanipour, S. H. Hoseinian, and H. Rastegar, "Price restricted optimal bidding model using derated sensitivity factors by considering risk concept," *IET Generation, Transmission & Distribution*, vol. 10, no. 2, pp. 310–324, Feb. 2016.

[21] S. Ahmadian, H. Malki, and A. R. Sadat, "Modeling time of use pricing for load aggregators using new mathematical programming with equality constraints," in *2018 5th International Conference on Control, Decision and Information Technologies (CoDIT)*. IEEE, 2018, pp. 38–44.