# Data-driven Cyber-attack Detection of Intelligent Attacks in Islanded DC Microgrids

**Wan, Yihao; Dragicevic, Tomislav**

*Published in:*
IEEE Transactions on Industrial Electronics

*Link to article, DOI:*
10.1109/TIE.2022.3176301

*Publication date:*
2022

*Document Version*
Peer reviewed version

*Citation (APA):*
Wan, Y., & Dragicevic, T. (2022). Data-driven Cyber-attack Detection of Intelligent Attacks in Islanded DC Microgrids. *IEEE Transactions on Industrial Electronics*, *70*(4), 4293-4299. Article 9782082. https://doi.org/10.1109/TIE.2022.3176301

# Data-driven Cyber-attack Detection of Intelligent Attacks in Islanded DC Microgrids

Yihao Wan, *Student Member, IEEE*, Tomislav Dragičević, *Senior Member, IEEE*

***Abstract*—In this paper, a data-driven cyber-attack detection method for islanded DC microgrids is proposed. Data is collected by monitoring the behavior of an intelligent attacker who is able to bypass conventional cyber-attack detection algorithms and disrupt the operation of the system. Reinforcement learning (RL) algorithm emulates the actions of such intelligent attacker, who exploits the vulnerability of index-based cyber-attack detection methods, such as discordant detection algorithm. The data is then used to train a neural network based detector that complements the conventional method with additional capability to detect a larger set of possible attacks. Through experiments, the effectiveness of the proposed method is validated.**

***Index Terms*—DC microgrids, data-driven cyber-attack detection, reinforcement learning, discordant detection algorithm, neural network based detector.**

## I. Introduction

DC microgrids facilitate smart grid applications in an efficient and cost-effective way due to the natural matching with different distributed generation resources [1]. For the control of microgrid, distributed control has become popular as it offers better scalability, reliability, and efficiency compared with centralized control, which also suffers from the single point of failure [2]. As the distributed control rests on the communication network, it makes the DC microgrids cyber-physical systems, which are vulnerable to cyber-attacks. Among different types of cyber-attacks [3], [4], the most common are false data injection attacks (FDIAs). FDIAs alter the system states by injecting data into the sensors or communication links and disrupt the operation of the system [5]. Such attacks can destabilize the DC microgrids if not detected and mitigated properly.

To mitigate the vulnerability of DC microgrids, different cyber-attack detection methods are proposed, which could be broadly classified into model-based and model-free methods [6]. The model-based detection methods rely on the accuracy of the system model, which is challenging to implement in practical applications due to its unavoidable mismatch with the complex real-world power electronic systems. On the other hand, the model-free methods utilize the measurements without prior knowledge of the system. In [7], based on the estimated outputs of the system using an artificial neural network, the stealthy FDIA detection method is proposed. However, since the data used to train the given neural network based detector is generated from the healthy model of the microgrid, it does not explicitly incorporate knowledge about different types of attacks. As a result, the method is not able to detect some types of attacks, such as destabilization attacks. Signal temporal logic (STL) is proposed to detect FDIAs by monitoring the output voltage and current with defined specifications [8] while the performance under stealthy attack is not verified. A discordant detection algorithm is proposed to detect both destabilization and stealthy attacks by calculating the discordant element (DE) term [9]. In addition, similar methods are proposed in [10]–[12] utilizing different indices for attack detection. However, these index-based detection methods will still fail when intelligent attackers introduce novel attack patterns and a wider range of coordinated or uncoordinated attacks by injecting false signals into the sensors, communication links of multiple nodes, or concurrently both of them [11]. Such intelligent and deceptive behaviors can be emulated via reinforcement learning.

Of different forms of machine learning, reinforcement learning is the learning paradigm closest to the human learning process as it can learn through experience by exploring and exploiting the dynamic and unknown environment [13]. RL can model an intelligent agent to take sequential optimal actions without or with limited knowledge of the environment, which makes it particularly adaptable and feasible in real-time systems. Therefore, reinforcement learning demonstrates excellent suitability for application in cyber-security areas, where cyber-attacks become increasingly sophisticated [14]–[16]. Based on this, the reinforcement learning agent is able to serve as an intelligent attacker, who exploits the vulnerability of DC microgrids system protected with conventional discordant detection scheme by generating novel attack patterns. To generate sophisticated cyber-attacks in DC microgrids, in reinforcement learning algorithm, deep neural networks that represent the attacker are trained over many system rollouts and autonomously discover the deficiency of the index-based cyber-attack detection method in DC microgrids. The deep neural network afterward interacts with the real system, injects false signals into multiple nodes coordinately, nullifies the indices that are used for detection, such as discordant elements in [9], and crafts stealthy attacks that can bypass conventional cyber-attack detection methods.

To solve the aforementioned issue, in this letter, a data-driven cyber-attack detection method for DC microgrids is proposed. Particularly, the RL-based intelligent cyber-attacker can uncover the deficiencies of the DE-based detection algorithm but could also expose other index-based detection methods if trained in such a way. In view of this, the proposed data-driven cyber-attack detection method is to complement

Yihao Wan and Tomislav Dragičević are with the Department of Electrical Engineering, Technical University of Denmark, Copenhagen, Denmark (e-mails: wanyh@elektro.dtu.dk, tomdr@elektro.dtu.dk).

the conventional index-based detection methods by detecting the attacks and the attacked nodes under the RL-based cyber-attacks.

Unlike model-based cyber-attack detection methods, in the proposed data-driven cyber-attack detection method, only the current and voltage measurements are used. Moreover, the historical operation data is directly collected from a DC microgrid experimental testbed, which allows higher training precision compared to the previous data-based method that used simulation models for this purpose. Finally, to the best of the authors' knowledge, this paper is the first attempt to use the data from the system exposed to intelligent cyber-attacks to train the attack detector, while existing methods use data from normally operated microgrids. In real-time applications, the RL-based attacker will be implemented to generate sophisticated attacks to bypass the DE-based detection method, such a data-driven attack detector is then utilized to complement the DE-based detection method for detecting the attacks and identifying the attacked nodes in DC microgrids. The neural network based cyber-attack detectors are implemented in an experimental setup to verify the performance of the proposed method.

## II. RL-Based FDIA on Cooperative Control Based DC Microgrids

### A. Cooperative Control of DC Microgrids with Discordant Detection Algorithm

The configuration of the DC microgrid is shown in Fig. 1, where $n$ DC sources connected through dc-dc converters are linked via communication networks and form the cyber-physical network. The main objectives of the control are voltage regulation and proportional current sharing. The output voltage of each node is regulated by the primary control layer. The current sharing causes voltage error, which is compensated by the distributed secondary control [17].

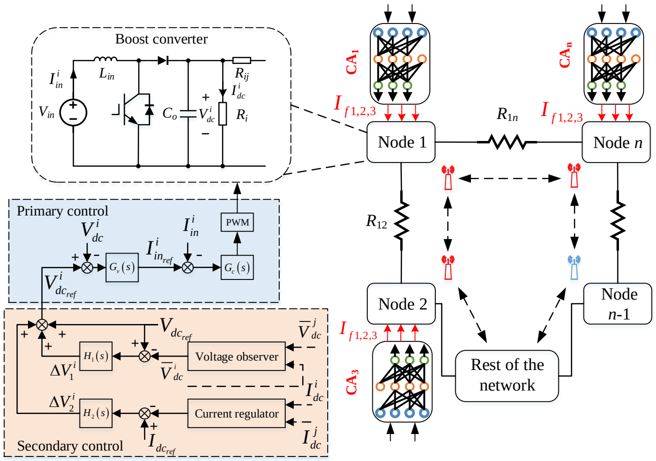


Fig. 1. Configuration of DC microgrid under multi-agent RL-based cyber-attack.

The distributed control rests on the communications between neighboring nodes. To achieve the control objectives, in the secondary control layer, two voltage correction terms $\triangle V_1^i$ and $\triangle V_2^i$ are calculated for the reference voltage of the primary control layer to regulate the output voltage of each node [17]. The reference voltage can thus be expressed as

$$V^i_{dc_{ref}} = V_{dc_{ref}} + \triangle V_1^i + \triangle V_2^i \tag{1}$$

where $V_{dc_{ref}}$ is the global reference voltage for all the nodes and $I_{dc_{ref}}$ is set to 0 to achieve current sharing.

Based on the distributed control in a fully-connected cyber network in DC microgrids, the control objectives will converge as follows [17]

$$\lim_{t\to+\infty} \bar{V}^i_{dc}(t) = V_{dc_{ref}}, \lim_{t\to+\infty} \bar{I}^i_{dc} = 0 \tag{2}$$

where $\bar{V}^i_{dc}$ and $\bar{I}^i_{dc}$ are the estimated average voltage and the normalized current regulation input for node $i$ respectively.

False signals could be injected into the sensor measurements, the distributed signals from the neighboring nodes, or both of them. The attacked signals can be expressed as

$$\boldsymbol{I_A} = \begin{cases} I^i_a = I^i_{dc} + k_i I_{fi}, & \text{sensor attack} \\ I^{ij}_a = I^i_{dc} + k_{ij} I_{fj}, & \text{link attack} \end{cases} \tag{3}$$

where $\boldsymbol{I_A}$ denotes the attacked vector at multiple nodes, $I^i_a$ is the output current from the node $i$, and $I^i_{dc}$ is the real measured value from the sensors, $I^{ij}_a$ is the distributed current received from neighboring node $j$ for node $i$, $I_{fi}$ and $I_{fj}$ are the injected false signals from the attacker, $k_i$ and $k_{ij}$ are the corresponding coefficients for the attack signals, of which $k_i = 1$ or $k_{ij} = 1$ denotes the presence of an attack in the corresponding signals, or 0 otherwise.

For different uncoordinated attacks, the control objectives will not converge as in equation (2). These events could destabilize the DC microgrids depending on the attack intensity. On the other hand, the attack that results in the control inputs converging as per equation (2) is considered as coordinated attack, which is generally formed by attacking sensors and communication links concurrently. Among different detection schemes, discordant detection algorithm, based on the synchrony between the neighboring reference current terms of a node, has shown excellent performance in detecting both types of cyber-attacks [9]. In the distributed control, the control objectives will converge as per (2), and the input current $I^i_{in}$ will always converge to zero under no attacks, thus the input current reference $I^i_{in_{ref}}$ quantities will also achieve consensus among themselves. However, when the attack occurs, for the attacked node, the secondary layer will maloperate due to the compromised current information, which causes different operation of the outer voltage loop, thus the current reference for the compromised node goes discordant with the remaining healthy nodes. The index for attack detection is calculated as

$$DE_i = l_i[\sum_{j\epsilon M_i}(I^j_{in_{ref}} - I^i_{in_{ref}})][\sum_{j\epsilon M_i}(I^j_{in_{ref}} + I^i_{in_{ref}})] \tag{4}$$

$$DE_i = \begin{cases} < DE_{min}, & if\ k_i\ \&\ k_{ij} = 0 \\ > DE_{min}, & if\ k_i \| k_{ij} \neq 0 \end{cases} \tag{5}$$

where $DE_i$ denotes the discordant term of node $i$, $I^i_{in_{ref}}$ and $I^j_{in_{ref}}$ are the input reference current from the outer voltage control loop for node $i$ and the neighboring node $j$ respectively,

$M_i$ denotes the set of neighbors of node $i$, $l_i$ is a positive coefficient to increase/decrease the value of $DE_i$. According to [9], based on (5), the attacks on the $i$th node can be determined by comparing the positive value of $DE_i$ terms with the minimum threshold, which is obtained in normal operation.

### B. Multi-agent Reinforcement Learning Based FDIA

To compromise the conventional detection method, multi-agent reinforcement learning is utilized. The autonomous attack generation process can be modeled as a Markov decision process (MDP), which consists of state space $S$, action space $A$, state transition probability function $P$ and reward function $R$ [13]. At each time step $t$, the agent $i$ observes the state $s_i^t$ from the environment, takes action $a_i^t$ based on the policy $\pi(a_i^t|s_i^t)$ and receives a reward $r_i^t$. The policy $\pi(a_i^t|s_i^t)$ maps the state $s_i^t$ to a probability distribution of action $a_i^t$. For the next time step, a new state $s_i^{t+1}$ is formed. The cumulative discounted reward could be expressed as

$$G_i^t = \sum_{k=0}^{\infty} \gamma_i^k r_i^{t+k} \tag{6}$$

where $\gamma \ \epsilon \ [0, 1]$ is the discounting factor.

In order to generate FDIA in DC microgrids to bypass the DE-based detection scheme, the target of the RL agents is to suppress the DE terms by injecting false signals into the sensors or communication links of multiple nodes. As is expressed in equations (4) and (5), any evident increase of the value will reflect an attack in the current counterparts of $i$th node. In addition, in the distributed cooperative control of the system, shown in Fig. 1, any deviation on current distributed terms $I_{dc}^j$ or local terms $I_{dc}^i$ would create an offset on term $\triangle V_2^i$ which deviates the voltage set-point $V_{dc_{ref}}^i$ from secondary control layer to the local control layer. This will lead to the deviation of the corresponding input reference current $I_{in_{ref}}^i$ and $I_{in_{ref}}^j$, which will change the value of DE terms. Thus, the RL-based intelligent attackers can harmonize and synchronize this offset between the neighboring nodes by attacking multiple nodes. In this way, the sophisticated attack that the conventional DE-based detection algorithm cannot detect is generated.

Particularly, in a DC microgrid shown in Fig. 1, for a node with $m$ incoming links, the list of cyber-attack agents is defined as $\{CA_1, ..., CA_{m+1}\}$. The observations of each agent are $\boldsymbol{S_i^t} = \{\{DE_1^t, ..., DE_{m+1}^t\}, \{\int DE_1^t, ..., \int DE_{m+1}^t\}\}$ of neighboring nodes. The corresponding actions of each agent are $\boldsymbol{A_i^t} = \{I_{f1}^t, ..., I_{fm}^t\}$, where $I_{f1}^t$ is the attack signals on local sensor and the rest are on incoming communication links from the neighboring nodes. The reward function is defined as

$$\begin{aligned} r_i^t = & - (k_{DE} \sum_{i=1}^{m+1} (DE_i^t)^2 + k_{\dot{DE}} \sum_{i=1}^{m+1} (\dot{DE}_i^t)^2 \\ & + k_{\dot{I}_f} \sum_{i=1}^{m} (\dot{I}_{fi}^{t-1})^2) + r_{dis}^t \end{aligned} \tag{7}$$

where $k_{DE}$ and $k_{\dot{DE}}$ are the coefficients for the summation of DE terms and the corresponding derivatives, respectively, which are adjusted to minimize the discordant terms and their variations. $k_{\dot{I}_f}$ is the coefficient for the summation of derivative of attack actions taken in the last time step $t-1$, which is tuned to minimize the variations of the generated attack signals especially when desirable stealthy attack performance is obtained. To help with the convergence of training the agents, a negative discrete reward term $r_{dis}^t$ is introduced, which is expressed as below

$$r_{dis}^t = -(k_1 \cdot r_1^t + k_2 \cdot r_2^t) \tag{8}$$

$$r_1^t = |\sum_{j=1}^{m} (I^{t-1}{}_{fj} - I^{t-1}{}_{fi})| < I_{f_{min}} \tag{9}$$

$$r_2^t = \Big((DE_1^t|...|DE_i^t) > DE_{max}\Big)\Big|_{i=1}^{m+1} \tag{10}$$

where the term $r_1^t$ indicates the intrusion terms on sensor and cyber links are canceling each other according to the calculation of the consensus current [17], $k_1$ is the coefficient to ensure the presence of the minimum non-canceling destabilizing cyber-attack as denoted by $I_{f_{min}}$ with the overall output actions, i.e. the generated attack signals, in time $t-1$, and $k_2$ is the coefficient for penalizing the detection of excessive value of DE terms during the training stage. The value of $I_{f_{min}}$ is chosen with a trade-off between the slope of the ramp for the destabilizing phenomenon under the generated attacks and the minimum discordant terms threshold, $DE_{max}$ is the upper threshold of discordant terms. Therefore, during the training process, the RL agents will learn autonomously to produce destabilization FDIAs to minimize the DE value. Thus, the attacks remain undetected by the discordant detection method.

The goal of the agents is to learn the policy $\pi(a_i^t|s_i^t)$ to maximize the reward $r_i^t$ and thus to maximize the discounted reward $G_i^t$. For a specific policy $\pi$, the action-value function $Q^\pi(s_i^t, a_i^t)$ is used in reinforcement learning algorithm to describe the expected return with the action $a_i^t$ with respect to the state $s_i^t$, which is estimated based on the Bellman equation as [18]

$$\begin{aligned} Q^\pi(s_i^t, a_i^t) \leftarrow & \ Q^\pi(s_i^t, a_i^t) + \\ & \alpha_i [r_i^{t+1} + \gamma_i max Q^\pi(s_i^{t+1}, a_i^{t+1}) - Q^\pi(s_i^t, a_i^t)] \end{aligned} \tag{11}$$

where $\alpha_i$ is the learning rate of agent $i$.

Deep Q-Network (DQN) is employed due to its computationally efficient characteristics [19]. To stabilize the training process, each experience tuple $e = (s_i^t, a_i^t, r_i^t, s_i^{t+1})$ of agent $i$ at each time step is first stored in an $R$-sized experience memory $\mathcal{D} = \{e_i^1, ..., e_i^R\}$. In each time step of training process, a minibatch of the tuples are randomly selected from $R$. Afterwards, as the DQN agent has a $Q$-network, which approximates the action-value function $Q_i^\pi(s_i^t, a_i^t|\theta^Q)$ with weights $\theta^Q$. In addition, to enhance the convergence of $Q$-network, a target network $\hat{Q}_i^\pi(s_i^t, a_i^t|\theta_i^{Q'})$ is used. The weights of the $Q$-network are optimized in the training process based on the loss function as below

$$L(\theta_i^Q) = \mathbb{E}[(Q_i^\pi(s_i^t, a_i^t|\theta_i^Q) - y^t)^2] \tag{12}$$

where

$$y^t = r_i^t + \gamma \hat{Q}_i^\pi(s_i', a_i'|\theta_i^{Q'}) \tag{13}$$

The parameters of the target network are updated as

$$\theta_i^{Q'} \leftarrow \tau\theta_i^{Q} + (1-\tau)\theta_i^{Q'} \tag{14}$$

where the smoothing factor $\tau \ll 1$.

The whole training process is shown as follows.

**Algorithm 1** Multi-agent RL-based FDIA

```
Input: $DE$ and $\int DE$ of neighboring nodes
Output: Attack signals $\{I_{f1}^t, ..., I_{fm}^t\}$
Initialize replay buffer $\mathcal{D}$ to capacity $R$
Initialize action-value function $Q$ with random weights $\theta_i^Q$
Initialize target action-value function $\hat{Q}$ with weights $\theta_i^{Q'} = \theta_i^Q$
for episode = 1 to $M$ do
    Receive initial observation at state $s_i^1$
    for iteration = 1 to $T$ do
        For each agent $i$, select and execute action $a_i^t$ with respect to policy $\pi(a_i^t|s_i^t)$, receive the reward $r_i^t$ calculated with (7) and transition into state $s_i^{t+1}$
        Store tuple $(s_i^t, a_i^t, r_i^t, s_i^{t+1})$ in the $\mathcal{D}$
        for each agent $i$ = 1 to $m$ do
            Randomly select the mini-batch $e$ from $\mathcal{D}$
            Set $y^t = \begin{cases} r_i^t, \text{if episode terminates at step } t+1 \\ r_i^t + \gamma_i max\hat{Q}_i^{\pi}(s_i^{'}, a_i^{'}|\theta_i^{Q'}), \text{otherwise} \end{cases}$
            Perform gradient descent on (12) with (13) regarding the network parameter $\theta_i^Q$
        end for
        Update the target network using (14)
    end for
end for
```

## III. PROPOSED DATA-DRIVEN CYBER-ATTACK DETECTION METHOD

We propose a data-driven detection method to complement the conventional DE-based detection method for the DC microgrids under multi-agent RL-based cyber-attacks. The proposed data-driven cyber-attack detection framework is illustrated in Fig. 2. The basis of the proposed method is to detect the attacks and identify the attacked nodes by extracting the mapping relationship between the inputs and the target labels, which consists of offline model training and online cyber-attack detection.

As the RL-based cyber-attacks can bypass the discordant detection algorithm, we aim to achieve attack detection as well as attacked nodes identification, which could be considered as a classification model. Since we use the measurement data of the system under both normal and attack conditions, considering the computational burden of the artificial neural network in the real-time application, we use the pattern recognition network (PRN), a type of feedforward neural network (FNN) for classifying the inputs to target classes. It is noteworthy that the paper is not focused on comparing the performance of different machine learning models or artificial neural networks for cyber-attack detection but to apply them to detect the cyber-attacks as well as identify the attacked nodes when the intelligent attacks occur in DC microgrids. The utilized PRN is constructed with input, output, and single hidden layer as shown in Fig. 3, and the mathematical description is as follows.

$$Y = F(X_{in}) = f_{out}[f_{hid}(b_{hid} + W_{hid}X_{in})W_{out} + b_{out}] \tag{15}$$

where $f$, $W$, and $b$ denote the activation function, weight matrix, and bias matrix respectively, $X_{in} = \{x_1, ..., x_n\}$ represents the input vector, the subscripts $hid$ and $out$ denote the hidden layer and output layer.

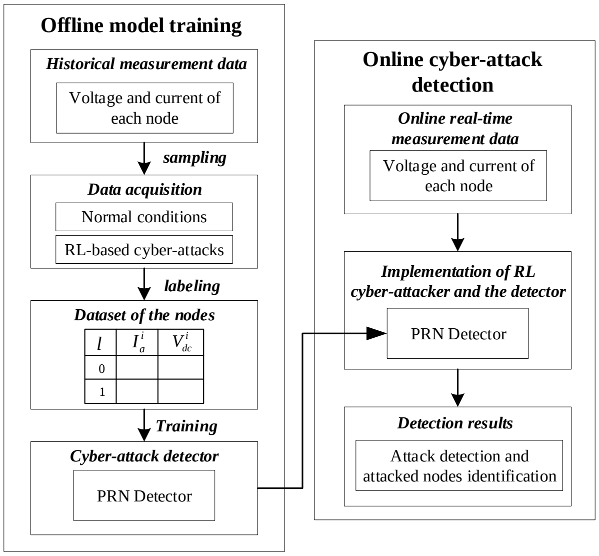

Offline model training
Historical measurement data
Voltage and current of each node
sampling
Data acquisition
Normal conditions
RL-based cyber-attacks
labeling
Dataset of the nodes
$l$
$I_a^i$
$V_{dc}^i$
0
1
Training
Cyber-attack detector
PRN Detector
Online cyber-attack detection
Online real-time measurement data
Voltage and current of each node
Implementation of RL cyber-attacker and the detector
PRN Detector
Detection results
Attack detection and attacked nodes identification

Fig. 2. Proposed data-driven cyber-attack detection method.

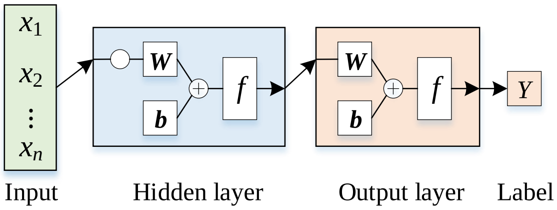


Fig. 3. Structure of PRN.

In the offline training process, to collect the data for training the neural network based cyber-attack detector, the historical operational data, consisting of measurements for $I_a$, $V_{dc}$ in the system under both normal and RL-based cyber-attack conditions for each node, is collected. By labeling the attacked nodes as 1 and the unaffected nodes as 0, the dataset for all the nodes is generated. Then the data is randomly split into training dataset, validation dataset, and test dataset. During the offline training process, the weights and biases of each layer are optimized. Subsequently, a well-trained cyber-attack detector with prior knowledge of multi-agent RL-based cyber-attack is attained.

In the online cyber-attack detection process, during the real-time system operation, the measured voltage $V_{dc}$ and current $I_a$ are input to the trained PRN based cyber-attack detector. The PRN based cyber-attack detector is implemented as a classifier for each node to extract the labels of corresponding input measurements, which will indicate whether the system is under attack and which nodes are attacked.

## IV. EXPERIMENTAL RESULTS

To validate the performance of the proposed data-driven cyber-attack detection method, an experimental setup of the DC microgrid with $n = 4$ nodes in Fig. 1 is implemented, shown in Fig. 4. The parameters of the system and controller are listed in Table I. The trained RL agents target the current sensor signals on three neighboring nodes 1, 2 and 4, and produce stealthy destabilization FDIAs on DE terms. The experimental testbed consists of dSPACE MicroLabBox DS1202 and a computer as the real-time control interface. For both the normal condition and RL-based attack condition, operation data was collected from the computer interface with a sampling time of 0.2 ms for 10 s respectively, which means the dataset comprising a total of 100000 samples of $I_a$, $V_{dc}$, and the corresponding labels for each node. The PRN is trained based on the dataset, where 80% randomly divided data was used to train the neural network, and 10% was used for validation and testing, respectively. The training was carried out on Intel (R) Core (TM) i5-10210U 1.60GHz processor with 8.00 GB RAM. And the training time for a PRN is around 8 s. The performance of the training can be observed in the confusion matrix, which is shown in Fig. 5. The classification accuracy of the trained PRNs for correctly classifying the inputs to the target labels is about 98.3%. In order to evaluate the performance of the used PRN, classical 10-fold cross-validation is carried out, where the collected dataset for training and testing is randomly and repeatedly assigned [20]. The average classification accuracy is about 98.5%, which verifies the effectiveness of the employed PRN for attack detection. The RL agents are applied via dSPACE with a sampling time of 50 $\mu$s. Then, the trained cyber-attack detectors are implemented in the experimental testbed to verify its performance in the DC microgrid under the multi-agent RL-based cyber-attack.

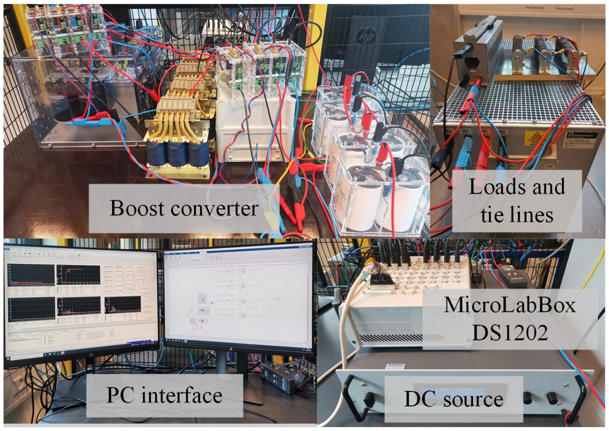


Fig. 4. Experimental setup of a DC microgrid with 4 nodes.

The experimental results are shown in Fig. 6. The RL agents target the sensor signals to produce destabilizing FDIAs on three neighboring nodes 1, 2, and 4. It can be observed when the RL-based attack is initiated, the agents generate the attacks on multiple nodes, and the system deviates from the normal operating condition. Moreover, from the DE terms shown in Fig. 6(c), it is evident that the RL algorithm successfully generates the sophisticated attacks that remain stealthy to the DE-based detection method as their values resemble the normal conditions counterparts and are maintained in their lower permissible range.

TABLE I. Experimental Setup Parameters

| Parameter sets | Values |
|---|---|
| Converter | $L_{in} = 0.86\ mH, C_o = 1.1\ mF, f_s = 10\ kHz, I_{rated} = 32\ A$<br>Loads: $R_1 = R_2 = R_3 = R_4 = 30.6\ \Omega$<br>Tie lines: $R_{12} = R_{23} = R_{34} = 0.5\ \Omega, R_{14} = 0\ \Omega$ |
| Controller | $V_{in} = 48\ V, V_{dc_{ref}} = 60\ V, I_{dc_{ref}} = 0$<br>Primary layer: $K_{pV} = 1, K_{iV} = 20, K_{pI} = 2.4, K_{iI} = 10$<br>Secondary layer: $K_p^I = 0.12, K_i^I = 0.15$ |

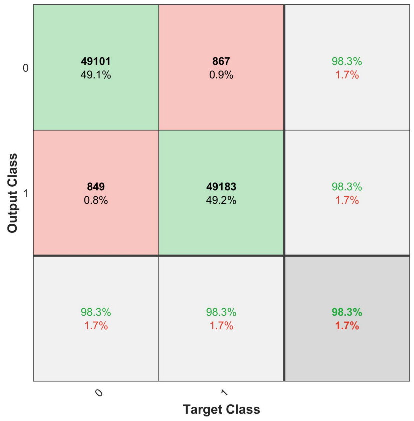


Fig. 5. Training results of the PRN for cyber-attack detection. (1:Attacked; 0:Healthy)

At the initial stage within 11 s, the system operates normally with well-tuned distributed control, where about 6 A current are shared proportionately among the four converters and output voltage for each converter converges to around 60 V. At around $t = 11$ s, the RL-based FDIA is initiated, which results in a 0.3 A current deviation on the compromised node and the unaffected node also experience a current rise to about 1.75 A, as shown in Fig. 6(b). It could also be observed in Fig. 6(a) that the output voltage of all the converters decreases by 0.05 V/s.

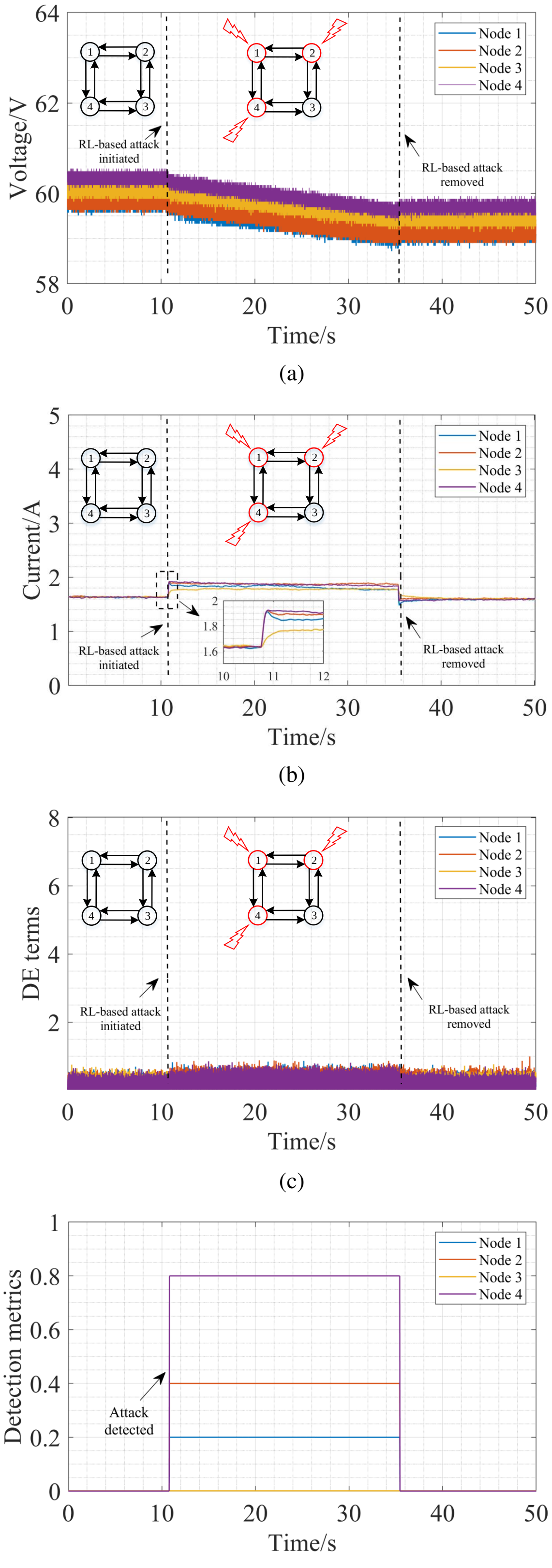
The performances of the conventional discordant detection method and proposed data-driven cyber-attack detection method are shown in Fig. 6(c) and (d). When the attack is initiated, it is observed in Fig. 6(c) that the DE terms for the compromised nodes are suppressed to the normal condition counterparts and within their minimum threshold. Neither of the attacked nodes is manifested with evident higher index values than the normal condition, according to the detection criteria in equation (5), clearly indicating that the conventional discordant detection method fails to detect the attacks and the attacked nodes under the RL-based cyber-attacks. For the proposed method, when the attack occurs, as shown in Fig. 6(d), with the labels scaled with the corresponding node index and 0.2 scaling factor, the well-trained PRN detectors signal out the compromised nodes 1, 2 and 4, while the output label for the node 3 is kept at 0, revealing that the cyber-attack occurs and the node 1, 2 and 4 are attacked. When the attack is removed, at around $t$ = 36 s, the system works normally at a new operating point and the detection metrics for all the four nodes are maintained at 0. From the above analysis, it can be concluded that the conventional discordant detection scheme is ineffective under the RL-based cyber-attack, and the proposed data-driven neural network based detector can detect the attacks and identify the attacked nodes for DC microgrids under the RL-based cyber-attacks.

## V. Discussion

Due to the system being regulated under the distributed cooperative control structure, the distributed multi-agent reinforcement learning algorithm can be applied to generate sophisticated attacks in larger systems. And the proposed data-driven detection method can still be employed to detect the attacks as long as the real operational data is collected. Moreover, in larger systems, as the distributed control is implemented independently in each microprocessor, the computational burden on the implementation of the proposed data-driven method and RL-based cyber-attack will not increase. Therefore, the RL-based cyber-attack and proposed data-driven detection method are scalable in larger systems with more nodes.

In addition, we can also design the RL-based attacker and train the proposed data-driven detector in an iterative way. In each iteration, by updating the model of the system protected with the new detection mechanism, i.e. the DE-based detection method and the up-to-date data-driven detector complementing each other, the RL-based attacker will explore in all the other attack types and exploit the vulnerability of the new detection mechanism, thus generating sophisticated attacks to bypass the new detection mechanism. In turn, by updating the database of the operation data under the new attacks, the data-driven cyber-attack detector is trained and implemented to detect the new attacks. Eventually, more attacks can be detected with the new detection mechanism.

To sum up, the proposed data-driven cyber-attack detection method can complement the DE-based detection method to detect the RL-based intelligent attacks and identify the attacked nodes which the conventional DE-based detection method fails to detect. Compared with the method in [7], which collects simulation data based on the healthy model of microgrids and can only detect a certain type of attack, the proposed method collects real data from system operation under both normal and attack conditions, which enable it to detect a wider range of cyber-attacks with high precision. Moreover, as the RL-based intelligent attacker can also learn to bypass other index-based detection methods [10]–[12], in the same way, the proposed data-driven attack detection method can be employed to complement other conventional cyber-attack detection methods.



(a)

(b)

(c)

(d)

Fig. 6. Experimental validation of the proposed data-driven detection method for the DC microgrid. (a) Output voltage, (b) Output current, (c) Performance of conventional DE-based detection method, and (d) Performance of proposed data-driven detection method.

## VI. Conclusions

This paper proposes a data-driven cyber-attack detection method to complement the DE-based detection method for DC microgrids under multi-agent RL-based cyber-attack. In particular, the multi-agent RL algorithm is employed to generate sophisticated attacks against the conventional DE-based detection method. The dataset of the DC microgrids operating under both normal and cyber-attack conditions is collected for offline training of the PRN based cyber-attack detectors. Then, the well-trained neural network based cyber-attack detectors are implemented in an experimental testbed to verify the performance of the proposed data-driven method. The experimental results show that the RL-based attacks remain undetected by the DE-based detection method as DE indices are maintained within their minimal permissible range, and the proposed data-driven detector works as a complementary detection scheme, detects the attacks and attacked nodes successfully. Moreover, the proposed detection mechanism could also be employed to complement other conventional cyber-attack detection approaches when they fail under the intelligent attacks.

## References

[1] T. Dragičević, X. Lu, J. C. Vasquez, and J. M. Guerrero, "Dc microgrids part i: A review of control strategies and stabilization techniques," *IEEE Transactions on power electronics*, vol. 31, no. 7, pp. 4876–4891, 2015.

[2] M. Yazdanian and A. Mehrizi-Sani, "Distributed control techniques in microgrids," *IEEE Transactions on Smart Grid*, vol. 5, no. 6, pp. 2901–2909, 2014.

[3] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Transactions on Information and System Security (TISSEC)*, vol. 14, no. 1, pp. 1–33, 2011.

[4] O. A. Beg, T. T. Johnson, and A. Davoudi, "Detection of false-data injection attacks in cyber-physical dc microgrids," *IEEE Transactions on industrial informatics*, vol. 13, no. 5, pp. 2693–2703, 2017.

[5] A. Karimi, A. Ahmadi, Z. Shahbazi, H. Bevrani, and Q. Shafiee, "On the impact of cyber-attacks on distributed secondary control of dc microgrids," in *2020 10th Smart Grid Conference (SGC)*, pp. 1–6. IEEE, 2020.

[6] A. S. Musleh, G. Chen, and Z. Y. Dong, "A survey on the detection algorithms for false data injection attacks in smart grids," *IEEE Transactions on Smart Grid*, vol. 11, no. 3, pp. 2218–2234, 2019.

[7] M. R. Habibi, S. Sahoo, S. Rivera, T. Dragičević, and F. Blaabjerg, "Decentralized coordinated cyber-attack detection and mitigation strategy in dc microgrids based on artificial neural networks," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, 2021.

[8] O. A. Beg, L. V. Nguyen, T. T. Johnson, and A. Davoudi, "Signal temporal logic-based attack detection in dc microgrids," *IEEE Transactions on Smart Grid*, vol. 10, no. 4, pp. 3585–3595, 2018.

[9] S. Sahoo, J. C.-H. Peng, A. Devakumar, S. Mishra, and T. Dragičević, "On detection of false data in cooperative dc microgrids:a discordant element approach," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 8, pp. 6562–6571, 2019.

[10] J. Zhang, S. Sahoo, J. C.-H. Peng, and F. Blaabjerg, "Mitigating concurrent false data injection attacks in cooperative dc microgrids," *IEEE Transactions on Power Electronics*, vol. 36, no. 8, pp. 9637–9647, 2021.

[11] D. Shi, P. Lin, Y. Wang, C.-C. Chu, Y. Xu, and P. Wang, "Deception attack detection of isolated dc microgrids under consensus-based distributed voltage control architecture," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 11, no. 1, pp. 155–167, 2021.

[12] S. Sahoo, T. Dragičević, and F. Blaabjerg, "An event-driven resilient control strategy for dc microgrids," *IEEE Transactions on Power Electronics*, vol. 35, no. 12, pp. 13 714–13 724, 2020.

[13] C. Wang, J. Wang, Y. Shen, and X. Zhang, "Autonomous navigation of uavs in large-scale complex environments: A deep reinforcement learning approach," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 3, pp. 2124–2136, 2019.

[14] Y. Chen, S. Huang, F. Liu, Z. Wang, and X. Sun, "Evaluation of reinforcement learning-based false data injection attack to automatic voltage control," *IEEE Transactions on Smart Grid*, vol. 10, no. 2, pp. 2158–2169, 2018.

[15] Z. Zhang, D. Zhang, and R. C. Qiu, "Deep reinforcement learning for power system applications: An overview," *CSEE Journal of Power and Energy Systems*, vol. 6, no. 1, pp. 213–225, 2019.

[16] Z. Ni and S. Paul, "A multistage game in smart grid security: A reinforcement learning solution," *IEEE transactions on neural networks and learning systems*, vol. 30, no. 9, pp. 2684–2695, 2019.

[17] V. Nasirian, S. Moayedi, A. Davoudi, and F. L. Lewis, "Distributed cooperative control of dc microgrids," *IEEE Transactions on Power Electronics*, vol. 30, no. 4, pp. 2288–2303, 2014.

[18] C. J. Watkins and P. Dayan, "Q-learning," *Machine learning*, vol. 8, no. 3-4, pp. 279–292, 1992.

[19] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski *et al.*, "Human-level control through deep reinforcement learning," *nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[20] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An introduction to statistical learning*, vol. 112. Springer, 2013.