Distributed host-based collaborative detection for false data injection attacks in smart grid cyber-physical system

Li, Beibei; Lu, Rongxing; Wang, Wei; Choo, Kim-Kwang Raymond

2016

Li, B., Lu, R., Wang, W., & Choo, K.-K. R. (2017). Distributed host-based collaborative detection for false data injection attacks in smart grid cyber-physical system. Journal of Parallel and Distributed Computing, 103, 32-41.

https://hdl.handle.net/10356/80762

https://doi.org/10.1016/j.jpdc.2016.12.012

Distributed Host-based Collaborative Detection for False Data Injection Attacks in Smart Grid Cyber-Physical System

Beibei Li[a], Rongxing Lu[b,*], Wei Wang[a], Kim-Kwang Raymond Choo[c,d]

[a]*School of Electrical and Electronic Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798*
[b]*Faculty of Computer Science, University of New Brunswick, Fredericton, Canada E3B 5A3*
[c]*Department of Information Systems and Cyber Security, University of Texas at San Antonio, San Antonio, TX 78249, USA*
[d]*School of Information Technology & Mathematical Sciences, University of South Australia, Adelaide, SA 5001, Australia*

Abstract

False data injection (FDI) attacks are a crucial security threat to smart grid cyber-physical system (CPS), and could result in cataclysmic consequences to the entire power system. However, due to the high dependence on open information networking, countering FDI attacks is challenging in smart grid CPS. Most existing solutions are based on state estimation (SE) at the highly centralized control center; thus, computationally expensive. In addition, these solutions generally do not provide a high level of security assurance, as evidenced by recent work that smart FDI attackers with knowledge of system configurations can easily circumvent conventional SE-based false data detection mechanisms. In this paper, in order to address these challenges, a novel distributed host-based collaborative detection method is proposed. Specifically, in our approach, we use a conjunctive rule based majority voting algorithm to collaboratively detect false measurement data inserted by compromised phasor measurement units

*Corresponding author.
Email addresses: bli012@e.ntu.edu.sg (Beibei Li), rlu1@unb.ca (Rongxing Lu), wei001@e.ntu.edu.sg (Wei Wang), raymond.choo@fulbrightmail.org (Kim-Kwang Raymond Choo)

(PMUs). In addition, an innovative reputation system with an adaptive reputation updating algorithm is also designed to evaluate the overall running status of PMUs, by which FDI attacks can be distinctly observed. Extensive simulation experiments are conducted with real-time measurement data obtained from the PowerWorld simulator, and the numerical results fully demonstrate the effectiveness of our proposal.

Keywords: Smart grid cyber-physical system (CPS), False data injection attack, Distributed host-based collaborative detection, Adaptive reputation system

---

1. Introduction

Smart grid cyber-physical system (CPS) is designed to facilitate highly efficient, accurate, and reliable power delivery as well as sustainable energy integration and utilization [1, 2]. Despite the potential benefits of a smart grid CPS, there are underlying threats that could jeopardize the security of the system and consequently, have a cascading effect on the stability of the society [1, 3, 4, 5].

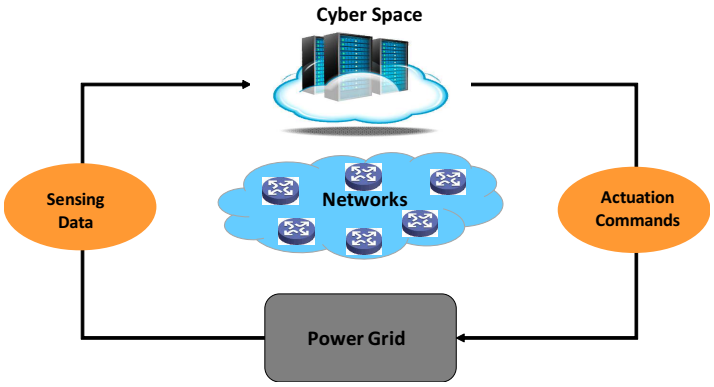


Fig. 1: The system view of smart grid CPS cyber-physical system.

In recent times, a number of high profile incidents targeting smart grid as well as other CPSs have been reported, e.g., Stuxnet [6], Conficker [7], and US drones hack [8]. Malicious attackers may attempt to falsify sensor measurements, embed fake control commands, delay or drop sensor readings or control commands [1, 9, 10, 11]. False data injection (FDI) attacks are increasingly recognized as a serious threat to smart grid CPS, and unsurprisingly, have been the focus of computer security researchers and industry practitioners. FDI attacks and mitigation strategies on smart grid CPS have been also evolved over the years. Conventional false data detection (FDD) approaches are generally based on system state estimation (SE) [12, 13, 14]. However, Liu *et al.* in [15] showed that smart FDI attackers armed with the knowledge of system configurations could easily bypass the traditional SE-based FDD schemes without detection.

Consequently, existing FDD approaches may be ineffective against newer or emerging FDI attacks. The major limitation of legacy FDD schemes is that they mainly focus on the inter-correlations among the measurement data (e.g., residuals and errors), rather than the malicious behaviors of meter devices, such as phasor measurement units (PMUs) and smart meters. Furthermore, in existing literature FDD is generally performed by the power system's centralized control center (CC), due to the demanding computational requirements [13, 14]. Although a small number of hierarchical or distributed FDD schemes are designed to reduce the computation requirements at the CC [16, 17], most of them are still based on SE; thus, vulnerable to smart attackers. Another limitation of legacy FDD methods is that some prevailing countermeasures against cyber intrusion only aim to detect the "bad" data without further evaluating the true running status of the meter devices that might already be compromised by malicious attackers [12, 16, 18]. These undetected hidden attackers can continue to launch or improve their attacks subsequently. Therefore, countering against FDI attacks in smart grid CPS remains a research challenge, and one that we seek to address in this paper.

Thus, we propose a distributed host-based collaborative detection (DHCD) method based on rule specifications, rather than SE. DHCD can not only reduce the computational burden of the CC, but also achieve fast FDD and the capability to evaluate the running status of meter devices. Specifically, in our method, each PMU is assigned a host monitor (HM) serving as the distributed local false data detector. Based on a set of pre-defined rule specifications, the monitors determine the anomalous levels of measurement data collected by their supervised PMUs. Then, by sharing and comparing the anomalous levels of the measurement data collected by the neighboring interconnected PMUs, these interconnected monitors collaboratively make a decision based on the majority

voting algorithm to determine whether their own measurement data is falsified. To evaluate the overall running status of the PMUs, a reputation system with an adaptive reputation updating (ARU) algorithm is designed, where a malfunction PMU can be easily identified. The contributions of our work are summarized as follows:

1. We develop a DHCD method to detect FDI attacks in smart grid CPS based on rule specifications, which can be used to effectively mitigate smart FDI attacks.
2. Our method can not only achieve fast and high accuracy of FDD, but also allow the identification of compromised PMUs using our designed reputation system.
3. Our distributed detection method will "displace" the computational burden of the CC by delegating FDD tasks to the local monitors.

The remainder of this paper is organized as follows. Section 2 reviews the related literature. Section 3 presents the system model, the threat model, and our design goals. The DHCD method is detailed in Section 4, followed by the performance evaluation in Section 5. Section 6 concludes the paper with future research directions.

2. Related Work

Intrusion detection has been extensively studied in the literature [19, 20], including for smart grids [1, 9, 11], wireless sensor networks [21, 22, 23], mobile ad hoc networks [24], etc.

Since the seminal work of Schweppe *et al.* who proposed a static SE-based approach to detect bad data in electric power systems [25], FDD has been the focus of research in the power system industry. Over the years, a number of FDD approaches based on SE designed to mitigate FDI attacks in smart grid CPS

have been proposed [12, 13, 18]. For example, Merrill and Schweppe presented a bad data suppression estimator based on a non-quadratic cost function to improve the performance of static SE [12]. Handschin *et al.* presented a method to detect and identify the bad data and structural error problems, and improved bad data analysis (detection probability, and effects of bad data) [18]. Cutsem *et al.* also proposed an identification method attempting to alleviate some existing difficulties, such as multiple and interacting bad data [26].

However, Liu *et al.* demonstrated that a new class of smart attackers armed with the knowledge of system configurations were capable of constructing a set of falsified data to circumvent the legacy SE-based FDD mechanisms [15]. Xie *et al.* also explained that some potential attackers were able to launch FDIs in deregulated electricity markets [2]. Thus, a small number of detection methods have been proposed to identify such "undetectable" attackers. Pasqualetti proposed a unified framework and advanced monitoring procedure to detect malfunctions or measurement corruptions of network components caused by an omniscient adversary [27]. Bobba *et al.* attempted to detect smart FDI attacks by protecting a strategically selected set of sensor measurements and finding a way to independently verify or measure these measurements [28].

Rather than using the static SE and to fully leverage the features of meter devices' anomalous behaviors, our proposed DHCD method mitigates FDIs by establishing a rule specification based behavior model and collaboratively verifying the measurement data. In addition, we design a novel reputation system with an ARU algorithm to evaluate the running status of PMUs, by which FDI attacks can be easily observed. Furthermore, our distributed detection system can significantly enhance the efficiency of FDD tasks.

3. Models and Design Goals

In this section, we introduce the system model, the threat model, and our design goals.

3.1. System Model

A smart grid CPS is a fully automated system capable of achieving self-healing, cost reduction, improved reliability and efficiency. These promising benefits are intensively grounded on the wide area measurement and control system (WAMCS), as it can provide high-level observability and controllability in power system operations [29, 30, 31]. Thus, in this paper, we consider the WAMCS as our system model.

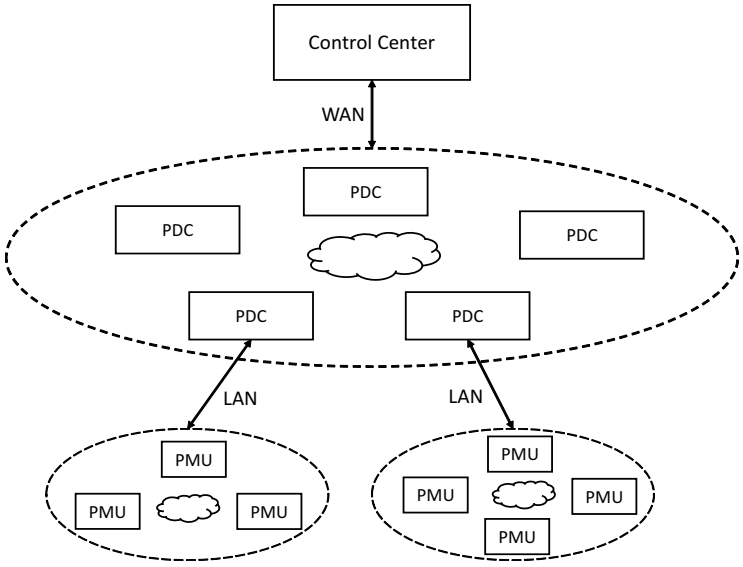


Fig. 2: The architecture of wide area measurement and control system.

As shown in Fig. 2, WAMCS is an integrated system consisting of PMUs, phasor data concentrators (PDCs), heterogenous communication networks, and a CC. Specifically, PMUs, located at the substations of the power generation and transmission system, are capable of measuring the real-time status of the

power system. For example, the real-time amplitude and phase angle of voltage at the bus, of current on the transmission line, and of the power at each branch, can be measured by the PMUs. These measurement data are then periodically transmitted to the the PDCs, usually in 50Hz, through the local area network (LAN). Then, the aggregated data at the PDCs are delivered to the CC via the wide area network (WAN) for further data analysis, such as state estimation, event diagnostics, and contingency analysis.

3.2. Threat Model

The real-time data provided by PMUs serve as the basis for automated, efficient, and reliable system control. However, adversaries seeking to intervene or manipulate system operations can attempt to inject false measurement data through compromised PMUs. Successful FDI attack may compromise the above-mentioned promising functionalities or even jeopardize the system operations.

In our threat model, we consider that PMUs in the WAMCS can be compromised by FDI attackers (e.g., rewriting the program settings, or stealing the secret information for data communication). Note that, in smart grid CPS, a single piece of false measurement data may not have significant impact on system operations, because the system is capable of correcting trivial faults or mistakes. However, the system may not be able to auto-correct in the event that consecutive false measurement data are received; consequently, resulting in system failures. As such, to successfully launch an FDI attack in practice, attackers usually recklessly and persistently inject false measurement data once they have an opportunity. This is the behavior pattern of FDI attackers we consider in the threat model.

3.3. Design Goals

Based on the aforementioned system model and threat model, our design goals are to develop an accurate, efficient, and scalable FDD method in smart

grid CPS. Specifically, the following specific objectives should be achieved.

Accuracy: The devised method is able to effectively detect smart FDI attacks, achieving both high detection rate and low false alarm rate.

Efficiency: The detection method should not introduce additional computational burden to the system, particularly to the CC inherent in traditional FDD schemes.

Scalability: The smart grid CPS needs to be scalable (similar to a cloud system) by allowing new devices to be added, etc, without incurring expensive (financial) costs.

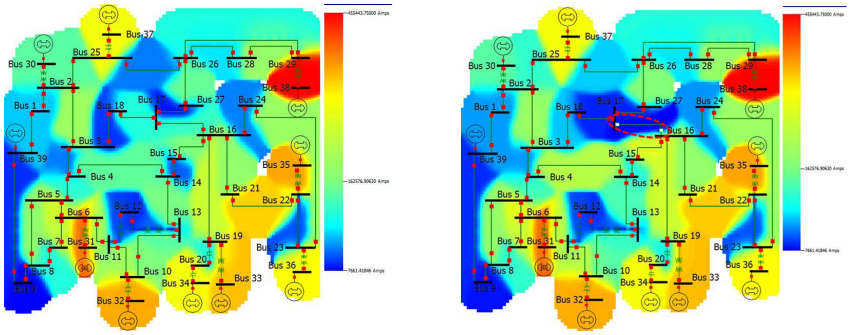
4. Proposed DHCD Method

In this section, we present the proposed DHCD method, which is composed of two steps (subsections): collaborative FDD and determination of compromised PMU. In the first step, we employ a set of rule specifications to identify anomalous measurement data reported by the PMU. Then, in the second step, we devise a reputation system with an ARU algorithm to monitor and assess PMUs' overall behaviors in order to further detect compromised PMU.

4.1. Collaborative FDD

In normal operational circumstances, the power grid operates in a stable status. In other words, all state variables vary in a mutual balanced manner according to Kirchhoff's law, demand-response constraints, etc. As such, any change of a variable state on one bus or transmission line, resulting from either the normal demand variation or system faults, would lead to corresponding state changes of the same and/or other variables on interconnected buses or transmission lines. For example, as shown in Fig. 3, the contouring maps with comparison are plotted, which describe the distribution of the current amplitude on each transmission line (a) before and (b) after an open circuit event on

transmission line from Bus 16 to Bus 17. As shown in Fig. 3(b), after the occurrence of this open circuit event, the current amplitude values near Line 16 to 17 shift. The closer to this line, the more the value changes.



(a) (b)

Fig. 3: Comparison of contouring maps describing the distribution of current amplitude on transmission lines: (a) before open circuit and (b) after open circuit on line from Bus 16 to 17 (marked by a red circle) in IEEE-39 bus system. As the bar shows, red area denotes high current amplitude while blue area denotes low current amplitude.

In contrast, if only some changes of variable states occur on one bus, without a corresponding shift in the parallel variables of interconnected buses, such changes can be regarded as anomalous. These anomalies may originate from either malfunction PMU devices or malicious activities due to compromised PMUs. In this paper, we only consider possible malicious activities rather than device malfunction, as there are many existing approaches to address issues relating to device malfunction. Based on the inter-correlations of power systems, we design a collaborative detection method to detect anomalous measurement data reported by PMUs [32, 33].

4.1.1. Normal Rule Specifications

When power system is under normal operation, all state variables must naturally follow some constraints and hold some properties. Let us take active power P as an example, which should obey the following rules:

- $P_{min} < P^t < P_{max}$: P at any time under stable status must vary within an experienced range $[P_{min}, P_{max}]$.
- $|P^t - P^{t-1}| < P_\Delta$: The variation of P within one time interval should be less than an experienced threshold P_Δ.
- $|P^t_{in} - P^t_{out}| < P_{loss}$: The difference of P flowing into a bus and flowing out the bus ought to be less than an experienced power loss threshold P_{loss}.
- Other more complicated rules.

As such, we pre-define some rule specifications as listed in Table 1 that PMUs have to coincide with in the stable status. These rule specifications serve as the basis of our method to identify the anomalous measurement data (for convenience, the superscript t is omitted).

Table 1: Rules specifications for PMUs in stable status

Index	Variable	Rule Description
1	Active Power Angle	$\Delta\delta < \delta_\Delta$
2	(Phase A) Voltage Amplitude	$\Delta V < V_\Delta$
3	Load Mvar	$\Delta L_{Mvar} < L_{Mvar\Delta}$
4	Load MW	$\Delta L_{MW} < L_{MW\Delta}$

To represent the results of whether the rule specifications have been violated, we employ a binary system, where "0" denotes that the measurement data of one variable follows the relevant rule specification and "1" indicates a violation. A binary sequence with length E (E is the number of rule specifications, and here E is 4) is utilized to represent the conjunctive results pertaining to the entire measurement data. For instance, "1001" denotes that both rules 1 and 4 are violated. A non violation of the conjunctive four rule specifications is represented by "0000", which is our *baseline* of PMUs' behaviors.

In order to assess to what extent each piece of measurement data is anomalous, we introduce a normalized Euclidean distance strategy to determine the

anomalous level l^t, which is shown as follows:

$$l^t = D_0(seq^t, seq_0), \tag{1}$$

where seq^t is the binary sequence representing the conjunctive results of measurement data at time t, while seq_0 = "0000" is the baseline. D_0 is the normalized Euclidean distance of the two sequences seq^t and seq_0. Euclidean distance is the square root of the sum of results that are different between two sequences. For example, the Euclidean distance between sequence "1001" and the *baseline* "0000" is $\sqrt{1^2+0+0+1^2} \approx 1.414$. Then, the *anomalous level* l is computed by the normalized distance, i.e., $1.414/\sqrt{1^2+1^2+1^2+1^2} \approx 0.707$.

4.1.2. FDD algorithm with Iterative Majority Voting

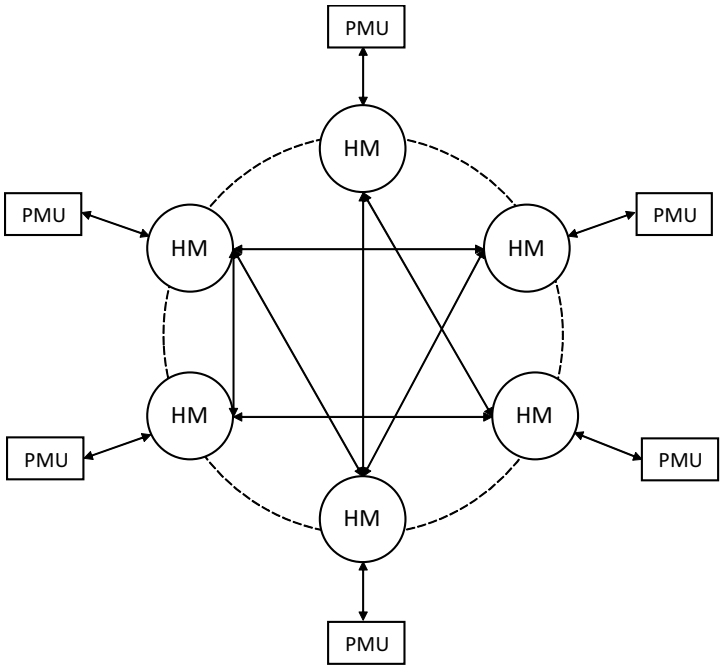


Fig. 4: The distributed host-based collaborative FDD system.

Figure 4 shows the distributed host-based collaborative FDD system, where

each host monitor (HM) is responsible for monitoring and assessing the behaviors of its administrated PMU. Let $\mathcal{M} = \{M_1, M_2, \cdots, M_N\}$ denote the set of monitors and $\mathcal{U} = \{U_1, U_2, \cdots, U_N\}$ the set of PMUs, where N is the total number of HMs or PMUs. HMs communicate among each other following the connection pattern of the PMUs, which means each HM only communicates with HMs that their monitored PMUs have interconnection relations.

As stated above, we utilize the inter-correlations between the state variables to build our detection method. Algorithm 1 outlines the FDD algorithm with iterative majority voting process. Concretely, set $\mathcal{M}$ is initialized as $\mathcal{M} = \{M_1, M_2, \cdots, M_N\}$, and a flag variable $repeat\_flag$ as '0'. Note that $repeat\_flag$ = '0' indicates that the procedure does not need to be repeated, while $repeat\_flag$ = '1' indicates the need to repeat the procedure. Next, each monitor $M_i \in \mathcal{M}$ determines the conjunctive result R_i^t of current piece of measurement data, and broadcasts the result to neighbouring connected monitors $\mathcal{M}_i = \{M_j | M_j \sim M_i\}$. An example is shown in Fig. 5.

$\mathbf{M_1}$:

Rule_Index	R1	R2	R3	R4
Rule_Result	0	0	0	0

$\mathbf{M_2}$:

Rule_Index	R1	R2	R3	R4
Rule_Result	0	0	0	1

•
•
•

$\mathbf{M_N}$:

Rule_Index	R1	R2	R3	R4
Rule_Result	0	1	0	0

Fig. 5: An example of the conjunctive results transmitted between HMs.

Then, M_i launches the false data identification process. If there is no bit "1" in the result R_i^t, then no false data is detected. Otherwise, M_i needs to determine

how many of its connected monitors have a bit "1" in their conjunctive results R_j^t. If more than or equal to half of the connected monitors have a bit "1" at the same position in R_j^t, M_i concludes that U_i has reported a piece of false measurement data; otherwise, R_i^t is tentatively considered suspicious. After all $M_i \in \mathcal{M}$ have concluded the first procedure, the termination criterion is determined. If $repeat\_flag ==$ '1', this procedure is repeated to further identify the false data; otherwise, the procedure goes to the end.

Algorithm 1 FDD Algorithm

initialization: $\mathcal{M} = \{M_1, M_2, \cdots, M_N\}$, $Upperbound = 5$, $Iteration = 0$, $repeat\_flag =$ '0'
procedure
  for each monitor $M_i \in \mathcal{M}$ **do**
    (1). determines the conjunctive result R_i^t of current piece of measurement data.
    (2). broadcasts the result R_i^t to the neighbouring connected monitors $\mathcal{M}_i = \{M_j | M_j \sim M_i\}$.
    (3). **identifies false data:**
    if there is no bit "1" in the result R_i^t **then**
      output: no false data detected.
    else if more than or equal to half of the monitors in $\mathcal{M}_i$ hold bit "0" at the same position in the result R_j^t **then**
      (a). **output**: false data detected.
      (b). removes M_i from $\mathcal{M}$ and its connections with other monitors.
    else
      (a). keeps R_i^t as suspicious result.
      (b). $repeat\_flag =$ '1'.
    end if
  end for
  (4). **judges the termination criteria:**
  if $repeat\_flag ==$ '1' and $Iteration < Upperbound$ **then**
    (a). repeats **procedure**.
    (b). $Iteration = Iteration + 1$.
  else
    ends the procedure.
  end if
end procedure

4.2. Determination of Compromised PMU

FDD step is a critical process to detect false data, but it is not sufficient to identify compromised PMUs. Therefore, in the second step, we employ a

reputation-based algorithm to monitor and assess the PMUs' overall behaviors over a period of time, which allows us to identify compromised PMUs if their reputation level drops below an acceptable threshold [34, 35].

Specifically, in this subsection, we first model the probability distribution of the anomalous level of measurement data with a Beta distribution. Then, we estimate its two shape parameters α and β using maximum likelihood estimation (MLE) and Newton-Raphson method. Then, a detailed description of an adaptive reputation updating (ARU) algorithm is presented.

4.2.1. Probability Distribution of Anomalous Level

Let random variable X be the anomalous level of a piece of measurement data, where X can either be 0 or 1 and it is determined by the normalized Euclidean distance (see section 4.1.1). Particularly, $X = 0$ represents compliance of the rule specifications, while $X = 1$ represents a violation. Here, to determine the exact distribution of the probabilities of different anomalous level and its future values, we model the random variable X using a $Beta(\alpha, \beta)$ distribution. Beta distribution family can represent a collection of probability distributions, and can be used to depict a prior distribution of an unknown distribution with only a series of collected observations.

The probability density function (pdf) of a Beta distribution is

$$f(x; \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \, x^{\alpha-1}(1-x)^{\beta-1}, \tag{2}$$

where α and β are the two shape parameters. The mean value of a Beta distribution is

$$\mu = E[X] = \int_0^1 x \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \, x^{\alpha-1}(1-x)^{\beta-1} dx = \frac{\alpha}{\alpha + \beta}. \tag{3}$$

To obtain the exact distribution of X, we estimate the parameters α and

β using a well-known method MLE. We suppose that the n independent and identically distributed observations $\{x_1, x_2, ..., x_n\}$ are from an unknown distribution with pdf $f_0(\cdot|\theta), \theta$ is a vector of parameters. As for our model, the Beta distribution, $\theta = [\alpha\ \beta]$. By using MLE, we formulate the joint density probability function of these n independent and identically distributed observations $\{x_1, x_2, ..., x_n\}$ as

$$f(x_1, x_2, \ldots, x_n \mid \alpha, \beta) = \prod_{i=1}^{n} f(x_i \mid \alpha, \beta). \tag{4}$$

Now we look at this equation from a different perspective by fixing the observed samples $\{x_1, x_2, ..., x_n\}$ of this function, then α, β are the variables of the function that we call the likelihood:

$$\mathcal{L}(\alpha, \beta \mid x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f(x_i \mid \alpha, \beta). \tag{5}$$

In most cases, it is easier to work with the natural logarithm of the likelihood function. We rewrite it as

$$\begin{aligned}
&\ln \mathcal{L}(\alpha, \beta \mid x_1, x_2, \ldots, x_n) = \ln \prod_{i=1}^{n} f(x_i \mid \alpha, \beta) \\
&= \sum_{i=1}^{n} \ln \left\{ \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\ {x_i}^{\alpha-1}(1-x_i)^{\beta-1} \right\} \\
&= n \ln \Gamma(\alpha+\beta) - n[\ln \Gamma(\alpha) + \ln \Gamma(\beta)] + (\alpha-1) \sum_{i=1}^{n} \ln x_i + (\beta-1) \sum_{i=1}^{n} \ln(1-x_i).
\end{aligned} \tag{6}$$

Then, we have to find the optimal values of α and β that maximize $\ln \mathcal{L}(\alpha, \beta \mid x_1, \ldots, x_n)$. Since logarithm is a strictly monotonically increasing function, the

maximum value, if it exists, could be calculated by

$$\begin{cases} \frac{\partial \ln \mathcal{L}}{\partial \alpha} = 0 \\ \frac{\partial \ln \mathcal{L}}{\partial \beta} = 0 \end{cases} \tag{7}$$

That is

$$g_1(\alpha, \beta) = \psi(\alpha) - \psi(\alpha + \beta) - \frac{1}{n} \sum_{i=1}^{n} \ln x_i = 0 \tag{8}$$

$$g_2(\alpha, \beta) = \psi(\beta) - \psi(\alpha + \beta) - \frac{1}{n} \sum_{i=1}^{n} \ln(1 - x_i) = 0 \tag{9}$$

where $\psi(x)$ is the digamma function defined as

$$\psi(x) = \frac{d}{dx} \ln \Gamma(x) = \frac{\Gamma'(x)}{\Gamma(x)}. \tag{10}$$

There is no closed-form solution to Equations (8) and (9), so we use the Newton-Raphson method to find the approximate roots. The parameters $\hat{\theta} = [\hat{\alpha}\ \hat{\beta}]$ can be iteratively estimated by [36]

$$\hat{\theta}_{i+1} = \hat{\theta}_i - \frac{\mathbf{g}(\hat{\theta}_i)}{\mathbf{J}_{\mathbf{g}}(\hat{\theta}_i)}, \tag{11}$$

where $\mathbf{g} = [g_1\ g_2]$, and $\mathbf{J}_{\mathbf{g}}(\hat{\theta}_i)$ is an 2×2 Jacobian matrix defined over the function vector $\mathbf{g}(\hat{\theta}_i)$ defined as

$$\begin{bmatrix} \frac{d\mathbf{g_1}}{d\alpha} & \frac{d\mathbf{g_1}}{d\beta} \\ \frac{d\mathbf{g_2}}{d\alpha} & \frac{d\mathbf{g_2}}{d\beta} \end{bmatrix} \tag{12}$$

with

$$\frac{d\mathbf{g_1}}{d\alpha} = \psi'(\alpha) - \psi'(\alpha + \beta) \tag{13}$$

$$\frac{d\mathbf{g_1}}{d\beta} = \frac{d\mathbf{g_2}}{d\alpha} = -\psi'(\alpha + \beta) \tag{14}$$

$$\frac{d\mathbf{g_2}}{d\beta} = \psi'(\beta) - \psi'(\alpha + \beta) \tag{15}$$

This Newton-Raphson method converges when the estimates of $\hat{\theta}$ and $\hat{\beta}$ change by less than a acceptable threshold with each successive iteration.

4.2.2. ARU Algorithm

With the exact probability distribution of the anomalous level, we can obtain its expectation value μ, which is the best indicator of the overall performance of the PMUs over the observation period. Here, we define the history reputation level of a PMU as

$$T = 1 - \mu = \frac{\beta}{\alpha + \beta} \tag{16}$$

While, a dependable reputation system should be able to adaptively adjust the reputation values according to dynamic behavioral changes [37]. Thus, in this paper, we incorporate the history reputation level and the subsequent behavior fluctuations of PMUs to assess their real-time reputation levels. In addition, adaptive parameters are used to allow different impacts due to the reputation levels with different behavior observations. The real-time reputation level of a PMU is then defined as

$$\begin{aligned} T^t &= \omega \cdot T_h + (1 - \omega) \cdot T_u^t \\ &= \omega \cdot \frac{\beta}{\alpha + \beta} + (1 - \omega) \cdot \frac{\lambda_g \cdot N_g^t + 1}{\lambda_g \cdot N_g^t + \lambda_b^t \cdot N_b^t + 1}, \end{aligned} \tag{17}$$

where T_h is the history reputation level of a PMU, and T_u^t is the updating reputation level at time instant t. ω is the weight assigned for the history reputation level to evaluate the importance of history experience to the real-time reputation level, while $1 - \omega$ is for the updating reputation level to evaluate the impacts of recent performance to the real-time reputation level [38]. N_g^t and N_b^t

denote the cumulative number of observations regarding "good" data (not false data) and "bad" data (false data) of a PMU, respectively. Correspondingly, λ_g and λ_b^t are designed as the impact factors for "good" data and "bad" data. It is natural that, from the social perspective, one needs to spend a longer period of time performing successive good behaviors to establish a high reputation level, yet only a few bad behaviors would adversely affect the reputation built over time [39]. As such, we penalize the PMUs when "bad" data are observed. In our algorithm, λ_b^t is designed relatively larger than λ_g, and λ_b^t will be increased if successive "bad" data are observed to amplify the impacts.

Algorithm 2 Adaptive Reputation Updating Algorithm

procedure
  Input: N_g^{t-1}, N_b^{t-1}, λ_g, λ_b^{t-1},S_b^{t-1}, τ
  if the judgement result of current data is "good" **then**
    $N_g^t \leftarrow N_g^{t-1} + 1$;
    $S_b^t \leftarrow 0$;
  else
    $N_b^t \leftarrow N_b^{t-1} + 1$;
    $S_b^t \leftarrow S_b^{t-1} + 1$;
    if $S_b^t > 1$ **then**
      $\lambda_b^t = \lambda_b^{t-1} \cdot e^{\tau}$;
    end if
  end if
  Compute updating reputation level by:
  $T_u^t = \frac{\lambda_g \cdot N_g^t + 1}{\lambda_g \cdot N_g^t + \lambda_b^t \cdot N_b^t + 1}$,
  and the overall reputation level by:
  $T^t = \omega \cdot \frac{\beta}{\alpha+\beta} + (1-\omega) \cdot \frac{\lambda_g \cdot N_g^t + 1}{\lambda_g \cdot N_g^t + \lambda_b^t \cdot N_b^t + 1}$.
  output: T^t.
end procedure

Algorithm 2 presents the ARU procedure, where S_b^t denotes the number of successive observations of "bad" data. They increment by 1 when corresponding behavior occurs. If successive "bad" data is observed, the corresponding impact factor λ_b^t will be increased by $\lambda_b^{t-1} \cdot (e^{\tau} - 1)$, otherwise, the counter for successive "bad" observations S_b^t will be reset to 0 and the impact factor λ_b^t remains unchanged. Here, τ is initialized as a small value (e.g., 0.0001) in our experiments,

and can be adjusted according to different application environments.

With the real-time reputation level of each PMU, it is easy to identify the compromised PMU by testing the following binary hypothesis:

$$\begin{cases} \mathbf{H_0}\text{: PMU } U_j \text{ is compromised,} & \text{if } T_j^t < D_{th} \\ \mathbf{H_1}\text{: PMU } U_j \text{ is not compromised,} & \text{otherwise.} \end{cases} \tag{18}$$

where D_{th} is an acceptable detection threshold. This hypothesis is tested once the reputation level is updated in order to ensure real-time detection.

5. Performance Evaluation

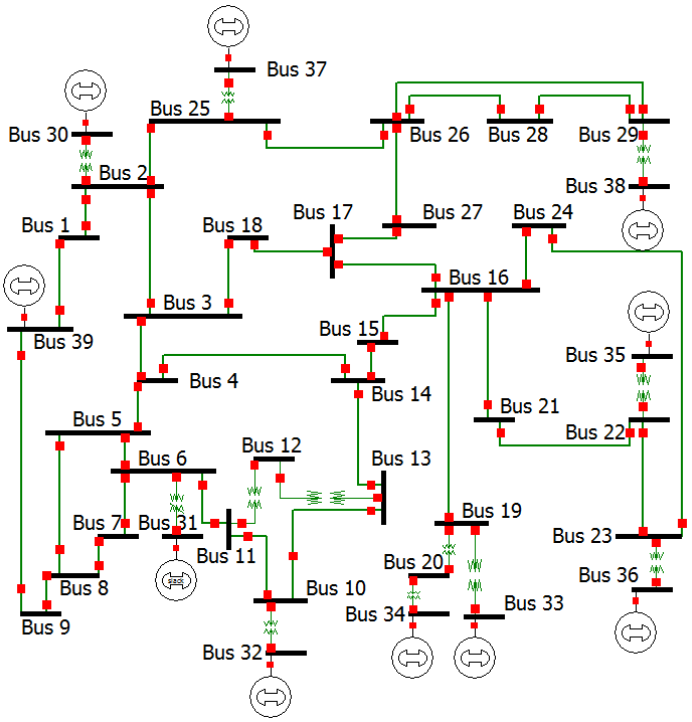


Fig. 6: IEEE 39-bus power system.

In this section, we present a set of simulation experiments and the results to

demonstrate the efficacy of our proposed DHCD method, including the collaborative FDD process and determination of compromised PMU process. Figure 6 shows the IEEE 39-bus power system that is used as a benchmark system in our simulation experiments. IEEE 39-bus power system is a well-known New England power system with 10 generators, 39 buses, and 46 transmission lines, which is commonly used as a benchmark system to test and verify new schemes [1, 9, 40]. Combined with the PowerWorld simulator [41], the power system can provide real-time, accurate and precise state information of the power system. Our experiments are conducted using the PowerWorld simulator on an IEEE standard 39-bus power system, where a number of scenarios are simulated and corresponding real-time measurement data from PMUs are collected. These data are then used to evaluate our proposed DHCD method in MATLAB. The key parameters are summarized in Table 2.

Table 2: Simulation Parameters

Parameter	Default setting
T_h	0.8
ω	0.4
λ_g	0.1
λ_b^0	0.5
S_b	10
τ	0.001
D_{th}	0.6
Number of PMUs: N	39
Number of samples each test: K	1000
State variables that collected	$\delta, V, L_{Mvar}, L_{MW}$

5.1. Efficacy of FDD Algorithm

In this section, we simulate two groups of simulation experiments. The first group shows that only one piece of the four rule specifications is violated (with a single "1" in R_j^t). In contrast, the second group shows that multiple pieces of the four rule specifications are violated (with multiple "1"s in R_j^t). Further, as

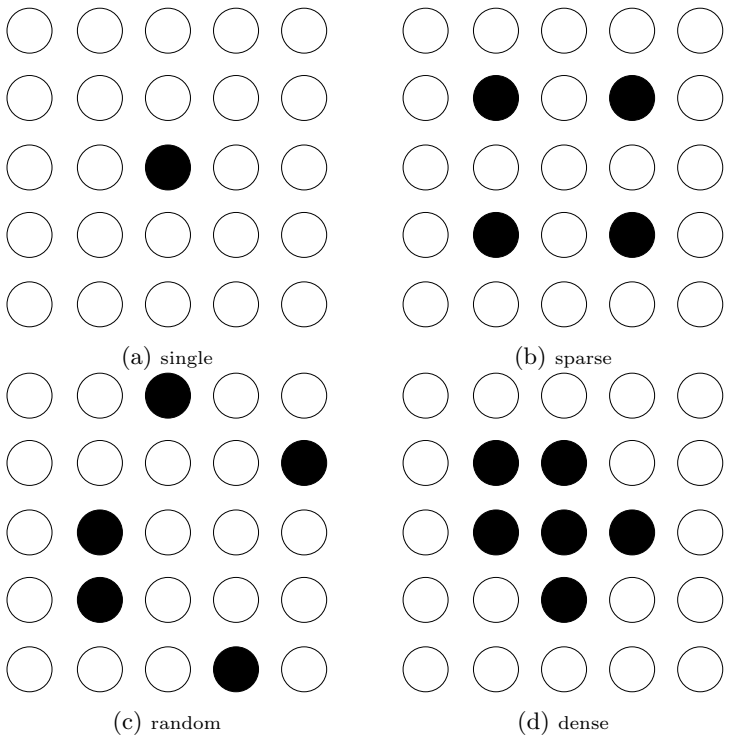


Fig. 7: Four different cases of the distribution of PMUs with inserted false measurement data: single, sparse, random, and dense.

shown in Fig. 7, each group is divided into four different cases: (a) single, (b) sparse, (c) random, and (d) dense, representing four distribution types of false measurement data. To be specific, case (a) describes that only single PMU is inserted with false measurement data; case (b) describes that multiple sparsely distributed PMUs are inserted with false measurement data; case (c) describes that multiple randomly distributed PMUs are inserted with false measurement data; and case (d) describes that multiple densely distributed PMUs are inserted with false measurement data.

Tables 3 and 4 show the simulation results in terms of the detection rate and the average iterations of the FDD algorithm for detecting false measurement data with single violated rule and multiple violated rules, respectively.

We observe from both Tables 3 and 4 that, either singly or sparsely distributed PMU(s) with inserted false measurement data can be easily detected by our FDD algorithm with a 100% detection rate. As for either randomly or densely distributed PMUs with inserted false measurement data, FDD has a high detection rate but not 100%. The reason is that, in most cases, the collaborative FDD performs well for detecting anomalous data when these corresponding PMUs are located near the inner regions of the grid. The anomalies can be identified by starting from the peripheral PMUs at the first iteration to the inner PMUs at the subsequent iterations. While, in some extreme and rare cases, if these anomalous PMUs are concentrated at the marginal regions of the grid, only peripheral PMUs in the vicinity of the inner regions can be identified. After the first or two iterations, the peripheral anomalous PMUs can be identified and their connections to other PMUs removed. Therefore, other anomalous PMUs in marginal regions may be isolated with only anomalous neighbouring PMUs. They can collude with each other to mutually protect each other by showing the same results R_i^t. Such extreme cases may occur in dense distribution type simulation experiments, so the dense type holds relatively lower detection rate in both group one and group two.

Table 3: The detection rate and the average iterations of FDD algorithm with single rule violated false measurement data under four different distribution types. The number of PMUs with false measurement data is 6.

Distribution Type	Detection Rate	Average Iterations
Single	100.0%	1.000
Sparse	100.0%	1.000
Random	97.1%	1.173
Dense	80.4%	2.071

The average iterations for either singly or sparsely distributed PMU(s) with inserted false measurement data in both group one and group two are 1.000, as the inserted anomalous data of these two types can be easily identified by

Table 4: The detection rate and the average iterations of FDD algorithm with multiple rules violated false measurement data under four different distribution types. The number of PMUs with false measurement data is 6.

Distribution Type	Detection Rate	Average Iterations
Single	100.0%	1.000
Sparse	100.0%	1.000
Random	97.9%	1.107
Dense	93.7%	1.520

collaborative detection with only one iteration. In random distribution type, the average iterations are 1.173 and 1.107 for the two groups, respectively. This means that one round FDD can successfully detect the inserted false data, but in some situations, it requires another one to two rounds to detect the false data. Note that, in our simulation experiments, for undetected false data, the number of iterations is set as 5, the upper bound of FDD algorithm. As for the densely distribution type, the average iterations are 2.071 and 1.0520 respectively. This shows that, compared with random distribution type, more cases require additional FDD iterations to detect the inner false data.

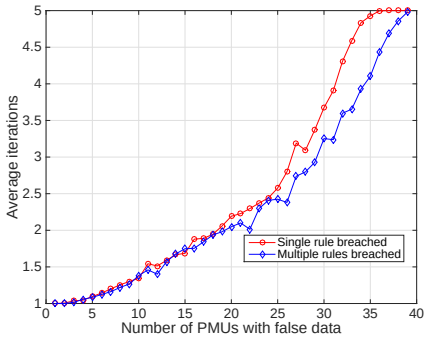


Fig. 8: The average iterations needed for FDD algorithm versus different numbers of PMUs with false measurement data. Two groups of false data: single rule violated and multiple rules violated are compared.

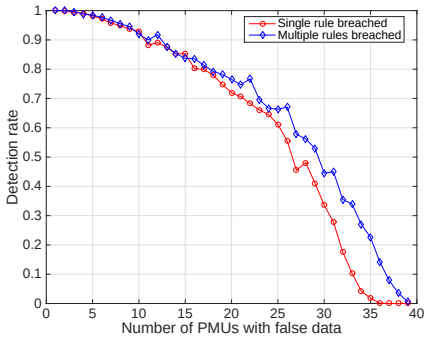


Fig. 9: The detection rate of FDD algorithm versus different numbers of PMUs with false measurement data. Two groups of false data: single rule violated and multiple rules violated are compared.

Interestingly, the simulation results also show that, group two simulations can achieve a higher or equal detection rate with fewer average iterations than

group one. This is because our FDD algorithm detects the false data when at least one rule is violated, so in group two it is much easier for FDD to detect the anomalous data.

In addition to the above results, we studied the relationship between the average iterations and the number of PMUs with false data under random distribution type as shown in Fig. 8, and the corresponding detection rate as well in Fig. 9. Clearly, the value of the average iterations increases, and eventually up to 5, the upper bound, as the increase in the number of PMUs with false data. Correspondingly, the value of the detection rate drops from 1 to 0 while the number of PMUs with false data increases. We also observe similar results in the sense that both values of the average iterations and the detection rate of multiple rules violation outperformed the single rule violated data.

5.2. Identification of Compromised PMUs with Our Reputation System

The performance of our reputation system can be affected by the following critical parameters: (1) ω, the weight assigned for weight assigned for the history reputation level; (2) D_{th}, the detection threshold; (3) λ_b, the impact factor; and (4) S_b^t, the number of successive observations of bad data.

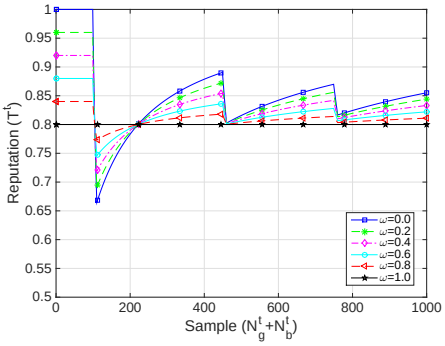


Fig. 10: The reputation level of a PMU under different ωs ($T_h = 0.8, D_{th} = 0.6, S_b = 10, \lambda_b^0 = 0.5$).

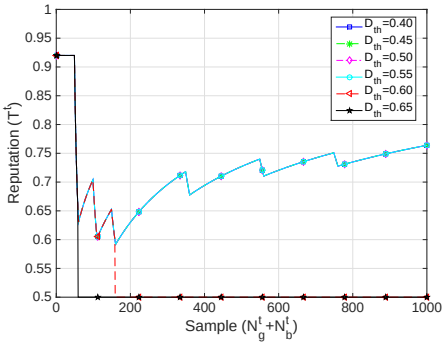


Fig. 11: The reputation level of a PMU under different under different D_{th}s ($T_h = 0.8, \omega = 0.4, S_b = 10, \lambda_b^0 = 0.5$).

Figure 10 shows the fluctuations of a PMU's reputation level under different ωs. Three FDI events, each lasting 10 samples, are inserted into the PMU's measurement data. This figure shows that, the higher the ω is, the more the current reputation level T^t relies on its history value T_h. Particularly, $\omega = 0.0$ indicates that $T^t = T_h$, and $\omega = 1.0$ indicates that $T^t = T_u^t$.

Figure 11 shows the fluctuations of a PMU's reputation level under different D_{th}s. Six FDI events, each lasting 10 samples, are inserted into the PMU's measurement data. We observe from this figure that a higher D_{th}s hold a lower tolerance to PMUs' "bad" behaviors, while lower D_{th}s have higher tolerance to PMUs' "bad" behaviors. In other words, higher D_{th}s are more sensitive than lower D_{th}s. For example, when $D_{th} = 0.65$, our reputation system raises an alarm when the first FDI event is inserted.

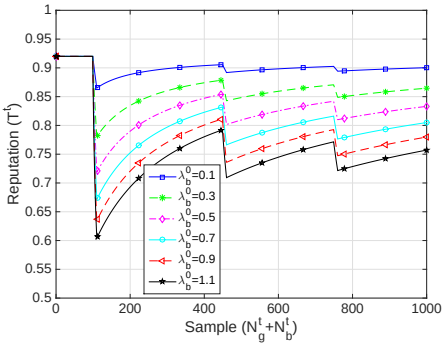


Fig. 12: The reputation level of a PMU under different λ_b^0s ($T_h = 0.8, \omega = 0.4, D_{th} = 0.6, S_b = 10$).

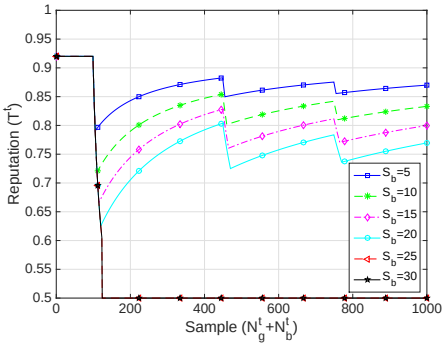


Fig. 13: The reputation level of a PMU under different S_b ($T_h = 0.8, \omega = 0.4, D_{th} = 0.6, \lambda_b^0 = 0.5$).

The relationship between the reputation level and the λ_b^0 is plotted in Fig. 12. Three FDI events, each lasting 10 samples, are inserted into the PMU's measurement data. Clearly, the higher the λ_b^0, the more adverse the consequence of penalty to the reputation level, which means that the reputation level decreases significantly.

A similar relationship between the reputation level and the S_b is plotted

in Fig. 13. Also, three FDI events but different lengths are inserted into the PMU's measurement data. Similar to Fig. 12, this figure shows that the larger the S_b, the more significance the penalty has on the reputation level, as large S_b results in more times of λ_b^t adjustment, i.e., $\lambda_b^t = \lambda_b^{t-1} * e^{\tau}$. For instance, with $D_{th} = 0.6$, the reputation level drops quickly below D_{th} if $S_b = 30$.

6. Conclusions

In this paper, we proposed a novel DHCD method to identify and mitigate FDI attacks in smart grid CPS. Specifically, a rule specification based real-time collaborative detection system was designed to identify the anomalies of measurement data. In addition, a new reputation system with an ARU algorithm was presented to evaluate the overall running status of the PMUs, which can be used to identify compromised PMUs. We then demonstrated the utility of the proposed approach using simulations of the IEEE 39-bus power system.

As previously discussed, our method is designed to detect the malicious activities resulting in the anomaly of measurement data. Future work would include extending the proposed approach to capture power system faults (e.g., voltage disturbance, open circuit, and short circuit).

References

[1] B. Li, R. Lu, W. Wang, K.-K. R. Choo, DDOA: A Dirichlet-based detection scheme for opportunistic attacks in smart grid cyber-physical system, IEEE Trans. Inf. Forensics Security 11 (11) (2016) 2415–2425.

[2] L. Xie, Y. L. Mo, B. Sinopoli, False data injection attacks in electricity markets, in: Proc. First IEEE International Conference on Smart Grid Communications (SmartGridComm), 2010, pp. 226–231.

[3] K.-K. R. Choo, A conceptual interdisciplinary plug-and-play cyber security framework, in: ICTs and the Millennium Development Goals, Springer, 2014, pp. 81–99.

[4] R. Jiang, R. Lu, Y. Wang, J. Luo, C. Shen, X. S. Shen, Energy-theft detection issues for advanced metering infrastructure in smart grid, Tsinghua Sci. Technol. 19 (2) (2014) 105–120.

[5] X. Li, X. Liang, R. Lu, X. Shen, X. Lin, H. Zhu, Securing smart grid: cyber attacks, countermeasures, and challenges, IEEE Commun. Mag. 50 (8) (2012) 38–45.

[6] N. Falliere, L. O. Murchu, E. Chien, W32. Stuxnet dossier, White paper, Symantec Corp., Security Response 5 (2011) 6.

[7] S. Shin, G. Gu, Conficker and beyond: A large-scale empirical study, in: Proc. 26th Annual Computer Security Applications Conference (ACSAC), 2010, pp. 151–160.

[8] S. Gorman, Y. J. Dreazen, A. Cole, Insurgents hack US drones, Wall Street Journal 17 (2009) 1–4.

[9] H. Bao, R. Lu, B. Li, R. Deng, BLITHE: Behavior rule based insider threat detection for smart grid, IEEE Internet Things J. 3 (2) (2016) 190–205.

[10] H. Fang, L. Xu, K.-K. R. Choo, Stackelberg game based relay selection for physical layer security and energy efficiency enhancement in cognitive radio networks, Appl. Math. Comput. 296 (2017) 153–167.

[11] J. Chen, L. Shi, P. Cheng, H. Zhang, Optimal denial-of-service attack scheduling with energy constraint, IEEE Trans. Autom. Control 60 (11) (2015) 3023–3028.

[12] H. M. Merrill, F. C. Schweppe, Bad data suppression in power system static state estimation, IEEE Trans. Power App. Syst. (6) (1971) 2718–2725.

[13] J. Chen, A. Abur, Placement of PMUs to enable bad data detection in state estimation, IEEE Trans. Power Syst. 21 (4) (2006) 1608–1615.

[14] W. W. Kotiuga, M. Vidyasagar, Bad data rejection properties of weughted least absolute value techniques applied to static state estimation, IEEE Trans. Power App. Syst. (4) (1982) 844–853.

[15] Y. Liu, P. Ning, M. K. Reiter, False data injection attacks against state estimation in electric power grids, ACM T. Inform. Syst. Se. (TISSEC) 14 (1) (2011) 13.

[16] M. E. Baran, A. W. Kelley, State estimation for real-time monitoring of distribution systems, IEEE Trans. Power App. Syst. 9 (3) (1994) 1601–1609.

[17] M. M. Nordman, M. Lehtonen, Distributed agent-based state estimation for electrical distribution networks, IEEE Trans. Power App. Syst. 20 (2) (2005) 652–658.

[18] E. Handschin, F. Schweppe, J. Kohlas, A. Fiechter, Bad data analysis for power system state estimation, IEEE Trans. Power App. Syst. 94 (2) (1975) 329–337.

[19] J. Peng, K.-K. R. Choo, H. Ashman, User profiling in intrusion detection: A review, J. Netw. Comput. Appl. 72 (2016) 14–27.

[20] S. Iqbal, M. L. M. Kiah, B. Dhaghighi, M. Hussain, S. Khan, M. K. Khan, K.-K. R. Choo, On cloud security attacks: A taxonomy and intrusion detection and prevention as a service, J. Netw. Comput. Appl. 74 (2016) 98–120.

[21] H. Kumarage, I. Khalil, Z. Tari, A. Zomaya, Distributed anomaly detection for industrial wireless sensor networks based on fuzzy data modelling, J. Parallel Distrib. Comput. 73 (6) (2013) 790–806.

[22] S. Rajasegarar, C. Leckie, M. Palaniswami, Hyperspherical cluster based distributed anomaly detection in wireless sensor networks, J. Parallel Distrib. Comput. 74 (1) (2014) 1833–1847.

[23] M. Xie, S. Han, B. Tian, S. Parvin, Anomaly detection in wireless sensor networks: A survey, J. Netw. Comput. Appl. 34 (4) (2011) 1302–1325.

[24] H. Otrok, N. Mohammed, L. Wang, M. Debbabi, P. Bhattacharya, A game-theoretic intrusion detection model for mobile ad hoc networks, Comput. Commun. 31 (4) (2008) 708–721.

[25] F. C. Schweppe, J. Wildes, D. B. Rom, Power system static-state estimation, parts I, II, and III, IEEE Trans. Power App. Syst. 89 (1) (1970) 120–135.

[26] T. V. Cutsem, M. Ribbens-Pavell, L. Mili, Hypothesis testing identification: A new method for bad data analysis in power system state estimation, IEEE Trans. Power App. Syst. (11) (1984) 3239–3252.

[27] F. Pasqualetti, F. Drfler, F. Bullo, Cyber-physical attacks in power networks: Models, fundamental limitations and monitor design, in: Proc. 50th IEEE Conference on Decision and Control and European Control Conference, IEEE, 2011, pp. 2195–2201.

[28] R. B. Bobba, K. M. Rogers, Q. Wang, H. Khurana, K. Nahrstedt, T. J. Overbye, Detecting false data injection attacks on DC state estimation, in: Preprints of the First Workshop on Secure Control Systems, CPSWEEK, Vol. 2010, 2010.

[29] W. Li, Risk evaluation of wide area measurement and control system, in: Risk assessment of power systems: models, methods, and applications, John Wiley & Sons, 2014, pp. 313–350.

[30] M. Qiu, W. Gao, M. Chen, J.-W. Niu, L. Zhang, Energy efficient security algorithm for power grid wide area monitoring system, IEEE Trans. on Smart Grid 2 (4) (2011) 715–723.

[31] M. Qiu, H. Su, M. Chen, Z. Ming, L. T. Yang, Balance of security strength and energy for a PMU monitoring system in smart grid, IEEE Commun. Mag. 50 (5) (2012) 142–149.

[32] A. Castiglione, R. Pizzolante, C. Esposito, A. De Santis, F. Palmieri, A. Castiglione, A collaborative clinical analysis service based on theory of evidence, fuzzy linguistic sets and prospect theory and its application to craniofacial disorders in infants, Future Gener. Comput. Syst. http://dx.doi.org/10.1016/j.future.2016.08.001.

[33] A. Patel, H. Alhussian, J. M. Pedersen, B. Bounabat, J. C. Jnior, S. Katsikas, A nifty collaborative intrusion detection and prevention architecture for smart grid ecosystems, Computers & Security http://dx.doi.org/10.1016/j.cose.2016.07.002.

[34] C. Esposito, A. Castiglione, F. Palmieri, M. Ficco, Trust management for distributed heterogeneous systems by using linguistic term sets and hierarchies, aggregation operators and mechanism design, Future Gener. Comput. Syst. http://dx.doi.org/10.1016/j.future.2015.12.004.

[35] U. S. Premarathne, I. Khalil, M. Atiquzzaman, Trust based reliable transmissions strategies for smart home energy consumption management in cognitive radio based smart grid, Ad Hoc Netw. 41 (2016) 15–29.

[36] K. Bowman, L. Shenton, Parameter estimation for the Beta distribution, J. Stat. Comput. Simul. 43 (3-4) (1992) 217–228.

[37] M. Srivatsa, L. Xiong, L. Liu, Trustguard: countering vulnerabilities in reputation management for decentralized overlay networks, in: Proc. 14th International Conference on World Wide Web (WWW), ACM, 2005, pp. 422–431.

[38] F. G. Mrmol, G. M. Prez, TRIP, a trust and reputation infrastructure-based proposal for vehicular ad hoc networks, J. Netw. Comput. Appl. 35 (3) (2012) 934–941.

[39] Y. L. Sun, Z. Han, W. Yu, K. J. R. Liu, A trust evaluation framework in distributed networks: Vulnerability analysis and defense against attacks, in: Proc. IEEE INFOCOM, 2006, pp. 1–13.

[40] D. Zhang, S. Li, P. Zeng, C. Zang, Optimal microgrid control and power-flow study with different bidding policies by using powerworld simulator, IEEE Trans. Sustainable Energy 5 (1) (2014) 282–292.

[41] Y. Brar, J. S. Randhawa, Optimal power flow using power world simulator, in: Proc. IEEE Electric Power and Energy Conference (EPEC), IEEE, 2010, pp. 1–6.