A Novel False Measurement Data Detection Mechanism for Smart Grids

Muhammad Awais Shahid 1, Fiaz Ahmad 1,*, Rehan Nawaz 1, Saad Ullah Khan 1, Abdul Wadood 2,*, and Hani Albalawi 3,4,*

1 Department of Electrical & Computer Engineering, Air University, Islamabad 44230, Pakistan; rehan._.nawaz@outlook.com (R.N.)
2 Department of Electrical Engineering, Air University Islamabad, Kamra Campus, Kamra 43570, Pakistan
3 Department of Electrical Engineering, Faculty of Engineering, University of Tabuk, Tabuk 71491, Saudi Arabia
4 Renewable Energy and Energy Efficiency Center (REEEC), University of Tabuk, Tabuk 71491, Saudi Arabia
* Correspondence: fiaz.ahmad@mail.au.edu.pk (F.A.); wadood@au.edu.pk (A.W.); halbala@ut.edu.sa (H.A.)

Abstract: With the growing cyber-infrastructure of smart grids, the threat of cyber-attacks has intensified, posing an increased risk of compromised communication links. Of particular concern is the false data injection (FDI) attack, which has emerged as a highly dangerous cyber-attack targeting smart grids. This paper addresses the limitations of the variable dummy value model proposed in the authors previous work and presents a novel defense methodology called the nonlinear function-based variable dummy value model for the AC power flow network. The proposed model is evaluated using the IEEE 14-bus test system, demonstrating its effectiveness in detecting FDI attacks. It has been shown that previous detection techniques are unable to detect FDI attacks, whereas the proposed method is shown to be successful in the detection of such attacks, guaranteeing the security of the smart grid’s measurement infrastructure.

Keywords: cyber-physical systems; power system state estimation; false data injection attacks; false data detection; cyber security

1. Introduction

A smart grid consists of smart infrastructure system, smart management system, and smart protection system. The two-way flow of electricity, as well as information, is used in a smart grid, and an automated as well as distributed energy delivery network is created in this way [1–4]. The smart grid also has self-healing features [5].

Smart meters are present for the measurement of power flow values. After measurement, those values are transmitted to utility companies as well as to the customers. These data are also sent through Advanced Metering Infrastructure (AMI) for the purpose of billing and taking certain control actions [6,7]. In a smart grid, the smart meters help enhance the different operational functionalities, such as generation scheduling, load management, power quality enhancement, fault prediction, load forecasting, etc. [8]. The control mechanism of a smart grid is far better than that of a traditional grid [9,10].

If a mismatch has occurred between the generation and the consumption of power in a power system, the electrical quantities will deviate from their values. For this purpose, in a smart grid, the two-way flow of power is achieved. For a reliable power flow, the communication should be secure. The communication links may be hacked by attackers in an attempt to change the power flow measurements [11]. The attacks can be carried out for multiple purposes. Attacks can be carried out to obtain financial benefits and creating technical problems [12–15]. On the basis of the target of the attacks, there are three types of attacks: attacks to target availability, attacks to target integrity, and attacks to target confidentiality [16,17].
To detect bad data in the system, the technique of power system state estimation (PSSE) is employed. Although the PSSE technique performs well in detecting bad data, it fails to detect a special kind of attack that is called a stealth attack. The attacker has full awareness of the Jacobian matrix $H$ of power system topology in this particular attack, and the attack can be constructed by using $H$. Stealth FDI is written as follows [18–29]:

$$z_a = z + a$$

(1)

The measurements received in the control room are put in a vector represented by $z$. In the above equation, vector $a$ consists of the false data, which are to be added to the measurement vector. The intruder hacks the communication line, and the measurement data is obtained. The attacker obtains the Jacobian matrix $H$ with the help of those data. The attack vector $a$ is injected into the measurement vector as $a = Hc$. The vector $c$ is taken as $c \sim N(0, \sigma^2 c)$, where $\sigma^2 c$ denotes the false state variance. The attacked measurement vector is represented by $z_a$.

The variable dummy value model (VDVM) proposed in [26] to detect FDI attacks becomes vulnerable to attacks when the attacker uses multivariate linear regression in order to find the function using which the dummy values were calculated. In that case, an attacker can make such attacks, which bypass the VDVM method. So, it is important to tackle these drawbacks of the VDVM technique.

Contributions

The paper has the following major contributions:

1. A nonlinear function-based the variable dummy value model (NF-VDVM) is proposed for the task of false data injection attack detection, and the proposed model has the ability to handle the limitations of the VDVM technique.
2. It has been validated through the results of the simulations that the NF-VDVM technique is able to successfully detect FDI attacks that were not detected by the bad data filter as well as the VDVM method, and the smart grid is protected against the FDI attacks.

The remaining paper is arranged as follows: Section 2 contains the related work present in the literature. Section 3 consists of the proposed model used to detect FDI attacks. The evaluation of the model is detailed in Section 4. The conclusion is presented in Section 5.

2. Related Work

In the literature, stealth false data injection (FDI) attacks are detected by using different types of algorithms and methods. The methods based on the deep learning [30] have been broadly used for this purpose. In [31], a single deep neural network model was designed to simultaneously perform the distribution system state estimation calculation and FDI attack detection using the real measurements. The FDI attacks were detected by proposing a hybrid deep learning model in [32]. The model used the convolutional neural networks-long short-term memory (CNN-LSTM) with particle swarm optimization (PSO). The measurements of the phasor measurement unit (PMU) were used by the model for the detection of abnormal measurement values. The PSO performed the task of optimizing the complex hyperparameter space of the CNN-LSTM. A graph convolutional network (GCN) framework was proposed in [33] to detect FDI attacks. In [34], an adversarial-resilient deep neural network detection framework was designed for the purpose of detecting the FDI attacks. In [35], FDI attacks were detected by proposing another method that was based on deep learning. The cyber-attacks in smart grids were classified by using the deep neural network model in [36].

In [37], a defense strategy was developed in order to improve the generalization ability as well as the detection performance in case of unseen FDI attacks. A graph autoencoder-based scheme was proposed for their detection that captured the spatio-temporal features.
from the power systems, and the detection performance was improved in this way. Moreover, in order to improve the generalization ability, training was performed on comprehensive graphs reflecting many realizations for power system topologies. A method for the localization detection of FDI attacks was proposed in [38]. The proposed method was based on graph data modeling and graph deep learning. In [39], a spatiotemporal deep network known as PowerFDNet was proposed to detect stealth FDI attacks. Two sub-architectures i.e., spatial architecture and temporal architecture formed the PowerFDNet. In [40], another deep learning-based method was adopted in order to detect attacks. The formulation of FDI attack detection was based on a machine learning based binary classification problem. A method based on deep neural networks was proposed for solving this binary classification problem. A spatio-temporal learning algorithm was developed in [41] to acquire the normal dynamics for the distribution systems in order to overcome FDI attacks. The usual dynamics were acquired by using a long short-term memory autoencoder. Then, the unsupervised trained model was employed to detect FDI attacks by using the residual error for each measurement sample. A spatiotemporal graph deep learning-based method was implemented in [42] in order to detect FDI attacks, and the attack samples were not required for the proposed scheme.

To detect FDI attacks, a method based on the gated graph neural network as well as the attention mechanism was employed in [43]. The spatial features of the grid operation data as well as the topology data were extracted by using the graph neural network, and the representation for every node was calculated. Moreover, the attention mechanism was used for assigning the aggregation weight for the neighbor nodes. In [44], for the detection of FDI attacks, a method was adopted that used the combination of blockchain, Fog Computing, as well as a linear support vector machine with Principal Component Analysis. In [45], FDI attacks were detected by using recurrent neural networks. Another technique based on the Gaussian mixture model was adopted in [46] for the detection of attacks.

A comprehensive analysis was provided in [47] for different machine learning algorithms for the detection of FDI attacks by two representative datasets, which were power system and water treatment. In order to detect FDI attacks, four methods based on machine learning were used in [48]. In [49], many machine learning-based methods were presented that have been employed in the literature in order to detect false data injection attacks. In [50], the classification of measurements as attacked or secure was achieved by using machine learning algorithms. For that purpose, the method of sparse logistic regression was used. Moreover, the methods of support vector machine (SVM), and k-nearest neighbor were also employed. The work of [51] consisted of a study on the classification of the data for the normal operation of a system as well as the data for FDI attacks. The detection model was based on the time series, and normal operating parameters as well as abnormal operation parameters in case of attack scenarios were collected as samples to train the detection model. A random forest algorithm was adopted for detection purposes.

Stealth FDI attacks were also detected in [52] by proposing two sensor encoding methods. FDI attacks towards the synchrophasor measurements were detected in [53] by using an online data-driven method. There was use of density-based local outlier factor analysis by the proposed algorithm for the purpose of detecting anomalies in data. In [54], another technique was employed, in which the modelling of the system was performed as a discrete-time linear dynamic system. The Kalman filter was used to perform the state estimation. The attacks were quickly detected by the generalized cumulative sum algorithm.

The research work presented in this paper is an improved form of our previously published work [26], in which the VDVM technique was proposed for the detection of FDI attacks. In the previous research work, the comparison of our proposed work was made with the existing techniques used in that area. So, in order to avoid repetition, the comparison of the proposed technique of NF-VDVM has not been made with those techniques. However, the NF-VDVM method has been compared to the previously published technique.
of VDVM in order to show the necessities and advantages of the proposed NF-VDVM scheme, and it has been validated that the NF-VDVM technique is capable of handling the limitations of the VDVM technique.

3. Proposed Model

In this paper, the proposed model is an extension of the work given in [26]. A novel model for power system has been introduced in [26] for an AC power flow network, in which stealth false data injection attacks are efficiently detected in the control room. The concept of a dummy measurement value is the basis of this model. The smart meters send the actual measurement values as well as the dummy values. For the task of efficient detection of FDI attacks, two techniques have been proposed, which are the fixed dummy value model (FDVM) and variable dummy value model (VDVM). The VDVM has detected even those attacks that had not been detected by the FDVM. However, the VDVM has a limitation while detecting stealth FDI attacks. When the intruder uses multivariate linear regression (MLR) for finding the coefficients used in the function implemented for the calculation of dummy measurement values, the intruder can make an attack that bypasses the VDVM technique. For this purpose, to tackle this limitation, a nonlinear function-based variable dummy value model (NF-VDVM) is proposed in this paper, which is robust against stealth FDI attacks.

In this work, the potential flaws of the communication networks in the smart grid supervisory control and data acquisition (SCADA) systems, like the basic bugs or the communication collapses, are not taken into account. Moreover, it is assumed that noise, interference, or estimation errors do not exist for data collection or detection processing. The main emphases of this research work is on the false data injection attacks. In such attacks, an intruder hacks the measurement data from the communication line and injects the malicious data into the measurement data before it reaches the control room.

In the case of an AC power flow network, where total buses is represented as \( b \) and the total transmission lines as \( l \), the measurement vector is written as follows:

\[
\mathbf{z}_t = \begin{bmatrix} z_{p(t)} & z_{q(t)} & z_{p(l)} & z_{q(l)} & z_{lq(l)} \end{bmatrix}^T \tag{2}
\]

Here \( \mathbf{z}_t \) denotes the measurement vector at \( t^{th} \) instant where \( t = 1, 2, 3, \ldots, tt \). The total number of instances are represented by \( tt \). The vectors \( z_{p(t)} \) and \( z_{q(t)} \) consist of the active powers and the reactive powers, which are injected into all the buses at the \( t^{th} \) instant, respectively. The vectors \( z_{p(l)} \) and \( z_{q(l)} \) contain the active as well as the reactive forward powers for the transmission lines at the \( t^{th} \) instant. Similarly, \( z_{lq(l)} \) consist of the active as well as the reactive backward powers for the transmission lines at the \( t^{th} \) instant. The dimensions of the measurement vector \( \mathbf{z}_t \) are \( m \times 1 \), where \( m \) represents the number of entries in \( \mathbf{z}_t \).

In the case of dummy measurement values, the measurement vector will become like this:

\[
\mathbf{z}_{dt} = \begin{bmatrix} z_{p(t)}(1); z'_{q(t)}(1); \ldots; z_{q(t)}(b); z'_{q(t)}(b); z_{p(l)}(1); z'_{p(l)}(1); \ldots; z_{q(l)}(l); z'_{q(l)}(l); z_{p(l)}(1); z'_{p(l)}(1); \ldots; z_{q(l)}(l); z'_{q(l)}(l) \end{bmatrix} \tag{3}
\]

Here \( \mathbf{z}_{dt} \) represents the measurement vector, which contains the actual as well as the dummy values. \( z_{p(t)}(1) \) denotes the first entry of vector \( z_{p(t)} \), and \( z_{q(t)}(b) \) shows the \( b^{th} \) entry of \( z_{q(t)} \). In the new measurement vector, even indexes contain the dummy measurement values of power. Dummy value vectors consisting of active as well as the reactive powers that are injected to buses at the \( t^{th} \) instant are \( z'_{p(t)} \) and \( z'_{q(t)} \), respectively. Similarly, the vectors \( z'_{q(l)} \) and \( z'_{q(l)} \) contain the dummy values of active as well as reactive forward and backward transmission line powers at the \( t^{th} \) instant. The dimensions of \( \mathbf{z}_{dt} \) will be \( 2 \times m \times 1 \).

In the proposed nonlinear function-based variable dummy value model (NF-VDVM), a dummy measurement value along with the actual measurement value is transmitted to
the control room, and that dummy value is calculated by using a nonlinear function. A fifth-order polynomial function is used for the purpose of calculating the dummy measurement values, and the actual measurement values are used in that function. So the dummy values have a dependency on the actual values and will be changed with the change in actual measurement values. Only the control room knows about the function, and its nonlinearity makes it really difficult for the attacker to successfully predict the function and to make an attack. This nonlinear function is inserted into the meters to calculate the dummy measurement values. The functions through which the dummy values are calculated for buses in the case of NF-VDVM are given below:

\[
z'_{p(i)}(i) = \alpha_{1pi}(z_{ppi(i)})^5 + \alpha_{2pi}(z_{ppi(i)})^4 + \alpha_{3pi}(z_{ppi(i)})^3 + \alpha_{4pi}(z_{ppi(i)})^2 + \alpha_{5pi}(z_{ppi(i)}) + \alpha_{6pi}
\]  

(4)

\[
z'_{q(i)}(i) = \alpha_{1qi}(z_{qqi(i)})^5 + \alpha_{2qi}(z_{qqi(i)})^4 + \alpha_{3qi}(z_{qqi(i)})^3 + \alpha_{4qi}(z_{qqi(i)})^2 + \alpha_{5qi}(z_{qqi(i)}) + \alpha_{6qi}
\]  

(5)

In Equations (4) and (5), \( i = 1, 2, 3, \ldots, b \). \( z'_{p(i)}(i) \) and \( z'_{q(i)}(i) \) denote the \( i^{th} \) entries for the vectors \( z'_{p(i)} \) and \( z'_{q(i)}' \), respectively, containing dummy measurement values for the active as well as the reactive powers, which are injected to buses at the \( i^{th} \) instant. Similarly, \( z_{p(i)}(i) \) and \( z_{q(i)}(i) \) denote the active and the reactive powers, respectively, for the first transmission line that is attached to the \( i^{th} \) bus at the \( i^{th} \) instant. \( \alpha_{1pi}, \alpha_{2pi}, \alpha_{3pi}, \alpha_{1qi}, \alpha_{2qi}, \) and \( \alpha_{3qi} \) represent the constants used in the equations. In Equations (4) and (5), the highest order of the expressions is set to 5. The purpose behind the selection of the highest order of the equations is that the highly nonlinear functions are selected for the calculation of the dummy values, so that the attacker is not able to predict the function. At the same time, the memory requirement is also kept in mind. A function having the highest degree of 5 contains the six constants for the calculation of a single dummy value of power. The higher the order of the polynomial function is, the more constants are utilized, and the more memory is required to implement that function. So, the highest degree of the functions used to calculate the dummy values of power is taken as 5.

The functions used for calculating the dummy values of forward as well as the backward active and reactive powers for transmission lines are given by:

\[
z'_{pl(i)}(i) = \alpha_{1pli}(z_{ppli(i)})^5 + \alpha_{2pli}(z_{ppli(i)})^4 + \alpha_{3pli}(z_{ppli(i)})^3 + \alpha_{4pli}(z_{ppli(i)})^2 + \alpha_{5pli}(z_{ppli(i)}) + \alpha_{6pli}
\]  

(6)

\[
z'_{ql(i)}(i) = \alpha_{1qli}(z_{qqli(i)})^5 + \alpha_{2qli}(z_{qqli(i)})^4 + \alpha_{3qli}(z_{qqli(i)})^3 + \alpha_{4qli}(z_{qqli(i)})^2 + \alpha_{5qli}(z_{qqli(i)}) + \alpha_{6qli}
\]  

(7)

\[
z'_{lp(i)}(i) = \alpha_{1lp(i)}(z_{prli(i)})^5 + \alpha_{2lp(i)}(z_{prli(i)})^4 + \alpha_{3lp(i)}(z_{prli(i)})^3 + \alpha_{4lp(i)}(z_{prli(i)})^2 + \alpha_{5lp(i)}(z_{prli(i)}) + \alpha_{6lp(i)}
\]  

(8)

\[
z'_{lq(i)}(i) = \alpha_{1lqi}(z_{qrli(i)})^5 + \alpha_{2lqi}(z_{qrli(i)})^4 + \alpha_{3lqi}(z_{qrli(i)})^3 + \alpha_{4lqi}(z_{qrli(i)})^2 + \alpha_{5lqi}(z_{qrli(i)}) + \alpha_{6lqi}
\]  

(9)

In Equations (6)–(9), \( i = 1, 2, 3, \ldots, l \). Here, \( z'_{pl(i)}(i) \) and \( z'_{ql(i)}(i) \) represent \( i^{th} \) entries for dummy measurement vectors \( z'_{pl(i)} \) and \( z'_{ql(i)} \), respectively, containing forward active as well as the reactive powers passing through the transmission lines at the \( i^{th} \) instant. Similarly, \( z'_{lp(i)}(i) \) and \( z'_{lq(i)}(i) \) denote the \( i^{th} \) entries for the dummy measurement vectors \( z'_{lp(i)} \) and \( z'_{lq(i)} \), respectively, which contain backward active as well as the reactive powers passing through the transmission lines at the \( i^{th} \) instant. To calculate the dummy measurement value of power flowing through a particular transmission line, \( z_{ppli(i)} \) and \( z_{prli(i)} \)
denote the active powers, and \(z_{qri(t)}\) and \(z_{qri(t)}\) denote the reactive powers injected into the bus from which that particular transmission line is coming at the \(t^{th}\) instant. These equations also use some constants, which are denoted by \(\alpha\)’s.

To learn the coefficients present in the equations used to calculate dummy values of power, a machine learning technique, known as polynomial regression, is used. It is a technique that models the relationship between a dependent and independent variable as a polynomial having a degree greater than 1. Polynomial regression is used here to find the best values of constants. The process of polynomial regression for finding the constants used in Equation (3) is explained here. The hypothesis can be written in this form:

\[
\alpha_k(z_{plk}) = \alpha_{1pk}(z_{pk})^5 + \alpha_{2pk}(z_{pk})^4 + \alpha_{3pk}(z_{pk})^3 + \alpha_{4pk}(z_{pk})^2 + \alpha_{5pk}(z_{pk}) + \alpha_{6pk} \quad (10)
\]

In the above hypothesis, \(\alpha_k(z_{plk})\) is the function of \(z_{plk}\) parameterized by \(\alpha_k\). \(z_{plk}\) denotes the \(k^{th}\) input vector, and \(k = 1, 2, 3, \ldots, b\). Similarly, \(\alpha_k\) represents the \(k^{th}\) parameter vector. \(\alpha_{1pk}, \alpha_{2pk}, \alpha_{3pk}, \alpha_{4pk}, \alpha_{5pk}, \) and \(\alpha_{6pk}\) denote the constants to learn. Depending on this hypothesis, a cost function is defined, and a method is adopted to minimize the cost function as discussed in detail for multivariate linear regression in [26]. The best values of constants are determined in this way, and the same process is used to find the constants of the remaining equations. Then these values are placed in their functions in order to calculate the dummy measurement values of active as well as reactive power. The nonlinear functions are then placed into meters for the purpose of calculating dummy measurement values of power. The actual values of power are measured by the meters, and then those embedded functions are used for calculating the dummy values of power. The actual measured values of power as well as the calculated dummy values are sent to the control room.

For the aim detecting FDI attacks in the control room, the recalculation of dummy measurement values is achieved by using those functions. In this case, the actual values of power that are obtained in the control room from the measurement vector are used in those functions for the recalculation of dummy values. For attack detection, a comparison is made between the recalculated dummy measurement value and the received dummy measurement value from the measurement vector. For comparing the recalculated dummy values with the received dummy values in the control room, the equations used for active as well as the reactive powers injected into buses are following:

\[
r_{z_{pl}(t)}(j) = z'_{pr(i)}(j) - \left(\alpha_{1pj}(z_{qri(t)})^5 + \alpha_{2pj}(z_{qri(t)})^4 + \alpha_{3pj}(z_{qri(t)})^3 + \alpha_{4pj}(z_{qri(t)})^2 + \alpha_{5pj}z_{qri(t)} + \alpha_{6pj}\right) \quad (11)
\]

\[
r_{z_{ql}(t)}(j) = z'_{qr(i)}(j) - \left(\alpha_{1ql}(z_{qri(t)})^5 + \alpha_{2ql}(z_{qri(t)})^4 + \alpha_{3ql}(z_{qri(t)})^3 + \alpha_{4ql}(z_{qri(t)})^2 + \alpha_{5ql}z_{qri(t)} + \alpha_{6ql}\right) \quad (12)
\]

\(z_d\) represents measurement vector that the control room receives at the \(t^{th}\) instant. \(z'_{pr(i)}(j)\) and \(z'_{qr(i)}(j)\) denote the \(j^{th}\) entries of the received dummy measurement vectors \(z'_{pr(i)}\) and \(z'_{qr(i)}\), respectively at the \(t^{th}\) instant, having dummy measurement values for active and reactive powers. Similarly, \(z_{qri(t)}\) and \(z_{qri(t)}\) are picked from the received measurement vector. The \(j^{th}\) entries of residue vectors \(r_{z_{pl}(t)}(j)\) and \(r_{z_{ql}(t)}(j)\) are denoted by \(r_{z_{pl}(t)}(j)\) and \(r_{z_{ql}(t)}(j)\). The residue for active as well as the reactive powers that are injected to buses at \(t^{th}\) instant are contained in these residue vectors. To calculate the residue of forward as well as the backward active and reactive powers of transmission lines, the following equations are used:

\[
r_{z_{pl}(t)}(j) = z'_{p_{tr}(i)}(j) - \left(\alpha_{1pj}(z_{p_{tr}(i)})^5 + \alpha_{2pj}(z_{p_{tr}(i)})^4 + \alpha_{3pj}(z_{p_{tr}(i)})^3 + \alpha_{4pj}(z_{p_{tr}(i)})^2 + \alpha_{5pj}z_{p_{tr}(i)} + \alpha_{6pj}\right) \quad (13)
\]
In the above equations, the used actual measurement values as well as the dummy values are taken from the measurement vector that the control room receives. The \( t \)th entries of residue vectors, \( r_{zpl}(t) \), \( r_{zql}(t) \), \( r_{zlp}(t) \), and \( r_{zql}(t) \), are represented by \( r_{zpl}(t)(j) \), \( r_{zql}(t)(j) \), \( r_{zlp}(t)(j) \), and \( r_{zql}(t)(j) \). These residue vectors contain the residue of forward and backward active and reactive transmission lines powers at the \( t \)th instant. The system is considered as attacked when any of the entries of residue vectors is not equal to zero. The hacking of the measurement vector \( z_{dt} \) is performed by the intruder, and the FDI attack is launched. After the attack, the control room receives the measurement vector \( z_{dt} \). The attacker also attacks the dummy measurement values. The attacker is unaware of that function through which the dummy measurement values are calculated. Moreover, it becomes difficult for an attacker to predict the nonlinear function used in this case. So, if the attack is made by the attacker, the residue will be some value other than zero. In this way, the control room detects the attack.

4. Evaluation

The implementation of the proposed model is achieved on the IEEE 14-bus system with the AC power flow network. The IEEE 14-bus system [55] is a commonly used benchmark in power systems as well as electrical engineering research. It is a simplified representation of an electrical power network having 14 buses (nodes) and 20 branches (transmission lines). The IEEE 14-bus test system is generally used for many purposes, such as testing and the validation of the algorithms as well as the techniques in the power system analysis. For the purposes of the implementation and evaluation of the proposed model, realistic data for the AC power flow network is used. The data generation for the power flow network is achieved by using those load curves, according to standard realistic load curves. The load curves taken from the PJM are used as a reference for the purpose of data generation. The realistic data are used as the basis of the load curves of PJM. So, our generated data, by using those load curves, are extremely close to the realistic data. The data are generated for the four seasons, winter, summer, spring, and fall, according to standard load curves.

4.1. Data Generation

The generation of data is achieved for the IEEE 14-bus test system for the four different seasons by following the standard realistic load curve for each of the seasons. The time interval between two consecutive measurement vectors is one hour. This means that we take the measurements on an hourly basis. So, for a single day, the number of times the measurement vectors are obtained in the control room is 24. Data generation is performed for one year, that is, 365 days. So, the measurement vectors are obtained for 8760 different instances. The simulations of the model are performed by using MATPOWER 7.0. MATPOWER [57] is an open source, MATLAB-based tool used for the purpose of simulation as well as the optimization of the power system. It is commonly used in the field of electrical engineering, mainly for studying and analyzing power systems. In standard reference load
curves, load varies from 61% to 118.5% of its average value. So, for our data generation task, the load is also varied between these two values. After the variation of the overall load of the power system, the load curves for the four different seasons are given in Figure 1. These load curves are shown for a single day. There are 24 different instances, during which the load values are shown in the figure. These load curves are used for the generation of realistic data.

Figure 1. The seasonal load curves, which are used for the generation of data.

4.2. Method to Launch an Attack

In this manuscript, a certain method is adopted to launch an attack for the simulation purposes, and whenever attacks are made, this method is used for their construction. The attacks are carried out in 25% of measurement vectors for a single day. This means that the measurement vectors obtained at 6 instances out of 24 are taken as attacked. To make it generalized, those six instances are randomly chosen for one day. Out of the total measurement vectors chosen to be attacked, 50% are selected to be subjected to simple attacks, and in the remaining 50%, stealth attacks are carried out. This choice of instances for doing simple attacks or stealth FDI attacks is also carried out randomly. For the purpose of launching the stealth attacks, the attack vector is made by the method in which the Jacobian matrix is constructed first, and then the multiplication of that Jacobian matrix is achieved with vector \( c \). The selection of elements of vector \( c \) is carried out in such a way that the values between \(-1\) and 1 are randomly selected. So, its mean is zero, and the variance is 2. The attack vector is added to the measurement vector in order to launch the attack.

4.3. Limitations of Variable Dummy Value Model

The detection technique in [26] known as the variable dummy value model is able to successfully detect FDI attacks. As the dummy values are calculated through a linear function, it is possible for the attacker to find the exact function by using multivariate linear regression. The attacker uses this function in order to calculate and set dummy
measurement values while making the attack. In this case, the attacks launched by the attacker become undetectable for the VDVM technique. Figure 2 shows the results for the case in which the VDVM is bypassed by the attacks. The results for simple attacks as well as the stealth attacks are given in the figure. The figure also shows safe measurements. The measurements having simple attack are labelled with 0, stealth attack with $-1$, and the label used for the safe measurements is 1. All of the obtained dummy measurement values that are contained in the measurement vector are added to obtain a single combined power value at a particular instant. Similarly, all the recalculated dummy values are also added to obtain another combined power at that instant. The absolute value of both the combined powers at every instant are plotted for one complete day in Figure 2, and two curves are drawn, that is, one by using combined powers from obtained values and the second by drawing from the combined powers of the recalculated values. Both of the curves are the same at all the instances, which shows that the combined value of obtained powers is equal to the combined value of the recalculated powers for all the instances. For a secure system, both values should not be the same when the attack is made. But in this case, both of the values are the same even for simple attacks as well as for stealth attacks. So, the variable dummy value model fails to detect the simple as well as stealth attacks, as shown by the results.

In Figure 3, the combined power of the obtained dummy values and the combined power of recalculated dummy values at every instant for a complete day are plotted by using the bar graph. The first bar shows the absolute value of the combined power of the obtained dummy values, and the second bar displays the absolute value of the combined power of recalculated dummy values. The third bar gives the absolute value of residue. Here, the residue is calculated by taking the difference of the combined power of the recalculated dummy values from the combined power of the obtained dummy values. When the system is secure, if an attack is made, this residue must not give a zero value. But as the residue is zero in the cases of simple attacks and the stealth FDI attacks, VDVM is bypassed in this way.
4.4. Results of Nonlinear Function Based Variable Dummy Value Model

The limitation of VDVM is removed by proposing the NF-VDVM technique. It becomes extremely difficult for the attacker to attack the system due to the use of nonlinear functions in the proposed NF-VDVM technique. The constants used in the polynomial, through which the dummy measurement values are calculated, are learned via polynomial regression. In the case of the polynomial regression model, the dummy measurement values are selected as the outputs for the purpose of finding the relation between the input and output. For all of the equations implemented for calculating the dummy values of power, the polynomial regression model is run, and the best values are obtained for the constants of all the polynomial functions. The learning of the polynomial regression model for finding the parameters of the equation that is implemented to find the dummy measurement values of $P_1$ is shown in Figure 4. $P_1$ represents the active power that is injected to bus 1, and $P_{12}$ denotes the forward active power flowing through the first transmission line (active power flowing from bus 1 to bus 2 in the forward direction). The suitable constants for all of the equations are found, and then those equations are placed in meters to compute the dummy values.

In Figure 5, the results of the proposed NF-VDVM for simple as well as stealth attacks are shown by plotting the curves for the obtained dummy measurement values and the recalculated dummy measurement values of $P_1$ (active power that is injected to bus 1) for a complete day. In the figure, the measurements that have a simple attack are labelled with 0, stealth attack with $-1$, and safe measurements are labelled with 1. The absolute value of the obtained dummy values as well as the recalculated dummy values is used here to draw the curves. For safe measurements, the obtained dummy measurement value as well as the recalculated dummy measurement value of $P_1$ should be the same, and they should be different in case of simple and stealth FDI attacks. From the figure, it is shown that both the curves are same at the instances of safe measurements. However, for the cases of simple attacks as well as the stealth FDI attacks, the curves are not the same showing that the proposed NF-VDVM technique successfully detects the simple attacks as well as the stealth FDI attacks. In the figure, at some of the instants when the attacks are launched, the recalculated dummy values and the obtained dummy values are close to each other, but
they are not exactly the same. If there is no attack, both of the values should be exactly the same. So, even a small difference between them is enough to detect the attacks.

**Figure 4.** Training of polynomial regression model for finding the coefficients of the equation through which the dummy measurement values of $P_1$ are calculated.

**Figure 5.** Results of NF-VDVM in the form of curves for simple as well as stealth attacks by using the obtained dummy measurement values and recalculated dummy measurement values of $P_1$. 

**Figure 6.** Bar graph to display the results of NF-VDVM for simple as well as stealth FDI attacks. The first bar at every instant shows the absolute value of the obtained dummy value of $P_1$ at that particular instant. The second bar displays the absolute value of the recalculated dummy value of $P_1$ at that instant. However, the third bar shows the absolute value of residue, whose calculation procedure is given in Section 3. To find the value of residue, the recalculated dummy values that use the received actual values from the measurement vector, are subtracted from the received dummy values present in the measurement vector. The system is said to be secure when the residue contains some value other than zero at the instances of attacks. It is shown in Figure 6 that the value of residue is zero for the instances of safe measurement, and nonzero for simple attacks as well as stealth attacks. The results show the success of our proposed method NF-VDVM against simple attacks and stealth attacks.
Figure 6 shows a bar graph to display the results of NF-VDVM for simple as well as stealth FDI attacks. The first bar at every instant shows the absolute value of the obtained dummy value of $P_1$ at that particular instant. The second bar displays the absolute value of the recalculated dummy value of $P_1$ at that instant. However, the third bar shows the absolute value of residue, whose calculation procedure is given in Section 3. To find the value of residue, the recalculated dummy values that use the received actual values from the measurement vector, are subtracted from the received dummy values present in the measurement vector. The system is said to be secure when the residue contains some value other than zero at the instances of attacks. It is shown in Figure 6 that the value of residue is zero for the instances of safe measurement, and nonzero for simple attacks as well as stealth attacks. The results show the success of our proposed method NF-VDVM against simple attacks and stealth attacks.

Figure 6. Results of NF-VDVM in the form of bar graph for simple as well as stealth attacks by using the obtained dummy values and recalculated dummy values of $P_1$.

The results of NF-VDVM are also shown in Figure 7 by using the combined power values of received dummy values and recalculated dummy values. The combined power of the received dummy values at a particular instant is calculated by adding all the obtained dummy values of power present in the measurement vector at that instant. Similarly, the combined power of the recalculated dummy values at a particular instant is calculated by adding all the recalculated dummy values of power at that instant. The absolute value of the combined power of the obtained dummy values and the absolute value of the combined power of the recalculated dummy values are plotted in Figure 7 at all the instances for a complete day. The curves are drawn by using the combined power values at all the instances. For a secure system, both of the curves (one from the combined power values of received dummy values and the other from the combined power values of recalculated dummy values) should not be same at the instances where the attack is made. It is shown from Figure 7 that at the instances of simple attacks as well as stealth FDI attacks, the curves are not the same, which proves that the simple attacks as well as the stealth attacks are detected by the proposed NF-VDVM method.

Figure 8 displays the results of NF-VDVM in the form of a bar graph for the combined power values of the received dummy values as well as the combined power values of the recalculated dummy values. The first bar shows the absolute value of the combined power of the obtained dummy values present in the measurement vector at a particular instant. The second bar gives the absolute value of the combined power of recalculated dummy
values at that instant. The third bar shows the absolute value of residue, which is calculated here by subtracting the combined power of recalculated dummy values from the combined power of obtained dummy values at that particular instant. A system is said to be secure when the value of this residue is not equal to zero for the instances at which the attacks are carried out. Figure 8 shows that the value of the residue is zero at the instances of the same measurements. However, in the case of simple attacks as well as stealth FDI attacks, the residue is not zero. It proves that the simple attacks and stealth FDI attacks are successfully detected by our proposed method NF-VDVM.

![Diagram showing the results of NF-VDVM in the form of curves for simple as well as stealth attacks by using the combined dummy values of power.](image)

**Figure 7.** Results of NF-VDVM in the form of curves for simple as well as stealth attacks by using the combined dummy values of power.

The computational cost of the VDVM and NF-VDVM techniques is calculated. The simulation time of both techniques is computed and is given in Table 1.

The simulations are performed on the system, which has the following characteristics: Intel Core i7 (8th Gen) @2.20 GHz 2.21 GHz, RAM 16 GB (HP Omen, Canada), System Type 64-bit operating system, ×64-based processor. In order to compute the simulation time, the simulations are run five times for the VDVM and NF-VDVM techniques. The time is noted for each of the technique at each trial. Finally, the average simulation time of VDVM and NF-VDVM is calculated, and is shown in Table 1. The average simulation time of VDVM is less than that of NF-VDVM. The NF-VDVM technique has higher computational cost compared to VDVM due to the use of higher order functions for the calculation of dummy values in NF-VDVM, but the efficiency of NF-VDVM in the detection of FDI attacks is higher than that of VDVM. Although VDVM has less computational cost, due to its limitations, the security of the smart grid is compromised in the case of VDVM, which is a thing that cannot be overlooked. So, the NF-VDVM technique is preferred over the VDVM in order to protect the smart grid against FDI attacks.

The NF-VDVM technique is capable of detecting FDI attacks in an efficient way, but it may have some limitations and validity threats. The noise in the measurement data can affect the performance of NF-VDVM. Moreover, if the intruder somehow becomes familiar with the functions implemented for the calculation of dummy measurement values, and then the attack is launched, it may be difficult for the technique of NF-VDVM to detect the attack.
In a smart grid, the hacking of the communication links is carried out by the attackers in order to make different types of attacks. A huge loss and damage can be suffered through these cyber-attacks. The intruder can obtain financial benefits from these attacks. So, the detection of these cyber-attacks is very important for the accurate and uninterrupted power flow. The bad data detection techniques such as the DC state estimation as well as the AC state estimation are incapable of detecting the stealth FDI attacks. The method of VDVM also fails in a certain scenario. To tackle the limitations, a new technique named NF-VDVM was implemented into the power system to detect stealth FDI attacks. NF-VDVM successfully detected the attacks, and power system was made secure.

**Figure 8.** Results of NF-VDVM in the form of bar graph for simple as well as stealth attacks by using the combined dummy values of power.

**Table 1.** The simulation time of the VDVM and NF-VDVM techniques.

<table>
<thead>
<tr>
<th>Trial No.</th>
<th>VDVM (Seconds)</th>
<th>NF-VDVM (Seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.568</td>
<td>0.756</td>
</tr>
<tr>
<td>2</td>
<td>0.449</td>
<td>0.535</td>
</tr>
<tr>
<td>3</td>
<td>0.484</td>
<td>0.636</td>
</tr>
<tr>
<td>4</td>
<td>0.394</td>
<td>0.862</td>
</tr>
<tr>
<td>5</td>
<td>0.528</td>
<td>0.752</td>
</tr>
<tr>
<td><strong>Average Time</strong></td>
<td><strong>0.485</strong></td>
<td><strong>0.708</strong></td>
</tr>
</tbody>
</table>

5. Conclusions

In a smart grid, the hacking of the communication links is carried out by the attackers in order to make different types of attacks. A huge loss and damage can be suffered through these cyber-attacks. The intruder can obtain financial benefits from these attacks. So, the detection of these cyber-attacks is very important for the accurate and uninterrupted power flow. The bad data detection techniques such as the DC state estimation as well as the AC state estimation are incapable of detecting the stealth FDI attacks. The method of VDVM also fails in a certain scenario. To tackle the limitations, a new technique named NF-VDVM was implemented into the power system to detect stealth FDI attacks. NF-VDVM successfully detected the attacks, and power system was made secure.
Author Contributions: Conceptualization, M.A.S., F.A. and R.N.; data curation, S.U.K.; formal analysis, F.A. and S.U.K.; funding acquisition, A.W.; methodology, M.A.S. and F.A.; resources, S.U.K. and H.A.; software, R.N.; supervision, F.A., A.W. and H.A.; validation, S.U.K. and H.A.; visualization, R.N. and A.W.; writing—original draft, M.A.S.; writing—review and editing, F.A., A.W. and H.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

2. Ghiasi, M.; Wang, Z.; Mehrandezh, M.; Jalilian, S.; Ghadimi, N. Evolution of smart grids towards the Internet of energy: Concept and essential components for deep decarbonisation. IET Smart Grid 2023, 6, 86–102. [CrossRef]
38. Han, Y.; Feng, H.; Li, K.; Zhao, Q. False data injection attacks detection with modified temporal multi-graph convolutional network in smart grids. *Comput. Secur.* 2023, 124, 103016. [CrossRef]
43. Li, X.; Wang, Y.; Lu, Z. Graph-based detection for false data injection attacks in power grid. *Energy* 2023, 263, 125865. [CrossRef]


51. Qu, Z.; Zhang, X.; Gao, Y.; Peng, C.; Wang, Y.; Georgievitch, P.M. Detection of False Data Injection Attack in AGC System Based on Random Forest. Machines 2023, 11, 83. [CrossRef]


Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.