IEEE Transactions on Power Electronics

# A Decentralized Cyber-Attack Detection in DC Microgrids Based on Local Recurrent Neural Networks

| | |
|---|---|
| Journal: | *IEEE Transactions on Power Electronics* |
| Manuscript ID | TPEL-Letter-2019-08-0296 |
| Manuscript Type: | Letter |
| Date Submitted by the Author: | 12-Aug-2019 |
| Complete List of Authors: | Habibi, Mohammad Reza; Aalborg University, Energy Technology<br>Sahoo, Subham; Aalborg University, Department of Energy Technology;<br><br>Dragicevic, Tomislav (GE); Aalborg University, Department of Energy Technology |
| Keywords: | |
| | |

# A Decentralized Cyber-Attack Detection in DC Microgrids Based on Local Recurrent Neural Networks

Mohammad Reza Habibi, *Student Member, IEEE*, Subham Sahoo, *Member, IEEE* and Tomislav Dragičević, *Senior Member, IEEE,*

***Abstract*—Although distributed control can provide certain level of resilience against cyber attacks, coordinated attacks can still be designed to bypass observability owing to the lack of global information in individual controllers/observers. To address this issue, this work introduces a nonlinear auto regressive exogenous (NARX) neural network-based decentralized detection scheme of coordinated attacks on current measurements in cooperative DC microgrids. As a decentralized attack detection approach, it entails higher flexibility and less complexity as compared to the detection strategies carried out in past works. This is verified under simulation and experimental conditions to conclude that the proposed method exhibits robust behavior under cyber-physical disturbances such as communication link failures and plug-and-play of additional units.**

***Index Terms*—DC microgrid, nonlinear auto-regressive exogenous neural network, false data injection attack.**

## I. Introduction

DISTRIBUTED control of DC microgrids offers a reliable, flexible and economic alternative to centralized approach [1]. It provides resiliency from single-point-of-failure, plug-and-play capability, and a reduced cost of communication infrastructure [2]. Using this control strategy, the measurements from neighbors are exchanged between each other to achieve consensus for average voltage regulation and proportional load sharing in DC microgrids [3]. However, these measurements can be manipulated by injecting malicious data into the sensors and communication links [4]. These intrusion attempts, usually termed as false data injection attacks (FDIAs), may lead to disorientation of the system control objectives, possibly even leading to shutdowns. In recent studies, it has been shown that attacks with high degree of coordination, also termed as *stealth* attacks, can easily bypass the detection metrics of distributed observers [5]. Such attacks have been studied in [5], but only for coordinated attacks using linear observers on voltage measurements. However, significant attention should also be given towards detecting coordinated attacks on current measurements, which may create power imbalance leading to increased voltages and potentially causing shutdown of converters.

An achievable degree of sophistication of cyber attacks in cooperative DC microgrids is primarily determined by sufficient knowledge about the communication graph and control structure. Since the control and communication layer is dynamically coupled in cooperative networks, it becomes fairly easy for the attacker to formulate coordinated attacks with minimal information about the system. Hence, any attack detection strategy that relies on communicated/two-hop measurements can be deemed unreliable, since the communicated detection metric could also be compromised [6]. To provide best results with enhanced reliability, design of decentralized attack detection strategies should be encouraged. However, this research direction has largely been ignored in the realm of cyber-attack detection strategies for DC microgrids.

This paper bridges down this research gap by exploiting the nonlinear mapping capability of NARX neural network to detect coordinated FDIAs on current measurements locally. Using this philosophy, a simple yet robust detection norm is provided to distinguish between common cyber-physical disturbances (such as communication link failure, outage of converter and load change) and cyber attacks in DC microgrids. Finally, the authenticity and accuracy of the proposed detection method is simulated in a multi-agent based DC microgrid and validated experimentally.

M.R. Habibi, S Sahoo and T. Dragičević are with the Department of Energy Technology Power Electronic Systems, Aalborg University, Aalborg 9220, Denmark (e-mail: mre@et.aau.dk, sssa@et.aau.dk and tdr@et.aau.dk).

## II. Coordinated FDIAs in Cooperative DC Microgrids

### A. Conventional Control Scheme of Cooperative DC Microgrids

As highlighted in Fig. 1, each DC/DC converter in a DC microgrid operates to follow these objectives:

$$\lim_{t\to\infty} \bar{V}^i_{dc}(t) = V_{dc_{ref}}, \ \lim_{t\to\infty} \bar{I}^i_{out}(t) = 0 \tag{1}$$

where $V_{dc_{ref}}$, $\bar{I}^i_{out}$ and $\bar{V}^i_{dc}$ denote the global voltage reference, proportionate current sharing input and the average voltage estimate of $i^{th}$ agent, respectively. The average voltage estimates of each unit is updated using a *dynamic consensus* algorithm using the actual voltage to provide:

$$\dot{\bar{\mathbf{V}}}_{dc} = \dot{\mathbf{V}}_{dc} - \mathbf{L}\bar{\mathbf{V}}_{dc} \tag{2}$$

where $\bar{\mathbf{V}}_{dc}$ and $\mathbf{V}_{dc}$ denote the vector representation of average and actual voltage, respectively. Further, $\mathbf{L}$ denotes the Laplacian graph which reveals the topology of cyber network. Similarly, the output current regulation update, $\bar{\mathbf{I}}_{out}$ is given by:

$$\bar{\mathbf{I}}_{out} = -\mathbf{L}\mathbf{I}_{out} \tag{3}$$

where $\mathbf{I}_{out}$ denotes vector representation of output currents. To achieve these objectives, the cooperative secondary controller, as shown in Fig. 1, produces two voltage correction terms, i.e. $V^i_{vsec}$ and $V^i_{csec}$ for $i^{th}$ agent using:

$$V^i_{vsec}(t) = H^i_V(s)\underbrace{[V_{dc_{ref}} - \bar{V}^i_{dc}(t)]}_{e^i_V} \tag{4}$$

$$V^i_{csec}(t) = -H^i_I(s)\bar{I}^i_{out}(t), \tag{5}$$

where $H^i_V(s)$ and $H^i_I(s)$ are the PI controllers in secondary layer used for average voltage regulation and proportionate current sharing, respectively. Finally, the voltage correction terms obtained for $i^{th}$ agent in (4) and (5) are added to the global voltage reference to obtain the final reference $V^i_{dc_{ref}}$ using:

$$V^i_{dc_{ref}} = V_{dc_{ref}} + V^i_{vsec} + V^i_{csec}. \tag{6}$$

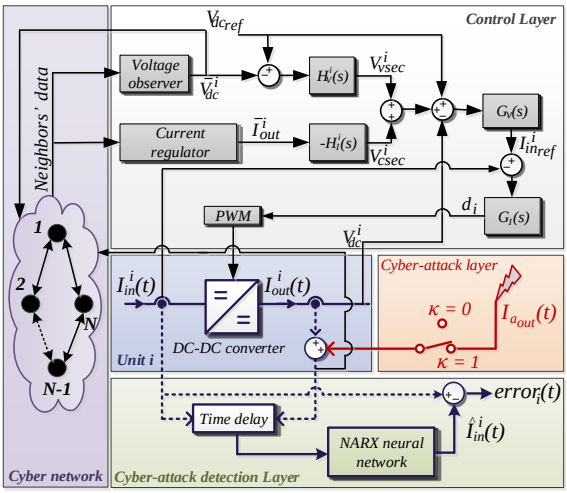



Fig. 1. Implementation of NARX neural network for identification of FDI attacks.

### B. Coordinated FDIAs on Current Sensors

A sufficient condition for any cyber attack in DC microgrid to be termed as *coordinated*, is when (1) holds true even in the presence of attack. Following detailed study on modeling and distributed detection of *stealth* attacks using voltage measurements in [5], this work shifts gears by focusing on modeling and decentralized detection of coordinated attacks on current measurements. Hence, the compromised output current regulation update $\bar{\mathbf{I}}_{out}$ in the presence of coordinated cyber attack is given by:

$$\bar{\mathbf{I}}_{out} = \mathbf{L}[\mathbf{I}_{out}(t) + \psi \mathbf{I}_{a_{out}}] = 0 \tag{7}$$

where $I^i_{a_{out}}$ denotes the vector representation of the attack element in $i^{th}$ agent. Further, $\psi = [\kappa_1, \kappa_2, ..., \kappa_N]$, where $\kappa_i$ = 1 denotes the presence of the attack element in $i^{th}$ agent.

In the presence of attack, the dynamics of voltage control loop as shown in Fig. 1 can be written as:

$$I^i_{in_{ref}} = G^i_v(s)[V^i_{dc_{ref}} - V^i_{dc}], \tag{8}$$

where $I^i_{in_{ref}}$ is the input current reference for $i^{th}$ agent and $G^i_V(s)$ is a PI controller. Using cooperative vulnerability factor (CVF) detection law [5], it can be concluded that $\mathbf{L}\mathbf{V}_{vsec}$ = 0 under no attacks. Using this equality and multiplying $\mathbf{L}^T$ with the vector representation of (8), we get:

$$\mathbf{L}^T\mathbf{I}_{in_{ref}} = \mathbf{G}_v[\mathbf{L}^T V_{dc_{ref}}\mathbf{1} - \mathbf{L}^T\mathbf{H}_I\bar{\mathbf{I}}_{out} - \mathbf{L}^T\mathbf{V}_{dc}], \tag{9}$$

where $\mathbf{1}$ is an identity matrix. Since average voltage estimate is regulated to the global reference, $\mathbf{L}^T V_{dc_{ref}}\mathbf{1} = \mathbf{L}^T\mathbf{V}_{dc}$. Simplifying (9), we get:

$$\mathbf{L}^T\mathbf{I}_{in_{ref}} = -\mathbf{G}_v(s)\mathbf{H}_I(s)\mathbf{L}^T\bar{\mathbf{I}}_{out}. \tag{10}$$

*Remark I:* Even though the attack on current measurements is coordinated such that (1) holds true, the symmetric attack element instills an asymmetric component in the left-hand side of (10) owing to the double order integration. Assuming apt current tracking performance by the current controller, the input currents will not follow consensus theory in the presence of attacks.

Hence, Remark I provides the key motivation to design a non-linear mapping tool to acquire the relationship between input and output currents during normal operating conditions, and use such relationship to detect anomalies, i.e. cyber-attacks.

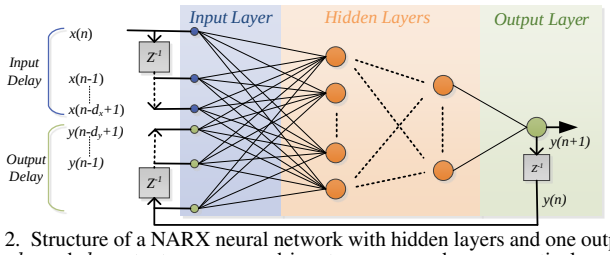



Fig. 2. Structure of a NARX neural network with hidden layers and one output with $dy$ and $dx$ output-memory and input-memory orders, respectively.

## III. Proposed Decentralized NARX Neural Network Based Attack Detection Mechanism

NARX is a subclass of the recurrent neural networks. NARX neural network has some advantages over conventional recurrent neural networks such as better accuracy, higher convergence speed and so on. NARX can be used for predicting nonlinear and time series systems [7], [8]. The basic mathematical representation of NARX in a discrete nonlinear time system is expressed as follows [9]:

$$y(n+1) = f(x(n), ..., x(n-d_x+1), y(n), ..., y(n-d_y+1)), \tag{11}$$

where $y(n)$ and $x(n)$ are $n^{th}$ output and input samples the system, respectively. Also, $f(.)$ is a nonlinear function. In addition, $d_y$ and $d_x$ are output-memory and input-memory orders, respectively, which are positive. Fig. 2 illustrates the general architecture of a NARX neural network with hidden layers and one output.

If $\hat{y}(t+1)$ is considered as the estimation of the output, (12) shows the mathematical formulation of a NARX neural network with a single hidden layer [10]:

$$\hat{y}(t+1) = f_o(W_o(f_h(W_h U(t) + b_h) + b_o), \tag{12}$$

where, $b_o$, $b_h$, $W_o$, $W_h$, $f_o$ and $f_h$ are bias vectors, weight matrices and activation functions of the output layer and hidden layer, respectively. Furthermore, $U(t)$ is as follows:

$$U(t) = [x(t), ..., x(t-d_x+1), y(t), ..., y(t-d_y+1)]^T. \tag{13}$$

For the training and operation of NARX neural network, a series-parallel mode is considered in this work, where the output's regressor is represented by the actual value of the system.

Using the necessary and sufficient conditions in Remark I, the idea behind this work is to create a historic dynamic relationship using a NARX neural network for each unit to estimate the input current of the DC/DC converter. If this prediction is not in accordance with the measured value, we can conclude that the system is under attack. The implementation of the proposed detection scheme is illustrated in Fig. 1, where only the local measurements are used for detection of cyber attacks on current measurements. With output current of the converter as input to NARX neural network, the estimated output $\hat{I}^i_{in}$ from the NARX neural network is compared with the measured input current $I^i_{in}$ of $i^{th}$ agent to give an error quantity:

$$error_i(t) = c_i(I^i_{in}(t) - \hat{I}^i_{in}(t)). \tag{14}$$

It is worth notifying that, a positive quantity $c_i = 100$ as a scaling factor is multiplied with $error_i$ in (14) to provide a clear indication of attacks on output plots.

*Remark II:* The estimation error in (14) will be non-zero in the case of FDIAs, thus providing a clear norm for detection of coordinated attacks on current measurements in cooperative DC microgrids.

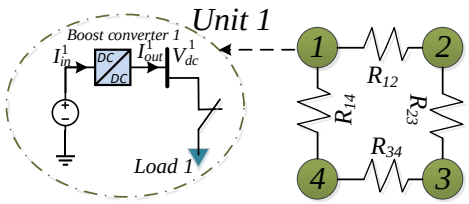



Fig. 3. Structure of the test DC microgrid system in Matlab/Simulink environment

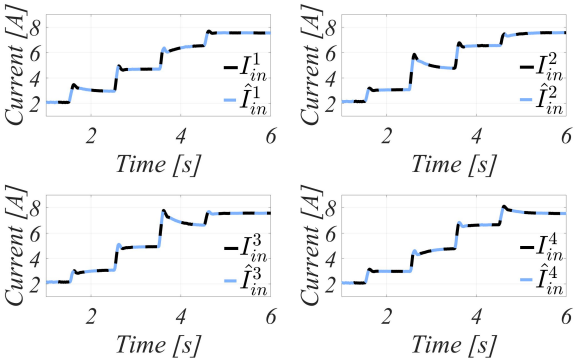



Fig. 4. Measured and estimated value of input current of converters.

## IV. Simulation Results

The performance of the proposed detection approach was evaluated in a DC microgrid with four units, as shown in Fig. 3, in MATLAB/Simulink environment. The NARX neural network for each unit was trained in a supervised manner from a detailed simulation model by acquiring samples of input and output currents of the converter to plot the map between input and output with high accuracy. In this study, in order to gather data to train the NARX neural network, the simulated DC microgrid was run and data at every time step was collected. In this work, the sampling time is 0.1 $ms$ and the selected time to collect data is considered 10 s so, 100000 samples of input and output currents of the converter is gathered to train the neural network. It is important to note that load changes were programmed in the simulated DC microgrid to reflect different conditions in the training phase. In addition, the NARX neural network consists of one hidden layer with 10 neurons and activation function $g(.)$:

$$g(x) = \frac{2}{1 + e^{-2x}} - 1. \tag{15}$$

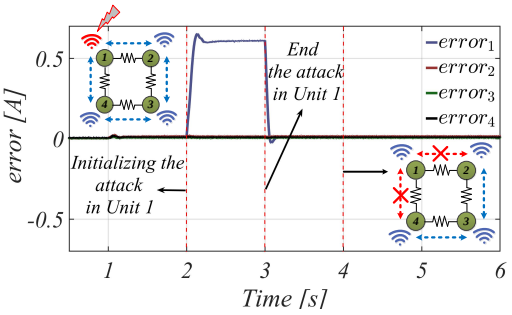
Further, in this study, the number of input- and output-memory orders are six. All the simulation parameters are provided in the Appendix.

### A. Scenario 1

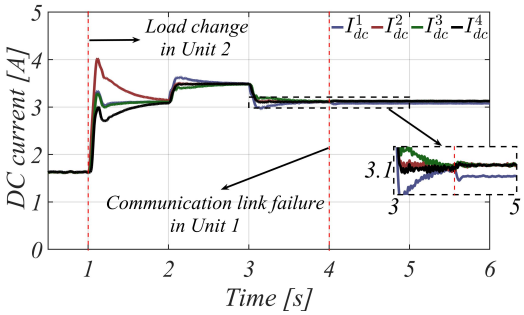
In Scenario 1, the estimation accuracy of the NARX neural network based estimator is tested. The performance of the NARX based estimator is evaluated for increase in load in the considered microgrid. In this scenario, loads are added in unit 1, unit 2, unit 3 and unit 4 at t = 1.5 s, 2.5 s, 3.5 s and 4.5 s, respectively. Fig. 4 indicates the measured input current of all units and also the estimated value of them by the NARX based estimator.

### B. Scenario 2

In Scenario 2, the performance of the decentralized detection scheme is tested under communication link failure in Fig. 5. When the attack is injected into unit 1 at t = 2 s, the output currents increase to obey the objectives in (1). This can be explained owing to the presence of resistive loads in the microgrid. For a positive-valued attack element, (5) results in increase of $V_{csec}$ setpoints for the non-attacked units. This increases the demand as well as the tie-line losses, which leads to an increase in the generation from each unit. However, a non-zero estimation error is encountered for unit 1 at t = 2 s, thereby concluding that the output current measurement unit 1 is attacked. Further, the estimation error returns back to zero when the attack is removed. It can also be seen that the proposed detection scheme is robust to cyber-physical disturbances, as the estimation error is zero when cyber link between units 1-2 and 1-4 fails. This behavior can be explained owing to the non-linear mapping by NARX neural network for different loading conditions.




(a) Difference between measured value and estimated value of input current of converters for all units in *Scenario* 2.




(b) DC output currents of DC/DC converters.

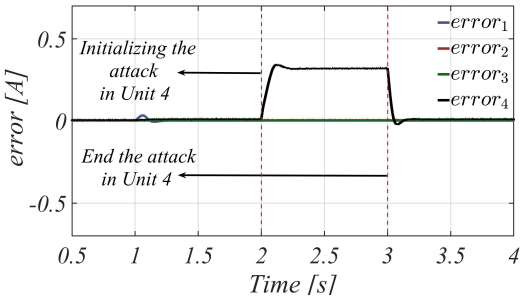
Fig. 5. Performance of proposed detection scheme under communication link failure : a) Estimation error, and b) Output currents from all units.

### C. Scenario 3

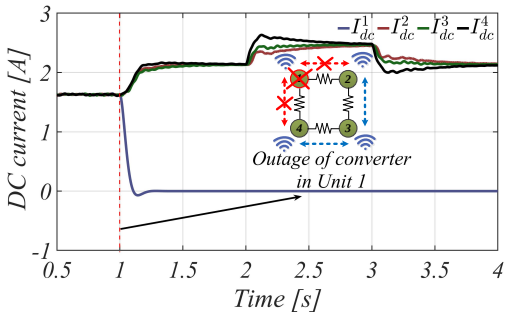
In Scenario 3, the performance of the proposed detection scheme is tested during outage of converter in unit 1. It can be seen in Fig. 6 that even though the estimation error is zero for each unit even though the converter 1 is plugged out. Since the relationship between input and output currents for the active units accords with the total demand, the estimation error is zero. However when an attack is injected into unit 4 at t = 2 s, the estimation error of unit 4 goes non-zero and returns back to zero as soon as the attack is removed at t = 3 s. During outage of converters, the cyber links are also disabled, as highlighted in Fig. 6, which affects the performance of a two-hop cyber attack detection schemes. This establishes the flexibility of operation of the proposed detection scheme whose performance remains unaffected under these conditions.

## V. Experimental Results

The proposed detection strategy has also been experimentally validated in a DC microgrid with $N$ = 2 agents, as shown in Fig. 7. Each converter is supplied by a DC source, and is operated to maintain output voltage across their respective buses. Using the local and neighboring measurements, the secondary




(a) Difference between measured value and estimated value of input current of converters for all units in *Scenario* 3.




(b) DC output currents of DC/DC converters.

Fig. 6. Performance of proposed detection scheme under outage of converter: a) Estimation error, and b) Output currents from all units.

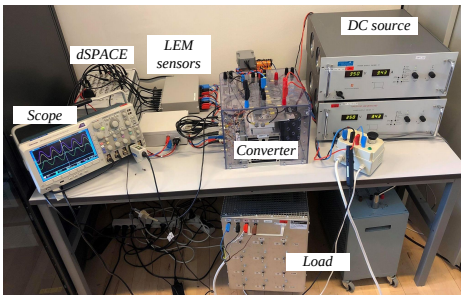



Fig. 7. Experimental prototype used to validate the proposed NARX-NN based attack detection mechanism in DC microgrids.

sublayers along with the proposed attack detection mechanism are modeled for each agent in dSPACE MicroLabBox DS1202. Further, the estimation error values are acquired using the AO (analog output) channels.

By the defined proposition, it can be seen in Fig. 8 that as a FDIA is conducted in Unit 2 during event A, $error_2$ immediately goes positive, despite the output currents are equally shared. To validate its operation during normal disturbances, a load change is conducted where the estimation errors remain unaltered. As a result, the proposed decentralized NARX neural network based attack detection framework is validated under practical experimental conditions.

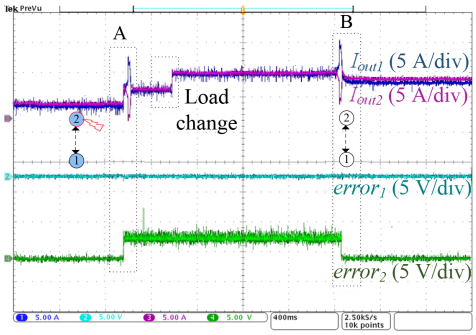



Fig. 8. Experimental validation of the proposed decentralized attack detection theory.

## VI. Discussions

This paper benefits from the nonlinear mapping capability of NARX neural network to use them as a local estimators to detect FDIAs in DC microgrids. The NARX based estimator can distinguish between FDIAs and disturbances such as load changing, outage of any units and communication link failure. The simulation and experimental results show the effectiveness of the proposed method. As the proposed method is a decentralized detection approach, there is no need to have data of other units to detect the FDIAs. It is important to note that this work focuses on FDIAs as a kind of cyber-attack on current sensors. The concept of this work will be implemented in the future also to detect other kind of cyber-attacks in microgrids and detection of other types of attacked instruments such as voltage sensors.

## VII. Conclusions

A novel decentralized NARX neural network based detection strategy is proposed to detect coordinated false data injection attacks on output currents, which usually goes undetectable since the output currents are shared proportionately. Since it operates on local measurements, the proposed strategy is more robust and flexible as compared to the conventional strategies, which rely on communicated measurements. The proposed mechanism has been validated experimentally to show the robustness and ease of implementation for any commercially available voltage controlled DC/DC converters.

## Appendix

The control and plant parameters used in simulation are provided below: $R_{12} = 1.8\ \Omega$, $V_{dc_{ref}} = 315\ V$, $R_{23} = 2.3\ \Omega$, $\Delta t = 0.1\ ms$, $R_{34} = 2.1\ \Omega$, $L_f = 5\ mH$, $R_{14} = 1.3\ \Omega$, $C_f = 50\ mF$

## References

[1] B. Abdolmaleki, Q. Shafiee, M. M. Arefi, and T. Dragicevic, "An instantaneous event-triggered hz-watt control for microgrids," *IEEE Trans. Power Syst.*, pp. 1–1, 2019.

[2] T. Dragičević, X. Lu, J. C. Vasquez, and J. M. Guerrero, "Dc microgrids—part i: A review of control strategies and stabilization techniques," *IEEE Trans. Power Elect.*, vol. 31, pp. 4876–4891, July 2016.

[3] S. Sahoo and S. Mishra, "A distributed finite-time secondary average voltage regulation and current sharing controller for dc microgrids," *IEEE Transactions on Smart Grid*, vol. 10, pp. 282–292, Jan 2019.

[4] D. Jin, Z. Li, C. Hannon, C. Chen, J. Wang, M. Shahidehpour, and C. W. Lee, "Toward a cyber resilient and secure microgrid using software-defined networking," *IEEE Transactions on Smart Grid*, vol. 8, no. 5, pp. 2494–2504, 2017.

[5] S. Sahoo, S. Mishra, J. C. Peng, and T. Dragicevic, "A stealth cyber attack detection strategy for dc microgrids," *IEEE Trans. Power Elect.*, vol. 34, pp. 8162–8174, Aug 2019.

[6] J. Duan, W. Zeng, and M.-Y. Chow, "Resilient distributed dc optimal power flow against data integrity attack," *IEEE Transactions on Smart Grid*, vol. 9, no. 4, pp. 3543–3552, 2016.

[7] L. Jin, S. Li, and B. Hu, "Rnn models for dynamic matrix inversion: A control-theoretical perspective," *IEEE Trans. Ind. Inform.*, vol. 14, pp. 189–199, Jan 2018.

[8] T.-N. Lin, C. L. Giles, B. G. Horne, and S.-Y. Kung, "A delay damage model selection algorithm for narx neural networks," *IEEE Trans. Sig. Proc.*, vol. 45, pp. 2719–2730, Nov 1997.

[9] J. M. P. Menezes and G. A. Barreto, "Long-term time series prediction with the narx network: An empirical evaluation," *Neurocomputing*, vol. 71, no. 16, pp. 3335 – 3343, 2008. Advances in Neural Information Processing (ICONIP 2006) / Brazilian Symposium on Neural Networks (SBRN 2006).

[10] S. A. Emami and A. Roudbari, "Multi-model elm based identification of an aircraft dynamics in the entire flight envelope," *IEEE Trans. Aerosp. Electron. Syst.*, pp. 1–1, 2018.