sensors

MDPI

Article

A Novel Diagnosis Scheme against Collusive False Data Injection Attack

Jiamin Hu [1], Xiaofan Yang [1,*] and Luxing Yang [2]

[1] Department of School of Big Data & Software Engineering, Chongqing University, Chongqing 400044, China; jiaminhu@cqu.edu.cn
[2] College of Information Technology, Deakin University, Melbourne, VIC 3125, Australia; y.luxing@deakin.edu.au
* Correspondence: xfyang1964@cqu.edu.cn

Abstract: The collusive false data injection attack (CFDIA) is a false data injection attack (FIDA), in which false data are injected in a coordinated manner into some adjacent pairs of captured nodes of an attacked wireless sensor network (WSN). As a result, the defense of WSN against a CFDIA is much more difficult than defense against ordinary FDIA. This paper is devoted to identifying the compromised sensors of a well-behaved WSN under a CFDIA. By establishing a model for predicting the reading of a sensor and employing the principal component analysis (PCA) technique, we establish a criterion for judging whether an adjacent pair of sensors are consistent in terms of their readings. Inspired by the system-level fault diagnosis, we introduce a set of watchdogs into a WSN as comparators between adjacent pairs of sensors of the WSN, and we propose an algorithm for diagnosing the WSN based on the collection of the consistency outcomes. Simulation results show that the proposed diagnosis scheme achieves a higher probability of correct diagnosis.

Keywords: wireless sensor network; collusive false data injection attack; diagnosis scheme; watchdog; autoregressive moving average model; principal component analysis; diagnosis algorithm

Citation: Hu, J.; Yang, X.; Yang, L. A Novel Diagnosis Scheme against Collusive False Data Injection Attack. *Sensors* **2023**, *23*, 5943. https://doi.org/10.3390/s23135943

Academic Editor: Alessandra Rizzardi

Received: 16 May 2023
Revised: 21 June 2023
Accepted: 25 June 2023
Published: 26 June 2023

1. Introduction

Wireless sensor networks (WSNs) are networks of wirelessly interconnected sensor nodes that collect data about the surrounding environment [1]. With the rapid popularization of Internet of Things (IoT) applications, WSNs have penetrated nearly all aspects of human life, ranging from industry and transportation to healthcare and military affairs [2]. Typically, sensors have limited energy, limited memory storage, and limited computing/communication capabilities, and are deployed in unattended and abominable environments. As a result, WSNs are vulnerable to a variety of cyber attacks. Consequently, the security of WSNs has received considerable attention from the network security community [3–6].

1.1. Problem Formulation

False data injection attacks (FDIAs) are cyber attacks on WSNs where false data are stealthily injected into the physically captured sensors [7,8]. FDIAs would render the compromised sensors to deliver wrong data to the base station. As a result, the decision-maker at the base station would make an incorrect decision, leading to serious consequences.

An attacker can obtain key information from a compromised sensor to gain control over it, which leads to a chance that proactive security mechanisms are useless in detecting FDIAs. Therefore, the best way to counteract FDIAs is detection by analyzing the measurements themselves. The spatiotemporal correlation of inter-measurements is a solution used in many studies to detect FDIAs [9–11]. Due to the continuity of physical phenomena, the measurements of each sensor are temporally correlative in time. Due to the high-density network topology of WSNs, the inter-measurements of adjacent sensors are

spatially correlative. When compromised and genuine sensors coexist, the inconsistency of measurements will lead to correlation failure.

However, attackers aim to minimize the risk of being detected by employing resourceful and sophisticated strategies. Most previous works on FDIA detection have been focused on situations where independent false data are injected into different captured sensors [10,12–16]. In this paper, we consider FDIAs in which the readings of some adjacent pairs of compromised sensors are modified in a coordinated manner so that the false readings still look spatially–temporally correlated. We refer to such FDIAs as collusive FDIAs or simply *CFDIAs*. As a result, the conventional methods for detecting an ordinary FDIA fail when used to detect a CFDIA. Consequently, it is crucial to investigate the following problem:

CFDIA diagnosis problem: Identify the compromised sensor nodes under a CFDIA in a WSN with no natural anomalies.

Inspired by the system-level fault diagnosis, a solution based on hybrid detection will be created for the CFDIA diagnosis problem. To our knowledge, this is the first time such an attempt has been made.

1.2. Main Contributions

Our main contributions are sketched below.

- We define a new kind of false data injection attack to WSNs, i.e., a conclusive false data injection attack (CFDIA), and we propose a new problem (i.e., the CFDIA diagnosis problem) aiming to identify the compromised sensors in a WSN under a CFDIA.
- We establish an autoregressive moving average (ARMA) model for predicting the current reading of a sensor using its historical readings. Based on the prediction model and by employing the principal component analysis (PCA) technique, we establish a model for judging if an adjacent pair of sensors are consistent in terms of their readings.
- Inspired by the system-level fault diagnosis, we introduce a set of watchdogs in the WSN under CFDIA as comparators between adjacent pairs of sensors within their respective communication range. These watchdogs deliver their respective collections of consistency outcomes to the base station. The base station collects all the received consistency outcomes to form a complete syndrome.
- We design an algorithm for identifying the abnormal sensors based on the complete syndrome. Through extensive simulation experiments, we conclude that the diagnosis algorithm achieves a higher probability of correct diagnosis.

The subsequent materials are organized in this fashion: In Section 2, the related works are reviewed. In Section 3, some terms, notations, and assumptions are introduced. The diagnosis scheme is described in detail in Section 4, and the effectiveness of the diagnosis scheme is corroborated through simulation experiments in Section 5. Finally, this work is summarized by Section 6.

2. Related Work

In this section, we make comments on the previous work that is related to the present paper, aiming to highlight the novelty of our work.

2.1. System-Level Fault Diagnosis

The system-level fault diagnosis aims to identify faulty units in a computer system based on a collection of test/comparison outcomes between adjacent units [17,18]. There are two different system-level diagnosis approaches: the test-based diagnosis and comparison-based diagnosis [17].

In the test-based diagnosis, for each adjacent pair of units, one unit serves as the tester and the other as the testee; the tester assigns a computational job to the testee, the testee performs the computation and returns the computational result to the tester, and the tester compares the result with its own result and judges the testee to be fault-free or faulty

according to whether the two results are identical or not. Finally, a diagnosis algorithm is performed based on the collection of test outcomes [17,19].

In the comparison-based diagnosis, each adjacent pair of units perform the same computation, and the computational results are compared. That the two results are identical implies that the two units are either both fault-free or faulty. On the contrary, that the two results are not identical implies that at least one of the two units is faulty. Finally, a diagnosis algorithm is performed based on the collection of comparison outcomes [18,20].

The system-level fault diagnosis has been applied to the detection of faulty sensors in WSNs [21–24]. Inspired by the system-level fault diagnosis, an algorithm for diagnosing the abnormal sensors in the WSN under a CFDIA based on the collection of the consistency outcomes is proposed without considering the existence of natural anomalies, and validates the effectiveness of the diagnosis algorithm through simulation experiments. To our knowledge, however, the system-level fault diagnosis has not been applied to the identification of compromised sensor nodes in WSNs under an FDIA, let alone the identification of captured sensors under a CFDIA.

2.2. FDIAs Detection of WSNs

FDIAs are known for their severe impact and are one of the widely studied cyber-attacks in smart grids, power systems, and WSNs [10,14]. In the research area of the FDIAs detection problem of WSNs, many studies focus on exploring techniques from the sensor measurements.

Ref. [25] proposed a generalized distributed anomaly detection scheme based on the spatio-temporal correlation of physical processes against FDIAs in WSNs. Ref. [26] presented a method combining a measurement check and authentication strategies to detect FDIAs in WSNs. Ref. [27] exploited the sensor data correlation in time and space to identify the falsified data in the industrial Internet-of-Things. Ref. [28] utilized the observed spatio-temporal and multivariate-attribute sensor correlations to detect FDIAs in WSNs. Ref. [29] addressed the issue of detecting FDIAs based on spatial correlation to dynamic WSNs. Ref. [30] suggested a method using temporal, spatial, and event-based correlations to prevent FDIAs in WSNs.

Few works cover detecting CFDIAs in WSNs. Ref. [31] revealed that wireless nodes usually have some correlation patterns in communication metrics, which can be used to defend against CFDIAs in WSNs. Ref. [32] proposed a wavelet transform method based on wavelet transform to detect CFDIAs. Ref. [33] exploited the spatio-temporal correlation of heterogeneous sensor data to detect CFDIAs in low-density WSNs. Since the detection of CFDIAs requires comparing measurements over a broader set of sensors, these efforts are based on a centralized detection scheme to provide a global view. However, a compromised node not only injects false data into itself, but it may also inject false data into routed packets, potentially leading to a higher false positive rate.

In this paper, we propose a hybrid detection scheme. We define watchdog as a kind of expensive dedicated hardware device that is deployed in the area monitored by the WSN in concern, is within the communication range of the base station serving the WSN, and can acquire the readings of the sensors that are within the communication range of the device. A set of watchdogs are deployed in the area for judging the consistency of adjacent pairs of sensors in terms of the spatial–temporal correlation of their readings. All the consistency outcomes are delivered directly to the base station. To our knowledge, a hybrid detection scheme has not yet been applied to the CFDIA diagnosis problem.

3. Preliminary Knowledge

3.1. WSNs with Watchdogs

Consider a WSN used for gathering a kind of environmental data within a specific area. Let R_1 denote the communication radius of the base station serving the WSN. Let R_2 denote the common communication radius of the sensors in the WSN. Let the undirected graph $G = (V, E)$ denote the topological structure of the WSN, i.e., $V = \{v_1, \cdots, v_N\}$

stands for the set of sensors in the WSN, and $\{v_i, v_j\} \in E$ if and only if v_i is within the communication range of v_j. Let $r_i(t)$ denote the measurement reading of the node v_i at time t.

Define watchdog as a kind of expensive dedicated hardware device that is deployed in the area monitored by the WSN, is within the communication range of the base station, and can acquire the readings of the sensors that are within the communication range of the device. Suppose a set of watchdogs are deployed within the monitored area. Let R_3 denote the common communication radius of the watchdogs. Suppose $R_2 < R_3 < R_1$. Let $W = \{w_1, \cdots, w_M\}$ denote the set of the watchdogs. Suppose the watchdogs are used for periodically acquiring the measurement readings of the sensors within their respective communication ranges. Let $S = \{1, 2, \cdots, K\}$ denote the set of time points at which the readings of the sensors are acquired by their respective neighboring watchdogs.

3.2. Collusive False Data Injection Attack

A collusive false data injection attack (CFDIA) is a false data injection attack in which the readings of some adjacent pairs of compromised sensors are modified in a coordinated manner so that the changed readings are still spatially–temporally correlated. Owing to the attacker's limited budget, assume (i) the compromised nodes are less than normal nodes, and (ii) the compromised sensors are concentrated in a small area. Figure 1 illustrates a CFDIA to a toy WSN of seven nodes.

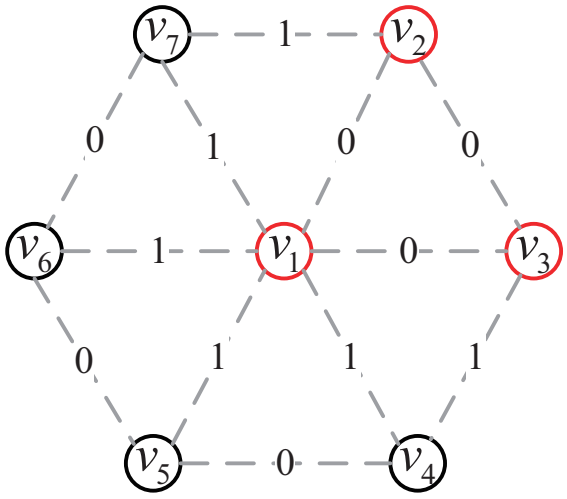

Figure 1. A CFDIA against a toy WSN, where each black circle (resp. red circle) represents a normal node (resp. compromised node), each edge represents the two associated nodes which are within each other's communication range, and "0" (resp. "1") represents that there is (resp. there is no) a spatial–temporal correlation between the two associated nodes.

3.3. Autoregressive Moving Average Models

An autoregressive model builds on the assumption that there is a linear relationship between the current value of a variable and its own historical values. A moving average model assumes that the current value of a variable depends not only on the current information but also on previous information. The model obtained by combining an autoregressive model with a moving average model is referred to as an autoregressive moving average (ARMA) model [34]. An ARMA model of order (p, q) is formulated as follows.

$$r(t) = \mu + \sum_{l=1}^{p} \phi_l r(t-l) + \sum_{l=1}^{q} \psi_l \epsilon(t-l) + \epsilon(t). \tag{1}$$

Here, $r(t)$ stands for the value of the variable r at time t; μ, ϕ_l, and ψ_l are model parameters; and $\epsilon(t)$ stands for the value of the independent error at time t, which follows a Gaussian distribution with zero mean.

3.4. Principal Component Analysis

The principal component analysis (PCA) is a commonly used technique for reducing the dimensionality of large datasets and increasing data interpretability. The PCA creates new uncorrelated variables (the principal components) by solving an eigenvalue/eigenvector problem [35]. The PCA has been applied to FDIA detection [36,37].

4. A Diagnosis Scheme against CFDIA

In this section, we propose a diagnosis scheme against a CFDIA. The scheme consists of two phases: the syndrome generation phase and the CFDIA diagnosis phase, which are stated as follows.

- *Phase I: Syndrome generation*. In this phase, each watchdog collects a set of readings of the sensors monitored by the watchdog and conducts a spatio-temporal correlation analysis between each adjacent pair of sensors, forming a partial syndrome. All the watchdogs deliver their own partial syndromes directly to the base station. A (complete) syndrome is generated by merging the partial syndromes.
- *Phase II: CFDIA Diagnosis*. Taking the syndrome as input, perform an algorithm for diagnosing a CFDIA. As a result, the compromised nodes are identified.

Next, let us discuss the two phases in detail.

4.1. Syndrome Generation

4.1.1. Consistency Criterion

The syndrome on a WSN under a CFDIA refers to the collection of the consistency outcomes between adjacent pairs of nodes of the WSN. The syndrome is the basis of CFDIA diagnosis. The key to generating the syndrome is to establish a consistency criterion. For this purpose, we need to discuss temporal correlation and spatial correlation between adjacent pairs of nodes, respectively.

First, there is a temporal correlation of each node in terms of their readings. Let $\widetilde{r}_i(t)$ denote the predicted value of $r_i(t)$. We assume that for $i = 1, \cdots, n$ and for all t, $\widetilde{r}_i(t)$ obeys the following ARMA model of order (p, q):

$$\widetilde{r}_i(t) = \mu_i + \sum_{l=1}^{p} \phi_l^i r_i(t-l) + \sum_{l=1}^{q} \psi_l^i \epsilon_i(t-l) + \epsilon_i(t). \tag{2}$$

The parameters in the model can be estimated using the historical data.

Remark 1. *ARMA is used to model linear relationships, which is suitable for stationary stochastic processes. However, the presence of seasonality and trends in the time-series sensor readings may introduce nonlinear non-stationary sequences. Therefore, the time-series sensor readings need to be pre-processed before extracting the spatio-temporal correlation of adjacent nodes. There is a need for stationary identification. If the time-series sensor readings are not stationary, we can employ the differencing method on the time-series sensor readings to remove seasonality and trends.*

Second, there is a spatial correlation between each adjacent pair of nodes in terms of their readings. We use the PCA to reveal the spatial correlation. Let $\bar{r}_i$(resp. $\bar{r}_j$) denote the mean of historical readings of v_i(resp. v_j). The correlation coefficient of the predicted readings of an adjacent pair of nodes, v_i and v_j, is calculated as follows.

$$\delta_{ij} = \frac{\sum_{t=1}^{K} (\widetilde{r}_i(t) - \bar{r}_i)(\widetilde{r}_j(t) - \bar{r}_j)}{K-1}. \tag{3}$$

The covariance matrix of the predicted readings of v_i and v_j reads

$$\Lambda_{ij} = \begin{bmatrix} \delta_{ii} & \delta_{ij} \\ \delta_{ji} & \delta_{jj} \end{bmatrix}. \tag{4}$$

Let λ_{ij}^1 (resp. λ_{ij}^2) denote the largest (resp. second largest) eigenvalue of the matrix Λ_{ij}. Let $\vec{\mu}_{ij}^1$ (resp. $\vec{\mu}_{ij}^2$) denote the unit eigenvector of Λ_{ij} associated with λ_{ij}^1 (resp. λ_{ij}^2).

$$\Lambda_{ij}\vec{\mu}_{ij}^1 = \lambda_{ij}^1\vec{\mu}_{ij}^1, \quad \Lambda_{ij}\vec{\mu}_{ij}^2 = \lambda_{ij}^2\vec{\mu}_{ij}^2. \tag{5}$$

Assume $\vec{\mu}_{ij}^1$ and $\vec{\mu}_{ij}^2$ are linearly independent. Then, $\vec{\mu}_{ij}^1$ is orthogonal to $\vec{\mu}_{ij}^2$.

The consistency ellipse at time t with the confidence degree $1-\theta$ (here, $\theta \leq 0.1$), denoted as $\Gamma_{ij}^{1-\theta}(t)$, can be calculated by taking $(\widetilde{r}_i(t), \widetilde{r}_j(t))$ as its center, taking $\vec{\mu}_{ij}^1$ (resp. $\vec{\mu}_{ij}^2$) as the direction of its major axis (resp. minor axis), taking λ_{ij}^1 (resp. λ_{ij}^2) as the ratio of its long radius (resp. its short radius), and choosing the confidence degree $1-\theta$. Let $\bar{\Gamma}_{ij}^{1-\theta}(t)$ denote the closed region surrounded by $\Gamma_{ij}^{1-\theta}(t)$. Then, $\bar{\Gamma}_{ij}^{1-\theta}(t)$ is the confidence region with the confidence degree $1-\theta$. Consequently, we present the following:

Consistency criterion: An adjacent pair of nodes, v_i and v_j, are consistent at time t with the confidence degree $1-\theta$ if $(\widetilde{r}_i(t), \widetilde{r}_j(t)) \in \bar{\Gamma}_{ij}^{1-\theta}(t)$. Otherwise, they are inconsistent with the confidence degree $1-\theta$.

See Figure 2a for a schematic explanation of the consistency criterion. For brevity, we remove "at time t" and "with confidence degree $1-\theta$" in the criterion.

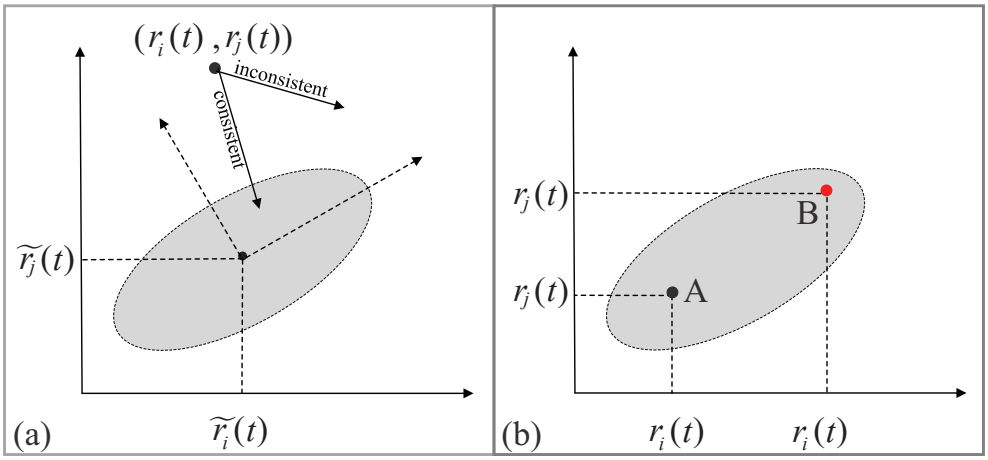

Figure 2. (**a**) The confidence region $\bar{\Gamma}_{ij}^{1-\theta}(t)$ with the confidence degree $1-\theta$. (**b**) A glance of a CFDIA, where the point A represents the values of $r_i(t)$ and $r_j(t)$, and the point B represents the false values of $r_i(t)$ and $r_j(t)$.

4.1.2. Syndrome and Partial Syndrome

Let $\sigma(u,v) = 0$ (resp. 1) denote that the adjacent pair of nodes, u and v, are consistent (resp. inconsistent). We refer to the collection

$$\sigma = \{\sigma(u,v) : \{u,v\} \in E\} \tag{6}$$

as the *syndrome* on the WSN. For each adjacent pair of nodes, u and v, we make the following reasonable assumptions:

1. If u and v are both normal, then $\sigma(u,v) = 0$ with probability $1-\theta$.
2. If one of u and v is normal and the other is abnormal, then $\sigma(u,v) = 1$ with probability $1-\theta$.
3. If u and v are both abnormal, then $\sigma(u,v) = 0$ or 1.

For each watchdog w_m, let V_m denote the set of nodes that are monitored by w_m. We refer to the collection

$$\sigma_m = \{\sigma(u,v) : u,v \in V_m, \{u,v\} \in E\} \tag{7}$$

as the *partial syndrome* associated with w_m.

Each watchdog can acquire the partial syndrome associated with it. All partial syndromes can be delivered by the watchdogs to the base station. Finally, the syndrome can be generated at the base station by merging the received partial syndromes.

4.2. CFDIA Diagnosis

We intend to identify all the abnormal nodes of a WSN by interpreting the syndrome. For this purpose, we need to introduce the following terms and notations.

Definition 1. *Let p be the a priori probability of a node of WSN being compromised.*

Definition 2. *Let σ be a syndrome on the WSN $G = (V, E)$, $e \in E$. The edge e is referred to as a 0-edge or a 1-edge according to $\sigma(e)$ = 0 or 1.*

Definition 3. *Let σ be a syndrome on the WSN $G = (V, E)$. The 0-subgraph of G is defined as a subgraph of G, denoted $G_0 = (V, E_0)$, such that E_0 is the set of 0-edges of G.*

Definition 4. *Let σ be a syndrome on the WSN $G = (V, E)$, $G_0 = (V, E_0)$ the 0-subgraph of G. Let $\mathcal{H} = \{H_i = (U_i, E_i) : 1 \leq i \leq r\}$ be the collection of connected components of G_0. The 1-condensed graph of G is defined as a graph $G^* = (U^*, E^*)$ such that (i) $U^* = \{u_1^*, \cdots, u_r^*\}$, (ii) $\{u_i^*, u_j^*\} \in E^*$ if and only if there is a 1-edge of G that connects a node in U_i with a node in U_j.*

Based on the previously introduced assumptions about the relationship between the states of two adjacent nodes and their consistency, we have the following results.

Theorem 1. *Let σ be a syndrome on the WSN $G = (V, E)$, $e = \{u, v\} \in E$. If $p << 0.5$, the following assertions hold true:*

1. *$\sigma(u, v) = 0$ implies u and v are either both normal with a higher probability (w.h.p.) or both abnormal w.h.p.*
2. *$\sigma(u, v) = 1$ implies at least one of u and v is abnormal w.h.p.*

The following theorem is a direct consequence of this theorem.

Theorem 2. *Let σ be a syndrome on the WSN $G = (V, E)$. Let $G_0 = (V, E_0)$ be the 0-subgraph of G. Let $\mathcal{H} = \{H_i = (U_i, E_i) : 1 \leq i \leq r\}$ be the set of connected components of G_0. If $p << 0.5$, the following assertions hold true:*

1. *For $1 \leq i \leq r$, either (i) the nodes in U_i are all normal w.h.p., or (ii) the nodes in U_i are all abnormal w.h.p..*
2. *If there is a 1-edge connecting U_i with U_j, then either (i) the nodes in U_i are all normal and the nodes in U_j are all abnormal w.h.p, or (ii) the nodes in U_j are all abnormal and the nodes in U_j are all normal w.h.p.*

Based on the theorem, we present an algorithm (i.e., the CFDIA algorithm given in Algorithm 1) for identifying the abnormal nodes in a WSN under a CFDIA. The correctness of the algorithm is guaranteed by the following observation.

Algorithm 1 CFDIA-DIAG.

Input: A WSN $G = (V, E)$ under CFDIA, a syndrome σ on G.
Output: A subset $V_a \subseteq V$ that is diagnosed to be the set of abnormal nodes.

Find the 0-subgraph of G, denoted $G_0 = (V, E_0)$;
Find all the connected components of G_0, denoted $H_i = (U_i, E_i)$, $i = 1, \cdots, r$;
Find the 1-condensed graph of G, denoted $G^* = (U^*, E^*)$, where $U^* = \{u_1^*, \cdots, u_r^*\}$;
// label all nodes of G^* with Z; //
for each $u^* \in U^*$ **do**
    $l(u^*) \leftarrow Z$;
// label all nodes of G^* with X or Y through depth-first search; //
Let Q be an empty queue;
Choose a node u^* with maximum degree from U^*;
$l(u^*) \leftarrow X$; $Q \leftarrow Q + u^*$;
while G^* has a node with label Z **do**
    fetch node u^* with the most nodes from Q; $Q \leftarrow Q - u^*$;
    for each $v^* \in U^*$ that is adjacent to u^* **do**
        if $l(v^*) = Z$ **then**
            $Q \leftarrow Q + v^*$;
            if $l(u^*) = X$ **then**
                $l(v^*) \leftarrow Y$;
            else
                $l(v^*) \leftarrow X$;
// determine V_a; //
if $|\bigcup_{l(u_i^*)=X} U_i| \leq |\bigcup_{l(u_i^*)=Y} U_i|$ **then**
    $V_a \leftarrow \bigcup_{l(u_i^*)=X} U_i$;
else
    $V_a \leftarrow \bigcup_{l(u_i^*)=Y} U_i$;
return V_a.

Theorem 3. *Consider a connected WSN under a CFDIA. If $p << 0.5$, then the CFDIA algorithm identifies the abnormal nodes correctly w.h.p.*

As the time overhead of the CFDIA-DIAG algorithm is dominated by the $O(|V| + |E|)$ time needed to perform the search-first search in the algorithm, we obtain that the worst-case time complexity of the diagnosis algorithm is $O(|V| + |E|)$. Additionally, the space complexity of the diagnosis algorithm is $O(|V| + |E|)$ as well.

5. Effectiveness of the Proposed Diagnosis Algorithm

This section is devoted to investigating the effectiveness of the CFDIA-DIAG algorithm through simulation experiments.

5.1. Metrics of Effectiveness of a Diagnosis Algorithm

In order to measure the effectiveness of the CFDIA algorithm, below let us introduce a pair of metrics of effectiveness of a diagnosis algorithm.

Definition 5. *Let $G = (V, E)$ be a WSN under a CFDIA, A be the set of abnormal nodes of G, and σ be a syndrome produced by A. Let $DIAG$ be a diagnosis algorithm. Let B be the set of nodes that are diagnosed to be abnormal by running $DIAG$ on (G, σ).*

1. *The diagnosis accuracy of $DIAG$ with respect to (w.r.t.) (G, A, σ) is defined as*

$$DA_{DIAG}(G, A, \sigma) = \frac{|B|}{|A|}. \tag{8}$$

2. *The false positive rate of DIAG w.r.t.* (G, A, σ) *is defined as*

$$FPR_{DIAG}(G, A, \sigma) = \frac{|B \bigcap (V - A)|}{|V - A|}. \quad (9)$$

3. *The false negative rate of DIAG w.r.t.* (G, A, σ) *is defined as*

$$FNR_{DIAG}(G, A, \sigma) = \frac{|A \bigcap (V - B)|}{|A|}. \quad (10)$$

5.2. Experiment Preparation

First, consider two additional diagnosis algorithms. The first one is almost the same as the CFDIA-DIAG algorithm, with the only exception of the sentence in line 10 of the CFDIA-DIAG algorithm being replaced with the sentence "arbitrarily choose a node u^* from U^*". We refer to this diagnosis algorithm as the *Random-Search algorithm*. The second is based on the *Correlation-Voting* solution proposed in Ref. [33].

Second, consider two different kinds of FDIAs: the simple FDIA (SFDIA) and the CFDIA. For the former, the readings of the compromised sensors are all enhanced by a larger fraction. For the latter, the readings of each adjacent pair of compromised sensors are changed in a coordinated manner.

Third, consider three synthetic WSNs, denoted G_1, G_2, and G_3, of sensor nodes that are with a communication radius of 20 m and are placed randomly in a square region of size 120×120 m^2, G_1; G_2, and G_3 have 50 nodes, 100 nodes, and 150 nodes, respectively. For each normal node v_i and any time t, assume $r_i(t)$ follows the Gaussian distribution $G(\mu_i, \sigma^2)$, where $\mu_i \in \{10, 15, 20\}$, $\sigma^2 = 1$, and the correlation coefficient of the readings of each adjacent pair of sensors is 0.9. Suppose a set of nine watchdogs with a communication radius of 40 m are deployed regularly in the region. See Figure 3 for the distribution of V_1 and nine watchdogs.

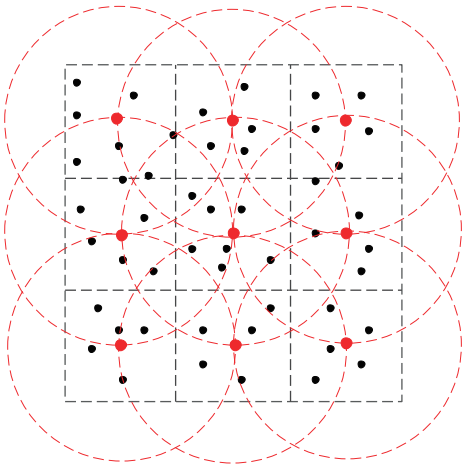

Figure 3. Distribution of V_1 and 9 watchdogs, where each red dot (resp. black dot) represents the watchdog (resp. sensor node), and each red circle represents the communication range of a watchdog.

Fourth, consider a real-world WSN, G_4, of 45 effective sensors used for gathering the environmental PM2.5 data, which is located in Krakow, Poland [38]. Here, the common sensing rate of the sensors is one reading per hour, and the average degree of the WSN is 21.16.

5.3. Experiments and Analysis of Experimental Results

Experiment 1. *Consider the WSN* G_2. *Let* $p \in \mathcal{P} = \{0.03, 0.06, \cdots, 0.3\}$. *Let* A_p *be a set of abnormal nodes randomly produced based on* p. *Let* σ_p^c *be the syndrome produced by* A_p *under the CFDIA that the readings of each of the compromised nodes are enhanced or reduced by 10%, and* σ_p^s

the syndrome produced by A_p under the SFDIA that the readings of all compromised are enhanced by 40%.

1. *For each $p \in \mathcal{P}$, running CFDIA-DIAG, Random-Search, and Correlation-Voting on (G_2, σ_p^c), we obtain their DA, FPR, and FNR, which are shown in Figure 4a–c. It is seen that (i) the diagnosis accuracy of CFDIA-DIAG is higher than those of the other two algorithms, and (ii) the false positive rate and false negative rate of CFDIA-DIAG is lower than those of the other two algorithms. Hence, we conclude that CFDIA-DIAG is more effective than the other two algorithms in the CFDIA situation.*
2. *For each $p \in \mathcal{P}$, running CFDIA-DIAG, Random-Search, and Correlation-Voting on (G_2, σ_p^s), we obtain their DA, FPR, and FNR, which are shown in Figure 4d–f. It is seen that (i) the diagnosis accuracy of CFDIA-DIAG is higher than those of the other two algorithms, and (ii) the false positive rate and false negative rate of CFDIA-DIAG is lower than those of the other two algorithms. Hence, we conclude that CFDIA-DIAG is more effective than the other two algorithms in the SFDIA situation.*

Experiment 2. *Consider the three WSNs: G_1, G_2, and G_3. Let $p \in \mathcal{P} = \{0.03, 0.06, \cdots, 0.3\}$. Let $A_{p,i}$ be a set of abnormal nodes of G_i randomly produced based on p. Let $\sigma_{p,i}^c$ be the syndrome on G_i produced by $A_{p,i}$ under the CFDIA that the readings of each of the compromised nodes are enhanced or reduced by 10%; $\sigma_{p,i}^s$ the syndrome on G_i produced by $A_{p,i}$ under the SFDIA that the readings of all the compromised are enhanced by 40%.*

1. *For each $p \in \mathcal{P}$ and each $i \in \{1, 2, 3\}$, running CFDIA-DIAG on $(G_i, \sigma_{p,i}^c)$, we obtain its DA, FPR, and FNR, which are shown in Figure 5a–c. It is seen that (i) the diagnosis accuracy of CFDIA-DIAG is higher when run on denser WSNs, and (ii) the false positive rate and false negative rate of CFDIA-DIAG is lower when run on denser WSNs. Hence, we conclude that in the CFDIA situation, CFDIA-DIAG is more effective when run on dense WSNs.*
2. *For each $p \in \mathcal{P}$ and each $i \in \{1, 2, 3\}$, running CFDIA-DIAG on $(G_i, \sigma_{p,i}^s)$, we obtain its DA, FPR, and FNR, which are shown in Figure 5d–f. It is seen that (i) the diagnosis accuracy of CFDIA-DIAG is higher when run on denser WSNs, and (ii) the false positive rate and false negative rate of CFDIA-DIAG is lower when run on denser WSNs. Hence, we conclude that in the SFDIA situation, CFDIA-DIAG is more effective when run on dense WSNs.*

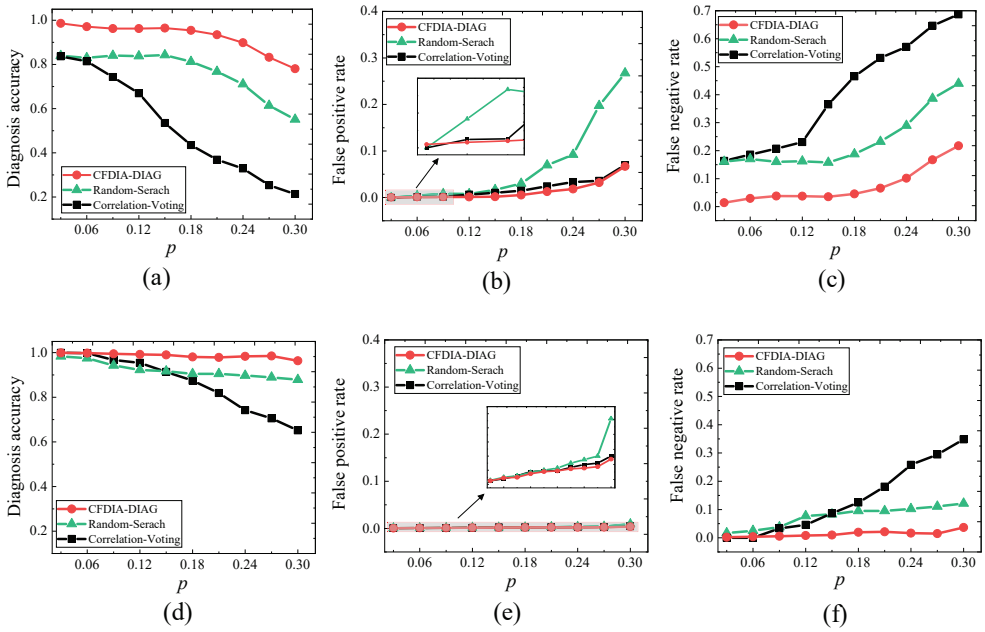

Figure 4. Effectiveness of the three diagnosis algorithms in experiment 1. (**a**) Diagnosis accuracy in the CFDIA situation. (**b**) False positive rate in the CFDIA situation. (**c**) False negative rate in the CFDIA situation. (**d**) Diagnosis accuracy in the SFDIA situation. (**e**) False positive rate in the SFDIA situation. (**f**) False negative rate in the SFDIA situation.

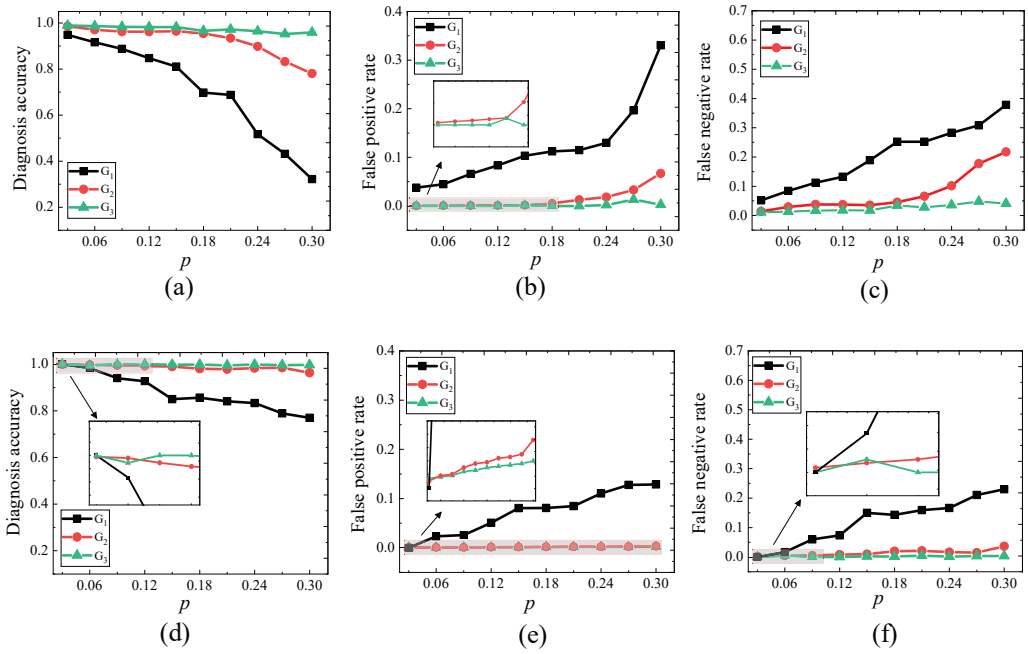

Figure 5. Effectiveness of the CFDIA-DIAG diagnosis algorithms in experiment 2. (**a**) Diagnosis accuracy in the CFDIA situation. (**b**) False positive rate in the CFDIA situation. (**c**) False negative rate in the CFDIA situation. (**d**) Diagnosis accuracy in the SFDIA situation. (**e**) False positive rate in the SFDIA situation. (**f**) False negative rate in the SFDIA situation.

Experiment 3. *Consider the WSN G_4. The real-world time-series sensor readings are stationary by first difference. Let $p \in \mathcal{P} = \{0.03, 0.06, \cdots, 0.3\}$. Let A_p be a set of abnormal nodes randomly produced based on p. Let σ_p^c be the syndrome produced by A_p under the CFDIA that the current reading of each of the compromised nodes is replaced with its largest reading in the past $K-1$ time points, and σ_p^s the syndrome produced by A_p under the SFDIA that the readings of all compromised are enhanced by 100%.*

1. *For each $p \in \mathcal{P}$, running CFDIA-DIAG, Random-Search, and Correlation-Voting on (G_4, σ_p^c), we obtain their DA, FPR, and FNR, which are shown in Figure 6a–c. It is seen that (i) the diagnosis accuracy of CFDIA-DIAG is higher than those of the other two algorithms, and (ii) the false positive rate and false negative rate of CFDIA-DIAG is lower than those of the other two algorithms. Again, we conclude that CFDIA-DIAG is more effective than the other two algorithms in the CFDIA situation.*
2. *For each $p \in \mathcal{P}$, running CFDIA-DIAG, Random-Search, and Correlation-Voting on (G_4, σ_p^s), we obtain their DA, FPR, and FNR, which are shown in Figure 6d–f. It is seen that (i) the diagnosis accuracy of CFDIA-DIAG is higher than those of the other two algorithms, and (ii) the false positive rate and false negative rate of CFDIA-DIAG is lower than those of the other two algorithms. Additionally, we conclude that CFDIA-DIAG is more effective than the other two algorithms in the SFDIA situation.*

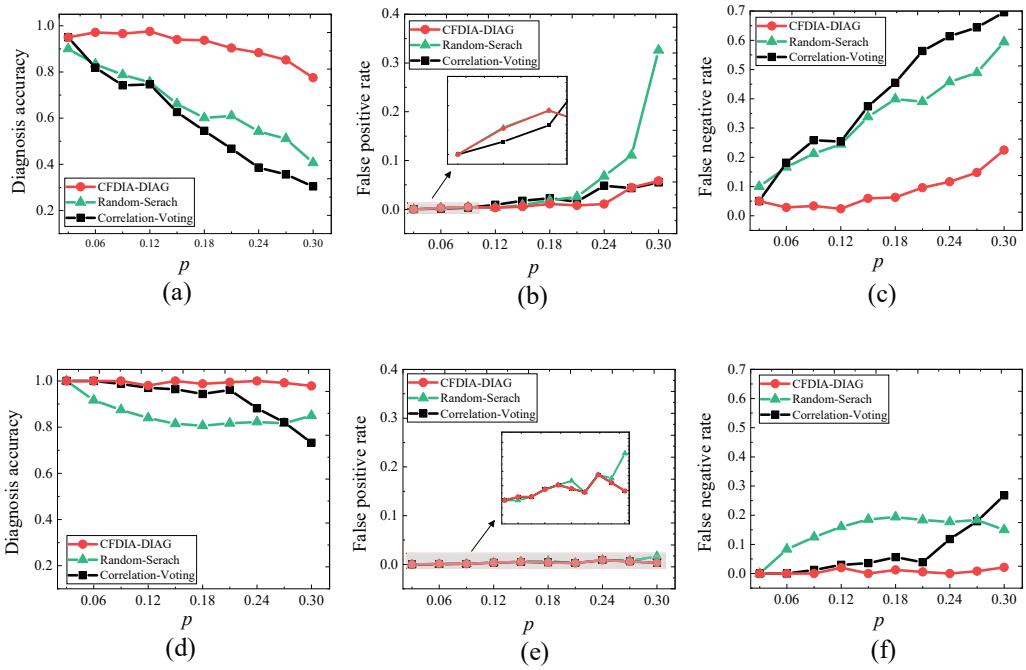

Figure 6. Effectiveness of the three diagnosis algorithms in experiment 3. (**a**) Diagnosis accuracy in the CFDIA situation. (**b**) False positive rate in the CFDIA situation. (**c**) False negative rate in the CFDIA situation. (**d**) Diagnosis accuracy in the SFDIA situation. (**e**) False positive rate in the SFDIA situation. (**f**) False negative rate in the SFDIA situation.

6. Conclusions and Future Work

A novel diagnosis scheme against a conclusive false data injection attack (CFDIA) has been proposed. First, a set of special watchdogs are deployed in a WSN under a CFDIA to collect consistency outcomes of adjacent pairs of sensor nodes and to deliver them to the base station, forming a syndrome. Second, inspired by the system-level fault diagnosis, a CFDIA diagnosis algorithm is presented. The effectiveness of the algorithm is corroborated through simulation experiments. By executing the diagnosis algorithm on the syndrome received by the base station, the set of compromised nodes is identified correctly with a higher probability.

According to the above three metrics of effectiveness, the method proposed in this paper is better than the compared method, but there are some limitations. Firstly, this paper considers only scenarios where attacks exist, and it is not yet designed to distinguish between malicious attacks and natural anomalies deviating from wide-sense stationary jointly Gaussian processes, including faults, disruptive events, and major disruptions. Secondly, the system-level fault diagnosis should be generalized to probabilistic system-level diagnoses to improve on the diagnosis accuracy and reduce the false positive rate and false negative rate of our proposed diagnosis algorithm [39]. Therefore, in future works, we should further optimize the method. Additionally, the proposed diagnosis algorithm should be extended to the diagnosis of mobile networks under a CFDIA [40,41]. Watchdogs can be made mobile to strike a balance between diagnosis accuracy and energy consumption, and the work can be done in the framework of game theory [42,43]. Finally, the methodology developed in the present paper may be applied to some other cybersecurity issues such as defense against advanced persistent threats [44,45].

Author Contributions: Conceptualization, J.H. and X.Y.; methodology, J.H., X.Y. and L.Y.; software, J.H.; validation, J.H., X.Y. and L.Y.; formal analysis, J.H., X.Y. and L.Y.; investigation, J.H. and L.Y.; resources, X.Y.; data curation, J.H. and X.Y.; writing—original draft preparation, J.H.; writing—review and editing, X.Y.; visualization, J.H.; supervision, X.Y.; project administration, X.Y.; funding acquisition, X.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (Grant No. 61572006).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are available upon request.

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

References

1. Forster, A. *Introduction to Wireless Sensor Networks*; Wiley-IEEE Press: Hoboken, NJ, USA, 2016.
2. El Emary, I.M.M.; Ramakrishnan, S. (Eds.) *Wireless Sensor Networks: From Theory to Applications*; CRC Press: Boca Raton, FL, USA, 2013.
3. Zhou, Y.; Fang, Y.; Zhang, Y. Securing wireless sensor networks: A survey. *IEEE Commun. Surv. Tutor.* **2008**, *10*, 6–28. [CrossRef]
4. Rani, A.; Kumar, S. A survey of security in wireless sensor networks. In Proceedings of the 3rd International Conference on CICT, Ghaziabad, India, 9–10 February 2017; pp. 1–5.
5. Mostefa, B.; Abdelkader, G. A survey of wireless sensor network security in the context of Internet of Things. In Proceedings of the 2017 4th International Conference on ICT-DM, Münster, Germany, 11–13 December 2017; pp. 1–8.
6. Sagar, V.B.B.; Munjul, M. Security issues in wireless sensor network-A survey. *J. Discret. Math. Sci. Cryptogr.* **2021**, *24*, 1415–1427.
7. Guan, Z.; Sun, N.; Xu, Y.; Yang, T. A comprehensive survey of false data injection in smart grid. *Int. J. Wirel. Mob. Comput.* **2015**, *8*, 27. [CrossRef]
8. Ahmed, M.; Pathan, A.K. False data injection attack (FDIA): An overview and new metrics for fair evaluation of its countermeasure. *Complex Adapt. Syst. Model.* **2020**, *8*, 4. [CrossRef]
9. Chen, P.; Yang, S.; McCann, J.A.; Lin, J.; Yang, X. Detection of false data injection attacks in smart-grid systems. *IEEE Commun. Mag.* **2015**, *53*, 206–213. [CrossRef]
10. Illiano, V.P.; Lupu, E.C. Detecting malicious data injections in wireless sensor networks: A survey. *ACM Comput. Surv.* **2015**, *48*, 24. [CrossRef]
11. Yang, L.; Ding, C.; Wu, M.; Wang, K. Robust detection of false data injection attacks for the data aggregation in Internet of Things-based environmental surveillance. *Comput. Netw.* **2017**, *129*, 410–428. [CrossRef]
12. Sood, K.; Nosouhi, M.R.; Kumar, N.; Gaddam, A.; Feng, B.; Yu, S. Accurate detection of IoT sensor behaviors in legitimate, faulty and compromised scenarios. *IEEE Trans. Dependable Secur. Comput.* **2021**, *20*, 288–300. [CrossRef]
13. Agrawal, S.; Das, M.L.; Lopez, J. Detection of node capture attack in wireless sensor networks. *IEEE Syst. J.* **2019**, *13*, 238–247. [CrossRef]
14. Musleh, A.S.; Chen, G.; Dong, Z.Y. A Survey on the detection algorithms for false data injection attacks in smart grids. *IEEE Trans. Smart Grid* **2020**, *11*, 2218–2234. [CrossRef]
15. Poornima, I.G.A.; Paramasivan, B. Anomaly detection in wireless sensor network using machine learning algorithm. *Comput. Commun.* **2020**, *151*, 331–337. [CrossRef]
16. Liu, J.; Labeau, F. Detection of false data injection attacks in industrial wireless sensor networks exploiting network numerical sparsity. *IEEE Trans. Signal Inf. Process. Over Netw.* **2021**, *7*, 676–688. [CrossRef]
17. Kreutzer, S.E.; Hakimi, S.L. System-level fault diagnosis: A survey. *Microprocess. Microprogr.* **1987**, *20*, 323–330. [CrossRef]
18. Duarte, E.P.; Ziwich, R.P.; Albini, L.C.P. A survey of comparison-based system-level diagnosis. *ACM Comput. Surv.* **2011**, *43*, 22. [CrossRef]
19. Lin, L.; Xu, L.; Chen, R.; Hsieh, S.; Wang, D. Relating extra connectivity and extra conditional diagnosability in regular networks. *IEEE Trans. Dependable Secur. Comput.* **2019**, *16*, 1086–1097. [CrossRef]
20. Wei, C.; Chen, C.; Hsieh, S. Conditional (t,k)-diagnosis in regular and irregular graphs under the comparison diagnosis model. *IEEE Trans. Dependable Secur. Comput.* **2018**, *15*, 351–356. [CrossRef]
21. Khilar, P.M.; Mahapatra, S. Intermittent fault diagnosis in wireless sensor networks. In Proceedings of the 10th International Conference on Information Technology (ICIT 2007), Rourkela, India, 17–20 December 2007; pp. 145–147.
22. Weber, A.; Kutzke, A.R.; Chessa, S. Diagnosability evaluation for a system-level diagnosis algorithm for wireless sensor networks. In Proceedings of the ISCC, Riccione, Italy, 22–25 June 2010; pp. 241–244.
23. Saha, T.; Mahapatra, S. Distributed fault diagnosis in wireless sensor networks. In Proceedings of the 2011 International Conference on Process Automation, Control and Computing, Coimbatore, India, 20–22 July 2011; pp. 1–5.
24. Barros, M.d.; Weber, A. System-level diagnosis for WSN: A heuristic. In Proceedings of the 2016 17th Latin-American Test Symposium (LATS), Foz do Iguacu, Brazil, 6–8 April 2016; pp. 171–176.
25. Chen, P.-Y.; Yang, S.; McCann, J.A. Distributed real-time anomaly detection in networked industrial sensing systems. *IEEE Trans. Ind. Electron.* **2014**, *62*, 3832–3842. [CrossRef]
26. Illiano, V.P.; Steiner, R.V.; Lupu, E.C. Unity is strength! Combining attestation and measurements inspection to handle malicious data injections in wsns. In Proceedings of the 10th ACM Conference on Security and Privacy in Wireless and Mobile Networks, Boston, MA, USA, 18–20 July 2017; pp. 134–144.

27. Aboelwafa, M.M.; Seddik, K.G.; Eldefrawy, M.H.; Gadallah, Y.; Gidlund, M. A machine-learning-based technique for false data injection attacks detection in industrial iot. *IEEE Internet Things J.* **2020**, *7*, 8462–8471. [CrossRef]
28. Berjab, N.; Le, H.H.; Yokota, H. A spatiotemporal and multivariate attribute correlation extraction scheme for detecting abnormal nodes in wsns. *IEEE Access* **2021**, *9*, 135266–135284. [CrossRef]
29. Huang, D.-W.; Liu, W.; Bi, J.C. Data tampering attacks diagnosis in dynamic wireless sensor networks. *Comput. Commun.* **2021**, *172*, 84–92. [CrossRef]
30. Lai, Y.; Tong, L.; Liu, J.; Wang, Y.; Tong, T.; Zhao, Z.; Qin, H. Identifying malicious nodes in wireless sensor networks based on correlation detection. *Comput. Secur.* **2022**, *113*, 102540. [CrossRef]
31. Bhuiyan, M.Z.A.; Wu, J. Collusion attack detection in networked systems. In Proceedings of the 2016 IEEE 14th Intl Conf on Dependable, Autonomic and Secure Computing, 14th Intl Conf on Pervasive Intelligence and Computing, 2nd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech), Auckland, New Zealand, 8–12 August 2016; pp. 286–293.
32. Illiano, V.P.; Muñoz-González, L.; Lupu, E.C. Don't fool me!: Detection, characterisation and diagnosis of spoofed and masked events in wireless sensor networks. *IEEE Trans. Dependable Secur. Comput.* **2016**, *14*, 279–293. [CrossRef]
33. Hau, Z.; Lupu, E.C. Exploiting correlations to detect false data injections in low-density wireless sensor networks. In Proceedings of the 5th on Cyber-Physical System Security Workshop, Auckland, New Zealand, 8 July 2019; pp. 1–12.
34. Choi, B.S. *ARMA Model Identification*; Springer: Berlin/Heidelberg, Germany, 2012.
35. Jolliffe, I.T.; Cadima, J. Principal component analysis: A review and recent developments. *Philos. Trans. R. Soc. A* **2015**, *374*, 20150202. [CrossRef] [PubMed]
36. Yu, Z.; Chin, W. Blind false data injection attack using PCA approximation method in smart grid. *IEEE Trans. Smart Grid* **2015**, *6*, 1219–1226. [CrossRef]
37. Ding, Y.; Liu, J. Real-time false data injection attack detection in energy internet using online robust principal component analysis. In Proceedings of the 2017 IEEE Conference on EI2, Beijing, China, 26–28 November 2017.
38. Krakow Air Quality Data. Available online: https://www.kaggle.com/datascienceairly/air-quality-data-from-extensive-network-of-sensors (accessed on 4 July 2022).
39. Elhadef, M.; Abrougui, K.; Das, S.; Nayak, A. A parallel probabilistic system-level fault diagnosis approach for large multiprocessor systems. *Parallel Process. Lett.* **2006**, *16*, 63–79. [CrossRef]
40. Gritzalis, S.; Karygiannis, T.; Skianis, C. (Eds.) *Security and Privacy in Mobile and Wireless Networking*; Troubador Publishing Ltd.: Leicester, UK, 2009.
41. Bendale, S.P.; Prasad, J.R. Security threats and challenges in future mobile wireless networks. In Proceedings of the 2018 IEEE GCWCN, Lonavala, India, 23–24 November 2018; pp. 146–150.
42. Alpcan, T.; Basar, T. *Network Security: A Decision and Game-Theoretic Approach*; Cambridge University Press: Cambridge, UK, 2010.
43. Shi, H.Y.; Wang, W.L.; Kwok, N.M.; Chen, S.Y. Game theory for wireless sensor networks: A survey. *Sensors* **2012**, *12*, 9055–9097. [CrossRef]
44. Yang, L.X.; Li, P.; Zhang, Y.; Yang, X.; Xiang, Y.; Zhou, W. Effective repair strategy against advanced persistent threat: A differential game approach. *IEEE Trans. Inf. Forensics Secur.* **2019**, *14*, 1713–1728. [CrossRef]
45. Yang, L.X.; Li, P.; Yang, X.; Tang, Y.Y. A risk management approach to defending against the advanced persistent threat. *IEEE Trans. Dependable Secur. Comput.* **2020**, *17*, 1163–1172. [CrossRef]