

## ORIGINAL RESEARCH

# An unsupervised cyberattack detection scheme for AC microgrids using Gaussian process regression and one-class support vector machine anomaly detection

Jeewon Choi<sup>1</sup>  | Behshad Roshanzadeh<sup>2</sup>  | Manel Martínez-Ramón<sup>2</sup>  | Ali Bidram<sup>1,2</sup> 

<sup>1</sup>Department of Mechanical Engineering, University of New Mexico, Albuquerque, NM, USA

<sup>2</sup>Department of Electrical and Computer Engineering, The University of New Mexico, Albuquerque, NM, USA

## Correspondence

Jeewon Choi, Department of Mechanical Engineering, University of New Mexico, Albuquerque, NM 87131, USA.  
Email: [chatchi923@unm.edu](mailto:chatchi923@unm.edu)

## Funding information

National Science Foundation, Grant/Award Numbers: ECCS-2214441, OIA-1757207; King Felipe VI endowed Chair of the University of New Mexico

## Abstract

This paper addresses the cybersecurity of hierarchical control of AC microgrids with distributed secondary control. The false data injection (FDI) cyberattack is assumed to alter the operating frequency of inverter-based distributed generators (DGs) in an islanded microgrid. For the microgrids consisting of the grid-forming inverters with the secondary control operating in a distributed manner, the attack on one DG deteriorates not only the corresponding DG but also the other DGs that receive the corrupted information via the distributed communication network. To this end, an FDI attack detection algorithm based on a combination of Gaussian process regression and one-class support vector machine (OC-SVM) anomaly detection is introduced. This algorithm is unsupervised in the sense that it does not require labelled abnormal data for training which is difficult to collect. The Gaussian process model predicts the response of the DG, and its prediction error and estimated variances provide input to an OC-SVM anomaly detector. This algorithm returns enhanced detection performance than the standalone OC-SVM. The proposed cyberattack detector is trained and tested with the data collected from a 4 DG microgrid test model and is validated in both simulation and hardware-in-the-loop testbeds.

## 1 | INTRODUCTION

Modern power systems are highly exposed to cyber threats due to the deployment of communication and control technologies for different applications like microgrid distributed control, energy management systems, wide area protection, monitoring, etc. [1]. In the cyber layer of power systems, both control and communication entities can be potential targets for cyber threats. Cyberattacks can be of different types and natures. In ref. [2], 63 different types of cyberattacks in power grids are listed. False data injection (FDI) attacks target the sensors and control and decision-making units which in turn corrupt the data transferred through the communication links and impact the data integrity [3–5]. Denial-of-service (DoS) attacks interrupt the availability of communication system services. Cyberattacks can endanger system voltage and frequency stability which in turn (i) cause cascading failures and power outages for customers, (ii) slow down the control system responses,

and (iii) overload and violate the system equipment thermal limits. On the other hand, power system intruders can cause a physical attack and create a cyberattack to mask the physical attack which later can result in a cascading failure across the power system. Some of the real-world cyberattacks include the Slammer worm of the David-Besse nuclear plant in Ohio in 2003, the Ukraine cyberattack in December 2015 which is considered the worst blackout caused by a cyberattack where 225,000 customers lost power, Malware Triton in the Saudi Arabian oil refinery in 2017, the cyberattack in the U.S. power utilities in March 2019, the cyberattack in Kudankulam Nuclear Power Plant in India in 2019, and Man-in-the-Middle attack in Nuclear Power Corporation of India [2]. The world economic forum has ranked large-scale cyberattacks as risks threatening countries in the next 10 years [6].

Microgrids as small-scale power grids that can operate in both grid-connected and islanded modes, also extensively utilize communications and therefore are highly exposed to

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *IET Renewable Power Generation* published by John Wiley & Sons Ltd on behalf of The Institution of Engineering and Technology.

cyberattacks. Especially, the secondary control level of microgrids in their hierarchical control structure [7–12] uses the centralized or distributed communication networks to perform voltage regulation and frequency restoration. The focus of this paper is on the distributed secondary control of microgrids, where the impact of an FDI attack on one of the distributed generators (DGs) can be propagated along the whole microgrid system to other DGs jeopardizing the microgrid's stable operation.

The bulk of the research in the cybersecurity of microgrids focuses mainly on attack detection based on data-driven techniques. FDI attacks are very common in microgrids and are addressed by different techniques including adaptive cumulative sum using Markov-chain analysis, Kalman filters, graphical method, model-based scheme, matrix separation technique, Chi-square detector, cosine similarity matching approach, and nonlinear internal observer are introduced for the attack detection in power systems with centralized control structure [13–19]. Attack detection strategies for a distributed control system using signal temporal logic [20] and Kalman filter-based unknown input observer [21] have been introduced in some recent works. In refs. [4, 5], the attack detection mechanism deploys Kullback–Liebler (KL) divergence to measure the discrepancy between the Gaussian distributions of the actual and expected measurements. More recently, machine learning and artificial intelligence have been utilized for cyberattack detection in electric power grids. Most of these techniques rely on supervised artificial neural networks or deep learning algorithms [22–24]. The existing cyberattack prevention techniques have the following limitations and challenges: (i) The conventional data-driven techniques rely on a fixed threshold to detect attacks that can be manipulated by intelligent attacks, and (ii) the supervised machine learning algorithms require a large amount of data to train which can be hard to gather due to the low probability and occurrence of cyberattacks.

This paper addresses the cybersecurity of hierarchical control of AC microgrids with distributed secondary control. To this end, an unsupervised FDI cyberattack detection scheme based on a combination of Gaussian Process (GP) regression and support vector machine (SVM) based anomaly detection algorithm is introduced. The FDI attack is assumed to alter the operating frequency of inverter-based DGs in an islanded microgrid.

The uncertain nature of FDI attacks and the impossibility of collecting real data sufficient to train machine learning algorithms in a supervised fashion highlight the necessity of unsupervised strategies that are capable of detecting cyberattacks that were previously unseen. The proposed technology aims to design structures and algorithms that treat the cyberattack as an anomalous event. The algorithms must be unsupervised to apply them where no data is available for training, the data is uncategorized, or it does not contain the events of interest (cyberattacks). The methods must be trainable with a small amount of data. Therefore, they need to have strong regularization properties. They must be adaptive to account for nonstationary scenarios. Finally, the number

of free or nontrainable parameters must be low, and they must be chosen by inference from the available data or by reasonable heuristics. This paper uses a GP regression tool to estimate the DG's operating frequency using active power and rate of change of active power as inputs. The estimated DG's frequency is then compared against the actual DG frequency measurement to create an error term that can describe the presence of a cyberattack. The calculated error term is used as an input to a one-class SVM (OC-SVM) to detect cyberattacks.

The contributions of this paper are as follows:

- An unsupervised cyberattack detection scheme is proposed that does not rely on a fixed threshold to detect cyberattacks.
- The proposed unsupervised machine learning algorithm does not require labelled data for training which is advantageous due to the nature of cyberattacks that are unknown, unpredictable, and low in probability.
- The output of the cyberattack detection algorithm is directly used for point-of-attack isolation to mitigate the impact of an attack on the operation of the microgrid.

The rest of the paper is organized as follows. Section 2 demonstrates the control system of a grid-forming AC microgrid that is a potential target of cyberattacks. Section 3 describes how an FDI attack can harm the microgrid and introduces a machine learning-based detection algorithm for such cyberattacks, as well as a mitigation strategy against them. Section 4 elaborates on the training and testing procedures for an attack detector and exhibits the performance of the trained machine. Section 5 provides the validation of the proposed attack detection algorithm in MATLAB/Simulink simulation testbed and HIL simulation testbed. Section 6 is the conclusion.

## 2 | PRELIMINARIES OF MICROGRID DISTRIBUTED SECONDARY FREQUENCY CONTROL

This work targets an AC microgrid that is composed of inverter-based DGs operating in a grid-forming mode, i.e., islanded from the main grid. A microgrid is controlled by three control hierarchies, namely the primary, secondary, and tertiary controls introduced in refs. [7, 9]. The focus of this paper is on the primary and secondary control levels. The grid-forming inverters can form and maintain the stable voltage and frequency of an islanded microgrid initially by the primary control. One of the control objectives in the primary control layer is to allocate power (loads) over the DGs proportional to their power ratings. The droop method is deployed for this power-sharing control. In general, active power sharing and reactive power sharing are paired with frequency and voltage control, respectively. For the present work, only the frequency droop is considered.

The frequency droop control is locally implemented in each of the inverter's primary control layers. The frequency droop

characteristic of DG  $i$  can be described by

$$\omega_i = \omega_{ni} - m_{P_i} P_i \quad (1)$$

where  $\omega_i$  is the angular frequency of DG  $i$ ,  $\omega_{ni}$  is the reference for the droop control,  $m_{P_i}$  is the droop coefficient that is selected based on the power rating of the DG, and  $P_i$  is the active power. The goal of the power-sharing control is to maintain the power supply ratio equal for all of the DGs based on their ratings. For example, for a microgrid that is composed of  $N$  DGs, this can be expressed as

$$\frac{P_1}{P_{\max,1}} = \dots = \frac{P_N}{P_{\max,N}} \quad (2)$$

where  $P_{\max,i}$  is the maximum power rating of DG  $i$ . This can be equivalently written as

$$m_{P_1} P_1 = \dots = m_{P_N} P_N \quad (3)$$

where  $m_{P_i}$  can be considered as an interchangeable specification of  $P_{\max,i}$ . Along with  $P_i$  being set by the active power sharing regulation in Equation (3), the operating angular frequency  $\omega_i$  can be determined by Equation (1). As the equation implies,  $\omega_i$  is a linear function of  $P_i$  that is decreasing from the  $y$ -intercept  $\omega_{ni}$  with the slope  $-m_{P_i}$ . Therefore, during the power-sharing control,  $\omega_i$  is likely to deviate from the primary reference  $\omega_{ni}$ . This deviation can be restored by resetting  $\omega_{ni}$ , which will be implemented in the secondary control layer.

The secondary control construction can start from differentiating the droop characteristic (1)

$$\dot{\omega}_{ni}(t) = \dot{\omega}_i(t) + m_{P_i} \dot{P}_i(t) \quad (4)$$

and the auxiliary frequency and active power control inputs  $u_{\omega_i}$  and  $u_{P_i}$  are respectively defined as

$$\dot{\omega}_i(t) = u_{\omega_i}(t) \quad (5)$$

$$m_{P_i} \dot{P}_i(t) = u_{P_i}(t) \quad (6)$$

From Equation (4), the droop reference  $\omega_{ni}$  can be written as

$$\omega_{ni} = \int (u_{\omega_i} + u_{P_i}) dt \quad (7)$$

Now, we consider the two objectives of the secondary control. One of the objectives is to synchronize the DG frequencies to the reference frequency of  $\omega_{\text{ref}}$ , i.e.

$$\lim_{t \rightarrow \infty} \omega_i = \omega_{\text{ref}}, \quad \forall i \quad (8)$$

The other objective is to synchronize the active power ratios of DGs as already mentioned in Equation (3). Based on these objectives, the auxiliary frequency and active power control inputs (5) and (6) can be designed in a distributed fashion as in Equations (9) and (10), respectively.

$$u_{\omega_i}(t) = c_i \left( \sum_{j \in \mathcal{N}_i} a_{ij} (\omega_j(t) - \omega_i(t)) + g_i (\omega_{\text{ref}} - \omega_i(t)) \right) \quad (9)$$

$$u_{P_i}(t) = c_i \sum_{j \in \mathcal{N}_i} a_{ij} (m_{P_j} P_j(t) - m_{P_i} P_i(t)) \quad (10)$$

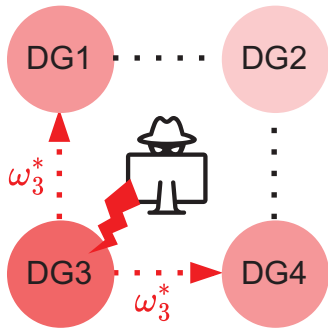
Here,  $c_i$  is a positive control gain.  $a_{ij}$  is a component that indicates whether there is a communication link between DG  $i$  and  $j$ . If DG  $i$  has an incoming communication from DG  $j$ , then  $a_{ij} = 1$  (or  $a_{ij} > 0$ ), otherwise  $a_{ij} = 0$ .  $\mathcal{N}_i$ , the in-neighbours of node  $i$  is a set of nodes that have an outgoing communication link to node  $i$ . A pinning gain  $g_i$  is an indication of whether the node  $i$  receives the reference point  $\omega_{\text{ref}}$  from the leader node or not. If the DG  $i$  has an incoming communication link from the leader node  $g_i = 1$  or  $g_i > 0$ , otherwise  $g_i = 0$ . For the hierarchical microgrid control system, the leader node would be the tertiary control layer. However, there will not be an update from the tertiary control during the grid-disconnected mode, therefore it does not need to be considered a communication link.

### 3 | ATTACK DETECTION AND MITIGATION METHODOLOGY

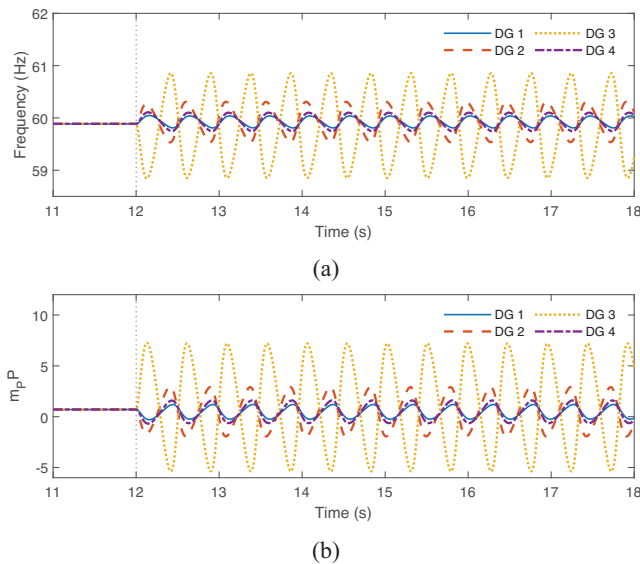
In this section, first, FDI attacks are introduced and the impacts of the FDI attack on the frequency droop with the distributed secondary control are demonstrated. Then, the attack detection scheme is proposed. Finally, the attack mitigation scheme is discussed.

#### 3.1 | Description of FDI attack

In this paper, the FDI attack under study targets the primary and secondary frequency control of AC microgrids. On an inverter-based DG, the primary and distributed secondary control protocols discussed in Section 2 are virtually implemented on a microprocessor that has communication ports. These communication ports and gateways are used by the distributed secondary control level to communicate DG's information to the neighbouring DGs as stated in Equations (9) and (10). However, an attacker can gain access to these communication ports and tamper with control protocols and parameters as discussed in ref. [3]. Herein, it is assumed that the attacker targets the primary frequency droop in Equation (1) and replaces  $\omega_i$  with  $\omega_i^*$  which is the false data injected by the attacker. As illustrated in Figure 1, in the control system that is operating in a distributed manner, an attack on one node does not only affect the corresponding DG but also the rest of the DGs in the network. The false information injected  $\omega_i^*$  can invade the other DGs' controllers through communication as described in Equation (9). In the microgrids with centralized control architecture, the attacker can directly target the central controller and the central controller should be equipped with attack detection



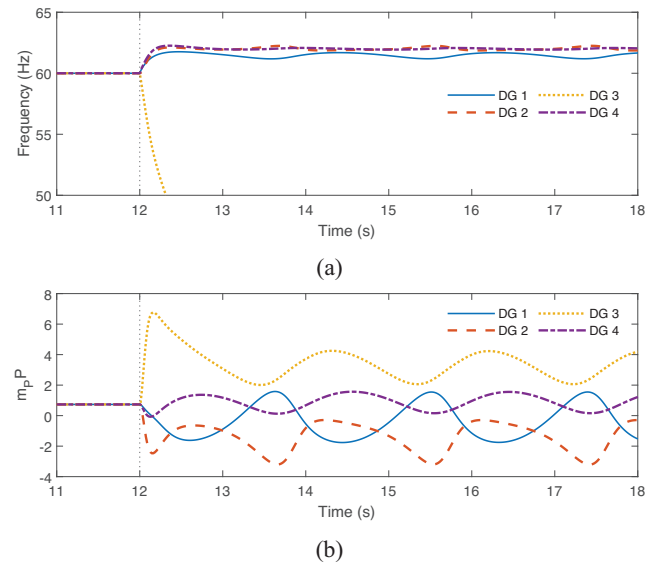
**FIGURE 1** FDI attack on one DG not only adversely impacts the DG under attack itself but also other DGs that are connected via a communication network.



**FIGURE 2** The impact of FDI attack on frequency droop without distributed secondary control activated: (a) DG frequencies; (b) active power ratios  $m_P P_i$ .

schemes in order to stop the spread of the attack to individual DGs. However, in the distributed control architecture, there is no central controller that the attacker can target and the attacker can directly target the gateway of microprocessors located on individual DGs. That is why in a distributed control architecture each DG should have its own attack detection and mitigation mechanism.

In order to show the impact of FDI attack on the frequency droop,  $\omega_i$  of one DG in the microgrid test system of Figure 5 is manipulated with the corrupted frequency of 62 Hz without the secondary control activated, as shown in Figure 2. Figure 2(a,b) shows the actual frequencies and active power ratios of DGs. Even though the corrupted information is not shared among the DGs for the moment as the secondary control is deactivated, the impact of the attack is spread out through the power network. As expected, the system loses its stability, and the frequencies and the active power ratios oscillate. In addition, the objective of primary control (3) is not achieved. On the other hand, when the secondary control is activated, the cor-



**FIGURE 3** The impact of FDI attack on frequency droop with distributed secondary control activated: (a) DG frequencies; (b) active power ratios  $m_P P_i$ .

rupted information is propagated through the communication network, adding to the impact induced by the power network.  $\omega_{ni}$  in Equation (1) is contaminated by the faulty information without satisfying the control objectives (3) and (8) as shown in Figure 3.

### 3.2 | Attack detection scheme

In summary, the proposed FDI attack detection algorithm is based on a combination of GP regression and OC-SVM anomaly detection. This algorithm is unsupervised in the sense that it does not require labeled abnormal data for training which is difficult to collect. The GP model predicts the operating frequency of the DG using its active power ratio and rate of change of active power as an input. This predicted value of DG's frequency is compared against the DG's actual frequency. DG's prediction error and estimated variances provide input to an OC-SVM anomaly detector. The OC-SVM detector highlights an attack scenario when the calculated error is large enough. In the following subsections, the details of GP regression, OC-SVM, and attack detection mechanisms using these tools are elaborated. A reasonable strategy to detect attacks is to estimate the likelihood of the observations or, likewise, to estimate how probable an observation is. The machine learning methods that implement this criterion are usually called anomaly or novelty detection algorithms. The classical ones are based on a direct probabilistic model of the data, that is used to measure the likelihood (see e.g. refs. [25, 26], that use the well-known K-nearest neighbours (KNN) algorithm), and others make an indirect measure that does not contain an explicit probabilistic model. Two prominent algorithms for novelty detection are the OC-SVM [27, 28] and the support vector data description (SVDD) [29], which make use of the maximum margin



criterion. In this approach, we will make use of a combination of a probabilistic model based on a GP regression for the observations and an OC-SVM, which is introduced in ref. [30].

### 3.2.1 | One class support vector machine

The OC-SVM algorithm is rooted in the principle of structural risk minimization (SRM) introduced by V. Vapnik in 1982 (see e.g. ref. [31]). An SVM machine can be optimized by minimizing a risk function that is a measurement of a discrepancy between the output and the label, i.e. the probability of error. However, such risk cannot be calculated if the distribution of the data is unknown. An empirical risk that can be measured over the data can be instead taken into consideration. However, minimizing the empirical risk increases the model complexity, then SRM is introduced in order to limit the growth of complexity. One can construct a criterion in a way that optimizes a trade-off between the empirical and structural error. The empirical and structural risks can be minimized through margin maximization, which is equivalent to

$$\begin{aligned} \text{minimize } L_p &= \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N \xi_i \\ \text{subject to } y_i \hat{y}_i &> 1 - \xi_i, \quad \xi_i \geq 0 \end{aligned} \quad (11)$$

where  $L_p$  is the primal loss function,  $\mathbf{w}$  is a vector of machine parameters,  $C$  is a free trade-off parameter,  $\xi_i$  is the slack variable that lies within the margin but at the wrong side of the hyperplane, and  $\hat{y}_i = \mathbf{w}^T \mathbf{x}_i + b$  is the estimated output of the input  $\mathbf{x}_i$  where  $b$  is a bias. It can be proven that minimizing the norm of vector  $\mathbf{w}$  under the conditions of functional (11) minimizes the structural error and it is equivalent to maximizing the so-called classification margin.

SRM was originally applicable to the linear model only, however, the problem can be easily transformed to nonlinear by kernel trick. The input data can be mapped into a higher or possibly infinite dimensional Hilbert space using a nonlinear transformation  $\varphi(\cdot)$ , where Hilbert spaces are provided with a kernel dot product  $k(\mathbf{x}, \mathbf{x}') = \varphi(\mathbf{x})^T \varphi(\mathbf{x}')$  [32].

Considering the nature of abnormal or novel events that are much less likely to happen than normal ones, it may not be easy to train a cyberattack detector in a supervised way as it will be hard to observe and label the anomalies. The fact that such novel behaviour is random and unpredictable makes it more unreasonable to use a supervised learning algorithm for cyberattack detection.

The OC-SVM can provide unsupervised learning for anomaly detection. As the name implies there is only one class in data, therefore no labeled training data are necessary. In this approach, the strategy is to separate the normal data from the origin. Therefore the machine can be optimized by maximizing the distance of the hyperplane to the origin. With the function  $f(\mathbf{x}_i) = \mathbf{w}^T \mathbf{x}_i - b$  where  $b$  is a bias, the formulation of the

primal optimization can be written as

$$\begin{aligned} \text{minimize } L_p &= \frac{1}{2} \|\mathbf{w}\|^2 + \frac{1}{\nu N} \sum_{n=1}^N \xi_n - \rho \\ \text{subject to } &\begin{cases} \mathbf{w}^T \varphi(\mathbf{x}_j) \geq \rho - \xi_j \\ \xi_j \geq 0 \end{cases} \end{aligned} \quad (12)$$

where  $0 < \nu < 1$ .  $\nu$  is the portion of outlier data in the training set.  $\frac{1}{\nu N}$  is a trade-off between the empirical and structural risks. Using the Lagrange optimization, the corresponding dual can be expressed as

$$\begin{aligned} \text{minimize } L_p &= \frac{1}{2} \boldsymbol{\alpha}^T \mathbf{K} \boldsymbol{\alpha} \\ \text{subject to } &\begin{cases} 0 \leq \alpha_n \leq \frac{1}{\nu N} \\ \sum_{n=1}^N \alpha_n = 1 \end{cases} \end{aligned} \quad (13)$$

where  $\boldsymbol{\alpha}$  is a vector of Lagrange multipliers  $\alpha_i$  and  $\mathbf{K}$  is a matrix of all kernel dot products  $\mathbf{K}_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$ .

For the present work, the square exponential kernel function,  $k(\mathbf{x}, \mathbf{x}') = \exp(-\gamma \|\mathbf{x} - \mathbf{x}'\|)$ , also known as radial basis function (RBF) kernel is considered. Then, the two free parameters to be adjusted are  $\nu$  and  $\gamma$ . While  $\nu$  can be adjusted to a small value that will be an upper bound of the outliers, parameter  $\gamma$  may largely vary depending on the nature of the observed data. Therefore, we propose to use a GP to standardize the data in an unsupervised way prior to its use in the OC-SVM. This way, the cross-validation of parameter  $\gamma$  is simplified.

### 3.2.2 | Gaussian process

The general form of the GP estimator with input transformed into a Hilbert space can be written as  $f(\mathbf{x}_i) = \mathbf{w}^T \varphi(\mathbf{x}_i)$ , and the observation is  $y_i = f(\mathbf{x}_i) + e_i$ . The estimation error  $e_i$  is many times referred to as a noise term in the literature. Note that bias  $b$  is included in  $\mathbf{w}$  and the input is accordingly formed as  $\mathbf{x}_i = \langle x_1 \dots x_d \ 1 \rangle^T$ . In the GP approach, the error is assumed to be Gaussian distribution, i.e.  $e_i \sim N(0, \sigma_n^2)$ .

The prediction  $f_*$  of new sample  $\mathbf{x}_*$  that does not belong to the training data, follows the distribution  $f_* \sim N(\bar{f}_*, \sigma_*^2)$ . Given the training data set  $[\mathbf{X}, \mathbf{y}]$ , the mean and variance of the prediction are found as

$$\begin{aligned} \bar{f}_* &= \mathbf{k}(\mathbf{x}_*) (\mathbf{K} + \sigma_n^2 \mathbf{I})^{-1} \mathbf{y} \\ \sigma_*^2 &= k(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{k}(\mathbf{x}_*)^T (\mathbf{K} + \sigma_n^2 \mathbf{I})^{-1} \mathbf{k}(\mathbf{x}_*) \end{aligned} \quad (14)$$

where  $\mathbf{k}(\mathbf{x}_*)$  is a vector whose elements  $k(\mathbf{x}_*, \mathbf{x}_i)$  are the kernel dot products of the test sample  $\mathbf{x}_*$  and  $N$  training samples  $\mathbf{x}_i$ ,  $\mathbf{K}$  is a  $N \times N$  matrix whose elements  $k_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$  are the dot kernel product between every training samples, and  $\mathbf{I}$  is a

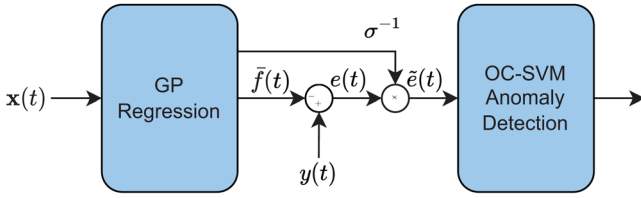


FIGURE 4 Structure of the proposed attack detection scheme.

$N \times N$  identity matrix. For the present work, a dot product kernel with scale factor  $\sigma_1^2$  and bias kernel  $\sigma_2^2$  is constructed as  $k(\mathbf{x}, \mathbf{x}') = \sigma_1^2 \mathbf{x}^\top \mathbf{x}' + \sigma_2^2$  [33].

The error variance  $\sigma_n^2$ , the kernel parameters  $\sigma_1^2$  and  $\sigma_2^2$  are the free parameters that need to be optimized. For optimization purpose, the loss function of the predictor can be defined by the negative log marginal likelihood, where the marginal likelihood is a likelihood function that is integrated over the parameter space. The loss function can be minimized using gradient descent. Note that the loss function may not always be convex, therefore initialization with different values may be desired. Thanks to the optimization available using the marginal likelihood, cross-validations over the parameters are not necessary.

### 3.2.3 | Attack detection algorithm

The attack detection algorithm proposed is structured in three steps, GP prediction, error standardization, and OC-SVM detection, as shown in Figure 4. At instant  $t$ , a GP block takes an input  $\mathbf{x}(t) \in \mathbb{R}^D$  and estimates  $\hat{y}(t) \in \mathbb{R}$ , which are respectively a vector of dimension  $D$  containing windows of samples of the active power ratio and its time derivative, i.e.  $\mathbf{x}(t) = [m_P P(t), m_P P(t-1), \dots, m_P \dot{P}(t), m_P \dot{P}(t-1), \dots]^\top$ , and the angular frequency estimation at instant  $t$ . The GP also outputs the predictive variance  $\sigma_*^2(t)$  of (14). The GP regression model is trained with sets of data collected during normal operations. The expectation here is that the prediction error is small during normal operation while it is larger when an attack is applied to the system.

After the training phase, the GP predictor enters the testing phase. More normal data is inputted to the GP predictor, and its prediction error  $e(t)$  is calculated. The error is the difference between the observed  $y(t)$  and the predicted mean of the estimation  $\hat{f}(t)$ . The error is standardized with the estimated predictive variance  $\sigma_*^2(t)$  and the error variance  $\sigma_n^2$ . The purpose of error standardization is to distinguish the high prediction error that occurs during normal operation from that of attack cases. For example, high prediction errors can be observed during drastic transients induced by normal disturbances such as load changes or DG connection. Such high errors during normal operations occur with high variance, while the high errors caused by an attack are expected to have a low variance. Therefore, after the standardization with the standard deviation, the high error with the attack will only remain high. A new vector  $\tilde{\mathbf{e}}(t) = [\tilde{e}(t), \tilde{e}(t-1), \dots]^\top$  consisting of a window of  $D_e$  samples

of the standardized error  $\tilde{e}(t) = \sigma^{-1}(t)e(t)$  is then used as an input for OC-SVM, where  $\sigma^2 = \sigma_*^2 + \sigma_n^2$ .

### 3.3 | Attack mitigation scheme

When the attack detector of DG  $i$  detects an attack, the corresponding DG will be shut down from the microgrid to ensure a stable operation of the microgrid. The information of attacked DG will as well be isolated from the communication network, by setting  $a_i^* = 0$  in Equation (15), which is by default  $a_i^* = 1$  before any attack is detected. This information is delivered to the neighbouring controllers so that the incoming contaminated data can be isolated in the neighbouring controllers as well.

$$u_{\omega_i}(t) = c_{\omega_i} \sum_{j \in \mathcal{N}_i} a_{ij} a_i^* a_j^* (\omega_j(t) - \omega_i(t)) + g_i (\omega_{\text{ref}} - \omega_i(t))$$

$$u_{P_i}(t) = c_{P_i} \sum_{j \in \mathcal{N}_i} a_{ij} a_i^* a_j^* (m_{P_j} P_j(t) - m_{P_i} P_i(t)) \quad (15)$$

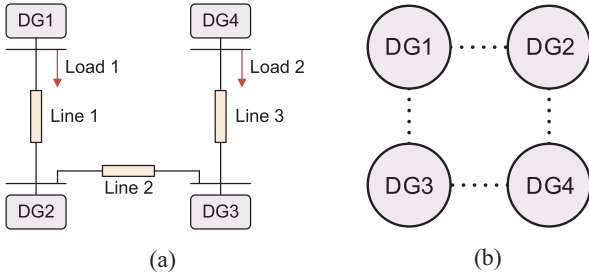
*Remark 1.* When the microgrid is operating in the normal condition, i.e. in the absence of cyberattack, the distributed control protocol (7) provides stable synchronization of the microgrid's frequency to the reference frequency and satisfies the sharing of active powers among DGs according to Equation (3), if the communication graph has a spanning tree and  $g_i$  in Equation (9) is nonzero for at least one root node. This is proved in ref. [12]. When the proposed cyberattack detection scheme detects an attack on a DG, the proposed attack mitigation algorithm ignores the information shared by that DG. In this case, if the communication graph connecting the remaining DGs still has a spanning tree and  $g_i$  is nonzero for at least one root node, then the distributed control protocol (7) provides stable synchronization of the microgrid's frequency and DGs' active power ratios.

## 4 | ATTACK DETECTOR TRAINING AND TESTING

The experiment first tested the effectiveness of the proposed methodology and then compared the results with a standard fault detection scheme. In the present work, 4 DG microgrid test system illustrated in Figure 5 with the specifications in Table 1 is considered. The training and testing data are collected from the simulation for each DG, and the attack detector is trained for each DG.

### 4.1 | Description of data

To collect training and testing datasets, a wide range of simulations are performed in Matlab/Simulink to account for different conditions of a microgrid. Due to the unsupervised nature of algorithms, both GP and OC-SVM algorithms are



**FIGURE 5** 4 DG microgrid test system: (a) circuit diagram; (b) communication graph.

**TABLE 1** Specifications of 4 DG microgrid test system.

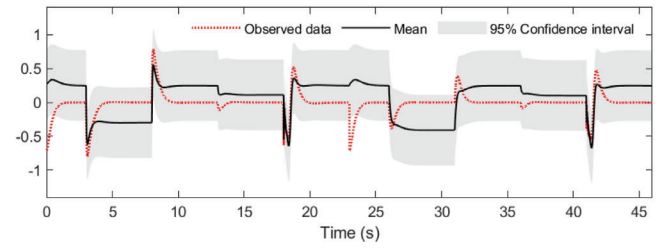
DGs	DG 1, 4	DG 2, 3	
$m_P$	$1 \times 10^{-4}$	$2 \times 10^{-4}$	
$n_Q$	$2 \times 10^{-3}$	$4 \times 10^{-3}$	
$R_c$	$0.05 \ \Omega$	$0.05 \ \Omega$	
$L_c$	4.8 mH	4.8 mH	
$R_f$	$0.1 \ \Omega$	$0.1 \ \Omega$	
$L_f$	1.35 mH	1.35 mH	
$C_f$	50 $\mu$ F	50 $\mu$ F	
$K_{PV}$	0.1	0.05	
$K_{IV}$	420	390	
$K_{PC}$	15	10.5	
$K_{IC}$	20000	16000	
Lines	Line 1	Line 2	Line 3
$R$	$0.2 \ \Omega$	$0.1 \ \Omega$	$0.2 \ \Omega$
$L$	3.6 mH	1.8 mH	3.6 mH
Loads	Load 1	Load 2	
$P$	12 kW	12 kW	
$Q_L$	5 kVAr	5 kVAr	

trained with system normal data (in the absence of cyberattacks). The system's normal scenarios include the simulation of a microgrid under islanding, load changes, and DG connection/disconnections. The simulations account for a load change in the range of 0 to 20 kW.

Each set of data is sampled every 100  $\mu\text{s}$ . Thirteen datasets without attack are collected, where each dataset includes the transient scenarios of islanding, load changes, and DG connection and re-connection. Two sets of data under random attack applied are also collected for evaluation purposes. Each normal dataset contains 180 to  $250 \times 10^3$  samples. For the OC-SVM test, 30% of faulty data are included out of 259,975 samples.

## 4.2 | Performance of the GP frequency predictor

For the GP frequency predictor, a window of two measurements of parameters  $m_{\beta}P_i(t)$  and  $m_{\beta}\dot{P}_i(t)$  is used, i.e.  $\mathbf{x}_i(t) =$



**FIGURE 6** The trained GP predictor tested with normal datasets. Test data fall within 95% confidence interval.

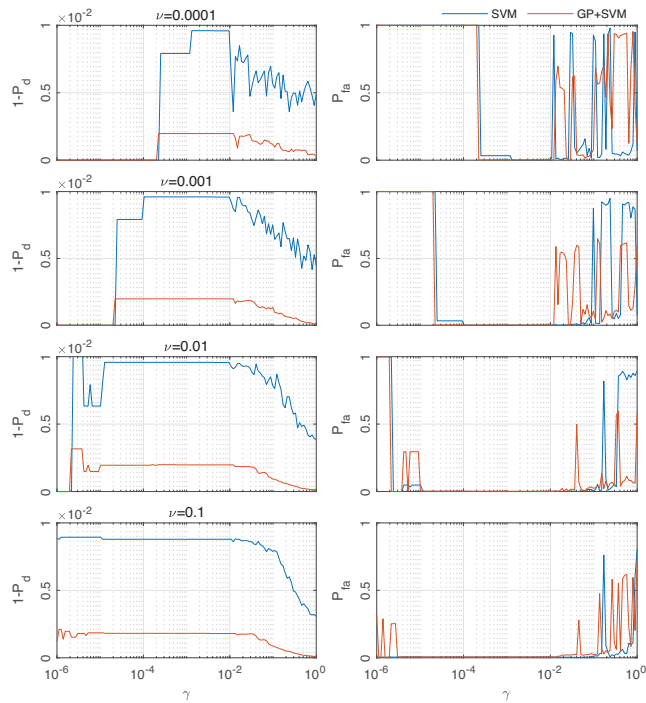
$[m_{\beta}P_i(t), m_{\beta}P_i(t-1), m_{\beta}\dot{P}_i(t), m_{\beta}\dot{P}_i(t-1)]^T$ . The error variance  $\sigma_n^2$ , the kernel parameters  $\sigma_1^2$  and  $\sigma_2^2$  are initialized to 0.1, 1, and 0, respectively, and optimized by using the negative log marginal likelihood minimization, therefore the cross-validation is unnecessary. The training is implemented in Python script and the Gpytorch library is used. We aim for the predictor to have 95% confidence. Three sets of normal data are used for GP training. The trained GP predictor is tested with two sets of normal data, and more than 95% of the test data fell within the 95% confidence interval as given in Figure 6. The predictor is additionally tested with the rest of the normal data, and in all of the datasets, more than 95% of the samples are observed within the confidence interval.

## 4.3 | Performance of OC-SVM attack detector

For the training of the OC-SVM attack detector, the window of 5 standardized errors is used, i.e.  $\tilde{\mathbf{e}}_i(t) = [\tilde{e}(t), \tilde{e}(t-1), \tilde{e}(t-2), \tilde{e}(t-3), \tilde{e}(t-4)]$ . Two of the normal datasets that were used to test the GP predictor are used to train OC-SVM. One of the normal data and two faulty data sets are used to test OC-SVM.

Cross-validation is required for the parameters,  $\gamma$ , and  $\nu$ . The probability of detection loss  $1 - P_d$ , where  $P_d$  is the probability of attack detection, and the probability of false alarm  $P_{fa}$ , are chosen as the metrics to be observed for parameter validation. In the OC-SVM module in Scikit-learn that is used, the normal and abnormal events are labeled 1 and -1, respectively. Following that convention, the probability of detection and false alarm can be calculated by  $P_d = \text{TN}/(\text{TN} + \text{FP})$ , and  $P_{fa} = \text{FN}/(\text{FN} + \text{TP})$ , respectively. Note that for the test purpose, the labelled normal and attack data are required, therefore from the test phase the problem can no longer be considered unsupervised.

OC-SVM is trained over the parameters spaced in logarithmic scale,  $10^{-6} \leq \gamma \leq 10^0$ , and  $10^{-4} \leq \nu \leq 10^{-1}$ , and the metrics are calculated for every combination of parameters with the test data. The proposed detector is compared with the standard OC-SVM detector without the regression entity. For the standalone OC-SVM detector, the window of the angular frequency of size 5 is used as input. The same training and test data are used for a fair comparison. The loss of detection and the false alarm observed for each detector, are shown in Figure 7. As shown in the figure, there is no significant difference in the rate of false alarms between the detectors; within certain ranges of  $\gamma$  where



**FIGURE 7** Probability of detection loss  $1 - P_d$  and probability of false alarm  $P_{fa}$  of standalone SVM and GP+SVM cases across  $\gamma$  with fixed  $\nu$ 's. Detector with GP shows lower loss whereas there is no significant difference in  $P_{fa}$ .

$P_{fa}$  is close to zero, in both cases  $P_{fa}$  is less than 0.2%. However, the probability of loss of the detector with GP predictor is five times smaller than the standalone OC-SVM machine.

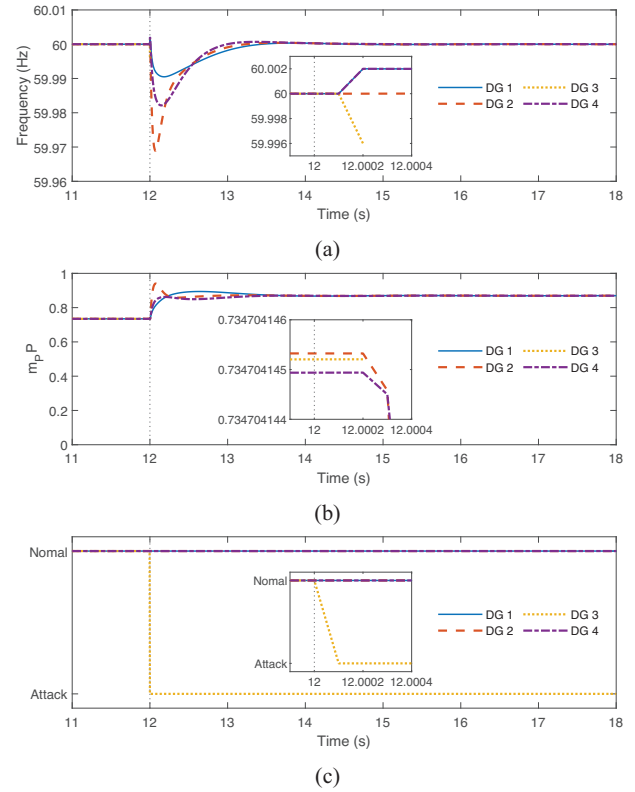
## 5 | CASE STUDIES

The effectiveness of the proposed FDI attack detection algorithm trained and tested from the data is validated on running simulations. In specific, the attack detection and false alarm performances are investigated. The validity of the attack mitigation scheme that uses the decision of detection algorithm is also examined. The attack detector is applied on the 4 DG microgrid test system in Figure 5 and Table 1 and is simulated in MATLAB/Simulink and HIL testbed for Case A and Case B, respectively.

### 5.1 | Case A: Simulation verification

#### 5.1.1 | Case A.1: Performance against FDI attack

In this test case, the FDI attack in Figure 3 is replayed with the detection scheme activated. The system is initially at stable operation, i.e. the frequencies are synchronized at 60 Hz and the active power sharing is maintained as shown in Figure 8. From  $t = 12$  s, a false value of 62 Hz is injected into DG 3 which is arbitrarily chosen as the point of attack. As shown in the figure, the frequency and the active power ratio consensus are broken



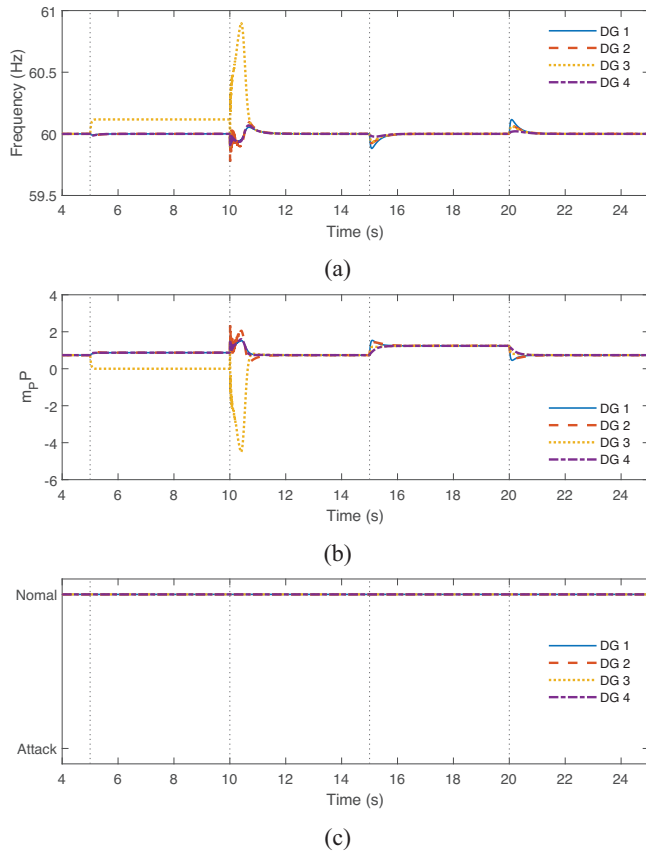
**FIGURE 8** A false data of 62 Hz is injected to DG 3 frequency droop in Case A.1: (a) DG frequencies; (b) active power ratios  $m_H P_i$ ; (c) attack detection.

after the attack is applied, and the system enters the transient state. The detector then locates the attack at  $t = 12.0001$  s and DG 3 is disconnected from both the power and communication network at  $t = 12.0002$  s. By isolating DG 3, which became the point of failure, the rest of the system gains back the steady state with the secondary control objectives satisfied. Here, note that the attack is detected while the frequency deviation from the attack is negligibly small as shown in Figure 8a. One might wonder how this is possible as this much deviation can be easily observed during normal disturbances. This can be partially contributed by taking a window of information and input to the detector, i.e. using the history of data, but the use of the estimation error instead of the direct use of angular frequency adds more contributions to this. As seen in Figure 8b, the active power ratios do not immediately respond to the attack as the frequencies do; the active power ratios remain synchronized until the isolation is triggered at  $t = 12.0002$  s, as opposed to the frequency consensus is broken right after the attack is applied at  $t = 12.0001$  s. By virtue of the delay in response to the attack, a larger estimation error of angular frequency can be generated and becomes the indication of the cyberattack on the system.

#### 5.1.2 | Case A.2: Response to normal disturbances

In general, when there is a fault in the power system, a manual fault-clearing process is required. Similarly, with the FDI attack



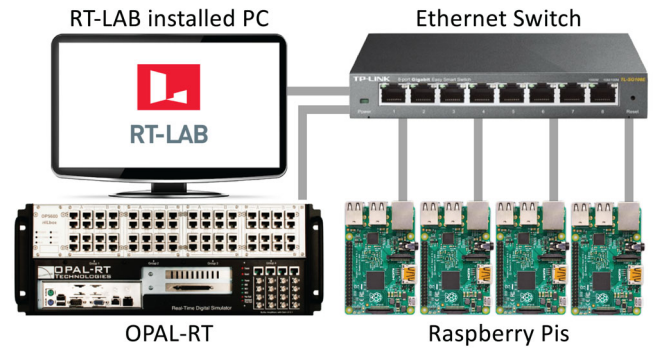


**FIGURE 9** The proposed detector is operated under normal disturbances in order to validate low false alarm rate *Case A.3*: (a) DG frequencies; (b) active power ratios  $m_P P_i$ ; (c) attack detection.

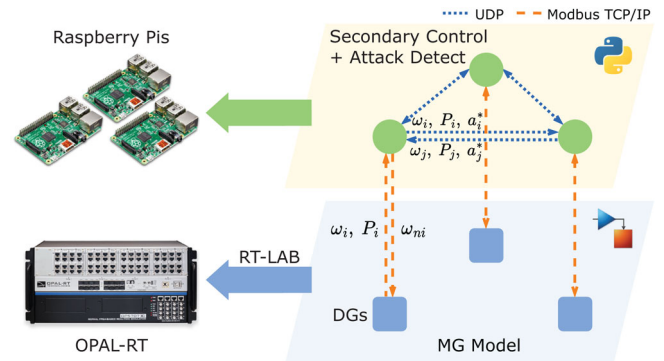
discussed in this paper, the isolation of the point of attack is required to interrupt the propagation of the attack. This means that once the attack is located, there is a loss in energy resources in the system. Therefore, it is also important that the detector does not falsely declare a normal transient as an attack to avoid degradation in system reliability. In order to validate the anti-false alarm performance, the detector is operated under normal disturbances. DG 3 is disconnected from and reconnected to the microgrid at  $t = 5$  s and  $t = 10$  s, respectively. An additional amount of  $P = 20$  kW is applied to and removed from the system at  $t = 15$  s and  $t = 20$  s, respectively. From Figure 9, it is seen that during the normal transients, the detector is not providing a false alarm.

## 5.2 | Case B: HIL verification

In Case B, the attack detection algorithm is validated in the HIL simulation testbed illustrated in Figure 10. The HIL testbed is composed of a digital real-time simulator OPAL-RT, RT-LAB installed PC and four Raspberry Pis. The communications between the hardware are established on Ethernet. The schematic of the HIL testbed is depicted in Figure 11. The detection algorithm for each DG is implemented in each Raspberry Pi in Python script. The communications between the



**FIGURE 10** Hardware setup required for the HIL testbed.

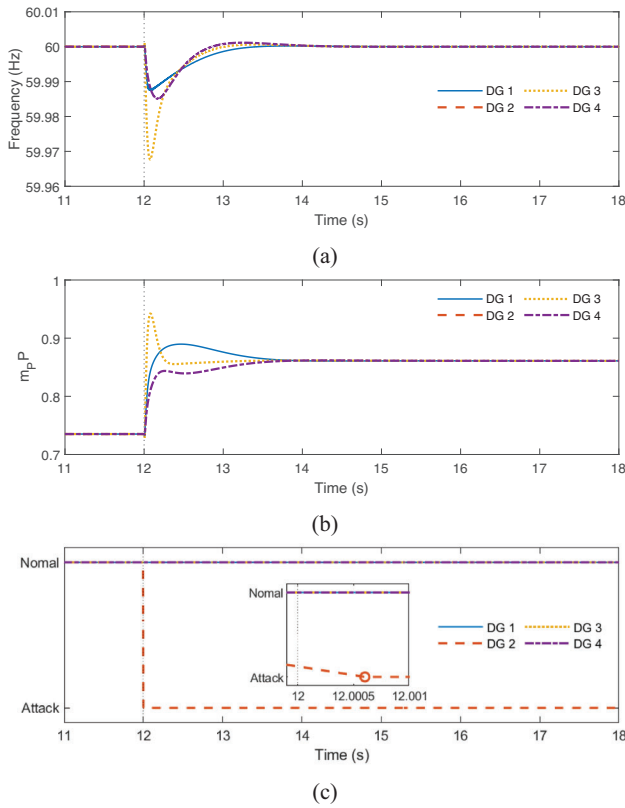


**FIGURE 11** The schematic of HIL testbed for FDI attack detector introduced. The detection algorithm is implemented in Python script and operated in Raspberry Pis.

controllers are established via UDP protocol that is represented in the blue dotted lines; each controller broadcasts to the sub-net to a designated port, then the neighbour of that controller listens to that port. The microgrid model in Figure 5 is simulated in real-time on OPAL-RT. The communication between the DGs (OPAL-RT) and the controller (Raspberry Pi) is established in Modbus TCP/IP, which is represented by the orange dashed lines.

### 5.2.1 | Case B.1. Performance against FDI attacks

In this case study, a similar attack presented in Case A.1 is applied and the detection algorithm is operated on a Raspberry Pi. A corrupted frequency of 61 Hz is continuously injected into the arbitrarily chosen point of attack DG 2 starting from  $t = 12$  s, and the detector locates the attack at  $t = 12.0006$  s as shown in Figure 12. Once the attack on DG 2 is detected, DG 3 is disconnected from the power network as well as from the communication network. As shown in Figure 12(a,b), the rest of the DGs in the microgrid maintain their stable operation, i.e. frequency synchronization at 60 Hz and active power ratio synchronization.



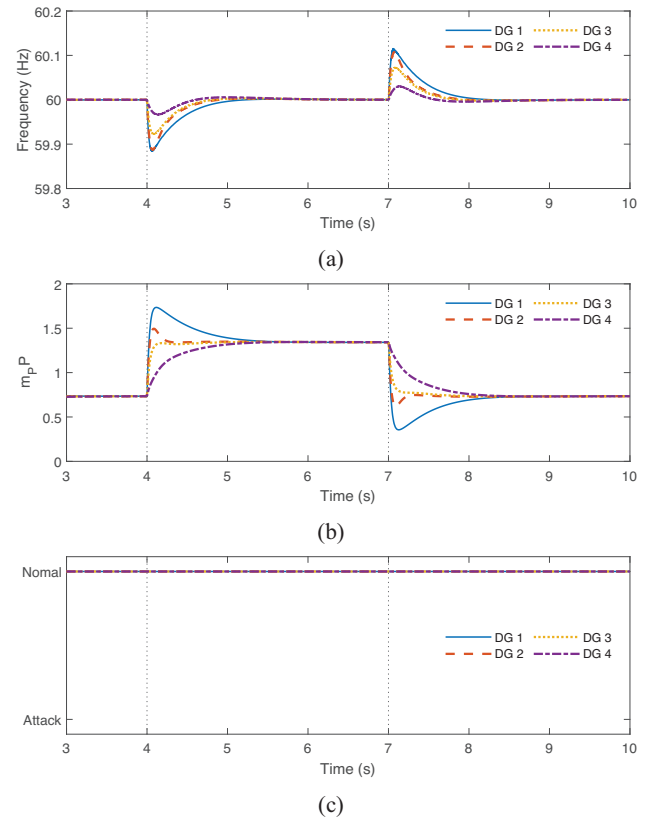
**FIGURE 12** DG 2 frequency is manipulated in *Case B.1*: (a) DG frequencies; (b) active power ratios  $m_P P$ ; (c) attack detection.

### 5.2.2 | Case B.2. Response to normal disturbances

The response of the proposed detector to the load changes which is a normal disturbance is validated in HIL testbed in this case study. As shown in Figure 13, the secondary control is activated at  $t = 1$  s, and 25 kW of load is added to and removed from the system at  $t = 4$  s and  $t = 7$  s, respectively. As expected, the detector does not locate false alarms.

## 6 | CONCLUSION

This work presented a FDI attack detection structure for a distributed microgrid control system based on the GP predictor and OC-SVM based anomaly detection algorithm. In this unsupervised approach, the GP regression is used to estimate the angular frequency of a DG from the active power ratio and its time derivative. The assumption is that the prediction error is low when there is no attack on the system while it is high when an attack is applied. In order to distinguish between a disturbance and an attack, in which they both have high errors but high and low variances, respectively, the predicted error is standardized with the predicted standard deviation, so that the standardized error is high only for the attack. The attack detector proposed was evaluated in terms of the probabilities of detection loss and false alarm. The OC-SVM algorithm using



**FIGURE 13** The proposed detector is operated under normal disturbances in order to validate low false alarm rate *Case B.2*: (a) DG frequencies; (b) active power ratios  $m_P P$ ; (c) attack detection.

the GP predictor returned a 5 times smaller false alarm probability than the standalone one, while no significant difference is observed for detection loss. The detector trained was validated in a running simulation of the 4 DG microgrid test model in both the simulation and HIL testbeds. The FDI attack on the frequency droop with distributed control system was considered and the effectiveness of the proposed attack detection and mitigation scheme was highlighted. The results show that the proposed scheme can effectively detect FDI attacks while avoiding false alarm locations.

### AUTHOR CONTRIBUTIONS

Jeewon Choi: Conceptualization, software, validation, visualization, writing - original draft, writing - review and editing. Behshad Roshanzadeh: Software. Manel Martínez-Ramón: Methodology, validation, writing - original draft, writing - review and editing. Ali Bidram: Conceptualization, methodology, supervision, writing - original draft, writing - review and editing.

### ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Awards OIA-1757207 and ECCS-2214441. Manel Martínez-Ramón is partially supported by the King Felipe VI Endowed Chair of the University of New Mexico.

## CONFLICT OF INTEREST


The authors declare no conflicts of interest.

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## ORCID

*Jeewon Choi*  <https://orcid.org/0000-0002-4611-0403>

*Behshad Roshanzadeh*  <https://orcid.org/0000-0003-2127-6076>

*Manel Martínez-Ramón*  <https://orcid.org/0000-0001-6912-9951>

*Ali Bidram*  <https://orcid.org/0000-0003-4722-4346>

## REFERENCES

- Zhang, L., Li, L., Wihl, L., Kazemtabrizi, M., Ali, S.Q., Paquin, J.N., et al.: Cybersecurity Study of Power System Utilizing Advanced CPS Simulation Tools. In: Proceedings of the 2019 PAC World Americas Conference, pp. 19–22. Keysight, Santa Rosa, CA (2019)
- Yohanandhan, R.V., Elavarasan, R.M., Manoharan, P., Mihet-Popa, L.: Cyber-physical power system (CPPS): a review on modeling, simulation, and analysis with cyber security applications. *IEEE Access* 8, 151019–151064 (2020)
- Bidram, A., Damodaran, L., Fierro, R.: Cybersecure distributed voltage control of AC microgrids. In: 2019 IEEE/IAS 55th Industrial and Commercial Power Systems Technical Conference (I&CPS), pp. 1–6. IEEE, Piscataway, NJ (2019)
- Poudel, B.P., Mustafa, A., Bidram, A., Modares, H.: Detection and mitigation of cyber-threats in the DC microgrid distributed control system. *Int. J. Electr. Power Energy Syst.* 120, 105968 (2020)
- Mustafa, A., Poudel, B., Bidram, A., Modares, H.: Detection and mitigation of data manipulation attacks in AC microgrids. *IEEE Trans. Smart Grid* 11(3), 2588–2603 (2019)
- Forum, W.E.: The global risks report 2019 14th edition. World Economic Forum, Geneva (2019)
- Bidram, A., Davoudi, A.: Hierarchical structure of microgrids control system. *IEEE Trans Smart Grid* 3(4), 1963–1976 (2012)
- Dehkordi, N.M., Sadati, N., Hamzeh, M.: Distributed robust finite-time secondary voltage and frequency control of islanded microgrids. *IEEE Trans. Power Syst.* 32(5), 3648–3659 (2017)
- Guerrero, J.M., Vasquez, J.C., Matas, J., De Vicuña, L.G., Castilla, M.: Hierarchical control of droop-controlled AC and DC microgrids—A general approach toward standardization. *IEEE Trans. Ind. Electron.* 58(1), 158–172 (2011)
- Bidram, A., Davoudi, A., Lewis, F.L., Guerrero, J.M.: Distributed cooperative secondary control of microgrids using feedback linearization. *IEEE Trans. Power Syst.* 28(3), 3462–3470 (2013)
- Bidram, A., Davoudi, A., Lewis, F.L., Qu, Z.: Secondary control of microgrids based on distributed cooperative control of multi-agent systems. *IET Gener. Transm. Distrib.* 7(8), 822–831 (2013)
- Bidram, A., Davoudi, A., Lewis, F.L.: A multiobjective distributed control framework for islanded AC microgrids. *IEEE Trans. Ind. Informat.* 10(3), 1785–1798 (2014)
- Huang, Y., Tang, J., Cheng, Y., Li, H., Campbell, K.A., Han, Z.: Real-time detection of false data injection in smart grid networks: an adaptive CUSUM method and analysis. *IEEE Syst. J.* 10(2), 532–543 (2014)
- Manandhar, K., Cao, X., Hu, F., Liu, Y.: Detection of faults and attacks including false data injection attack in smart grid using Kalman filter. *IEEE Trans. Control Network Syst.* 1(4), 370–379 (2014)
- Bi, S., Zhang, Y.J.: Graphical methods for defense against false-data injection attacks on power system state estimation. *IEEE Trans. Smart Grid* 5(3), 1216–1227 (2014)
- Mo, Y., Chabukwar, R., Sinopoli, B.: Detecting integrity attacks on SCADA systems. *IEEE Trans. Control Syst. Technol.* 22(4), 1396–1407 (2013)
- Liu, L., Esmalifalak, M., Ding, Q., Emesih, V.A., Han, Z.: Detecting false data injection attacks on power grid by sparse optimization. *IEEE Trans. Smart Grid* 5(2), 612–21 (2014)
- Rawat, D.B., Bajracharya, C.: Detection of false data injection attacks in smart grid communication systems. *IEEE Signal Process. Lett.* 22(10), 1652–1656 (2015)
- Pasqualetti, F., Dörfler, F., Bullo, F.: Attack detection and identification in cyber-physical systems. *IEEE Trans. Autom. Control* 58(11), 2715–2729 (2013)
- Beg, O.A., Johnson, T.T., Davoudi, A.: Detection of false-data injection attacks in cyber-physical DC microgrids. *IEEE Trans. Ind. Inf.* 13(5), 2693–2703 (2017)
- Ghafoori, M.S., Soltani, J.: Designing a robust cyber-attack detection and identification algorithm for DC microgrids based on Kalman filter with unknown input observer. *IET Gener. Transm. Distrib.* 16(16), 3230–3244 (2022)
- Sengan, S., V, S., V, I., Velayutham, P., Ravi, L.: Detection of false data cyber-attacks for the assessment of security in smart grid using deep learning. *Comput. Electr. Eng.* 93, 107211 (2021)
- Ferragut, E.M., Laska, J., Olama, M.M., Ozmen, O.: Real-time cyber-physical false data attack detection in smart grids using neural networks. In: 2017 Intl. Conf. on Comp. Sci. and Comp. Intelligence (CSCI), pp. 1–6. IEEE, Piscataway, NJ (2017)
- Qu, Z., Bo, X., Yu, T., Liu, Y., Dong, Y., Kan, Z., et al.: Active and passive hybrid detection method for power CPS false data injection attacks with improved AKF and GRU-CNN. *IET Renewable Power Gener.* 16(7), 1490–1508 (2022)
- Ramaswamy, S., Rastogi, R., Shim, K.: Efficient algorithms for mining outliers from large data sets. In: Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, pp. 427–438. ACM, New York (2000)
- Angiulli, F., Pizzuti, C.: Fast outlier detection in high dimensional spaces. In: European conference on principles of data mining and knowledge discovery, pp. 15–27. Springer, Piscataway, NJ (2002)
- Schölkopf, B., Williamson, R.C., Smola, A., Shawe-Taylor, J., Platt, J.: Support vector method for novelty detection. In: NIPS'99: Proceedings of the 12th International Conference on Neural Information Processing Systems, pp. 582–588. ACM, New York (1999)
- Schölkopf, B., Platt, J.C., Shawe-Taylor, J., Smola, A.J., Williamson, R.C.: Estimating the support of a high-dimensional distribution. *Neural Comput.* 13(7), 1443–1471 (2001)
- Tax, D.M., Duin, R.P.: Support vector data description. *Mach. Learn.* 54(1), 45–66 (2004)
- Van Every, P.M., Rodriguez, M., Jones, C.B., Mammoli, A.A., Martínez-Ramón, M.: Advanced detection of HVAC faults using unsupervised SVM novelty detection and Gaussian process models. *Energy Build.* 149, 216–224 (2017)
- Burges, C.J.: A tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Discov.* 2(2), 121–167 (1998)
- Aizerman, M.A., Braverman, E.M., Rozonoer, L.I.: Theoretical foundations of potential function method in pattern recognition. *Autom. Remote Control* 25(6), 917–936 (1964)
- Martínez-Ramón, M., Gupta, A., Rojo-Álvarez, J.L., Christodoulou, C.G.: Machine Learning Applications in Electromagnetics and Antenna Array Processing. Artech House, Norwood, MA (2021)

**How to cite this article:** Choi, J., Roshanzadeh, B., Martínez-Ramón, M., Bidram, A.: An unsupervised cyberattack detection scheme for AC microgrids using Gaussian process regression and one-class support vector machine anomaly detection. *IET Renew. Power Gener.* 17, 2113–2123 (2023).  
<https://doi.org/10.1049/rpg2.12753>