ORIGINAL RESEARCH PAPER

A novel secure diffusion Kalman filter algorithm against false data injection attacks

Yanan Du | Ning Li | Yonggang Zhang

College of Intelligent Systems Science and Engineering, Harbin Engineering University, Harbin, China

Correspondence
Ning Li, College of Intelligent Systems Science and Engineering, Harbin Engineering University, Harbin 150001, China.
Email: ningli@hrbeu.edu.cn

Abstract
This paper proposes a novel secure diffusion Kalman filter (dKF) algorithm to improve the estimation performance crippled by false data injection attacks on sensors in wireless sensor networks (WSNs). Different from the conventional dKF, each adjacent node in the WSNs is detected to ascertain its trustworthiness before local estimate fusion, so as to form a new secure network topology. Then the combination step is performed to fuse the information collected from the secure topology. The proposed secure dKF algorithm, having a better estimation performance, is robust to false data injection attacks on multiple sensors and partial elements of measurements. For the proposed secure dKF algorithm, its mean and mean-square performance are derived, based on which its convergence is analysed. Additionally, the estimating and tracking problem of projectile position is investigated to confirm the effectiveness of the proposed secure dKF algorithm. It is shown by simulations that the proposed secure dKF algorithm achieves a significant estimation performance gain.

1 | INTRODUCTION

Distributed estimation as an additional paradigm to centralized processing and decentralized processing with a fusion center is essential for the Internet of Things (IoT) and has flourished in past years. In distributed estimation, the fusion center is eliminated and a group of sensor nodes are interconnected to collaboratively realize state estimation from noisy measurements by sharing some information among them [1–3]. The distinctive talent eliminates the dysfunction issue of the central sensor node ultimately, and achieves higher flexibility, robustness and bandwidth and energy efficiency. As such, distributed estimation has been widely used in military and civilian fields, such as target tracking, auxiliary navigation, disaster forecast, incident investigation, and so on [4, 5].

Recent years have witnessed many recursive filtering algorithms [6] and distributed adaptive schemes, including diffusion least-mean squares (dLMS) [7], diffusion Kalman filtering (dKF) [8], dKF based on track fusion [9], diffusion nonlinear Kalman filter [10], distributed Bayesian estimation [11], distributed consensus strategy [12]. For all these works, it is assumed that the data transmission through networks is reliable. However, this assumption may not be realistic. That is primarily because wireless sensor nodes in sensing and communication may suffer from replay [13], false data injection [14], deception [15] and some other attacks [16]. Any attack may cause different kinds of severe consequences, such as customer information leakage, destruction of infrastructure, damages to state economics, endangering of human lives [17, 18].

In response to these adversarial attacks, some secure distributed algorithms were investigated to achieve reliable distributed estimation [19]. In [20], for the distributed consensus estimation attacked by false data injection, the relative entropy was leveraged as a stealthiness metric to detect whether the exchanged data in wireless sensor networks (WSNs) were attacked or not. In distributed consensus estimation, the convex combination coefficients are the same and cannot be adjusted. In [21], the stochastic coding detection scheme was proposed to detect malicious replay attacks without any performance loss in the normal system. In [22], reachability analysis was employed to construct a security layer enabling secure estimate information shares for the distributed dKF algorithm. In [23, 24], the secure distributed estimation problem over WSNs with malicious attacks imposed on sensor nodes and communications...
was considered. However, the reference estimate computed in [23] may be unreliable if there are multiple compromised neighbours in WSNs. There is no transmission of measurement information in dLMS subsystem in [24], which may degrade the distributed estimation performance to some extent. The secure state estimation in cyber-physical systems under attacks were reported in [25, 26]. But there is no information exchange between sensors for cyber-physical systems, which is relatively simple compared with dKF.

Different from the previous works, this paper develops a secure dKF algorithm under false data injection attacks for WSNs, which has a better robustness against attacks with an improved estimation performance. In the proposed secure dKF algorithm, each adjacent node is detected to ascertain its trustworthiness before local estimate fusion and form a new secure network topology. Then, the combination step is performed to fuse the information collected from the secure topology. The main contributions of this paper lie in two aspects: (i) a framework of secure dKF algorithm with false data injection attacks is presented; (ii) the performance of the proposed framework of secure dKF algorithm with false data injection is presented; (ii) the performance of the proposed secure dKF algorithm in mean and mean-square analyses is performed. Simulation results verify the effectiveness of the proposed algorithm.

2 | SYSTEM MODEL

Figure 1 depicts an attacked WSN comprising \( N \) spatially distributed sensor nodes, some of which are injected by false data, in a region. Two sensor nodes are assumed to be connected if directional communication occurs between them, with one sensor node always enabling connection to itself. All sensor nodes connected to the \( k \)th sensor node are termed as the \( k \)th sensor node's neighbourhood set, denoted by \( \mathcal{N}_k \) (notice that \( k \in \mathcal{N}_k \)). The size of the neighbourhood set \( \mathcal{N}_k \) is denoted by \( n_k \), which is termed as the degree of \( \mathcal{N}_k \). Whether the \( k \)th sensor node is connected to the other sensor nodes in this WSN or not can be mathematically described by the adjacency matrix of the \( k \)th sensor node, which is given by [8]

\[
[A]_{l,k} = \begin{cases} 
1 & \text{if } l \in \mathcal{N}_k, \\
0 & \text{otherwise}, 
\end{cases}
\]

Clearly, if the \( l \)th and \( k \)th sensor nodes are connected, then the \((l,k)\)th entry of the matrix \( A \) is one; otherwise, zero.

When no false data injection attacks exist, the following discrete-time linear system is given by

\[
\mathbf{x}_{i+1} = F_i \mathbf{x}_i + G_i \mathbf{w}_i, \\
\mathbf{y}_i = H_i \mathbf{x}_i + \mathbf{v}_i,
\]

where \( \mathbf{x}_i \in \mathbb{C}^M \) represents the system state variable and \( \mathbf{y}_i \in \mathbb{C}^{bN_i} \) denotes the system measurement variable, with \( M, N \) and \( b \) being the positive integers. For \( F_i, G_i, H_i \), they are known matrices. The process noise vector \( \mathbf{w}_i \) and measurement noise vector \( \mathbf{v}_i \) both follow independent and identical Gaussian distribution with \( \mathbf{w}_i \sim \mathcal{CN}(0, \mathbf{Q}_i) \) (\( \mathbf{Q}_i \geq 0 \)) and \( \mathbf{v}_i \sim \mathcal{CN}(0, \mathbf{R}_i) \) (\( \mathbf{R}_i > 0 \)). \( \{F_i, H_i\} \) and \( \{F_i, G_i, H_i\} \) are assumed to be detected and stabilized for every \( k \), respectively. Furthermore, the relationship of \( \mathbf{w}_i, \mathbf{v}_i, \mathbf{Q}_i \) and \( \mathbf{R}_i \) is given by

\[
\mathbb{E} \left[ \begin{bmatrix} \mathbf{w}_i \\ \mathbf{v}_i \end{bmatrix} \right] \left[ \begin{bmatrix} \mathbf{w}_i \\ \mathbf{v}_i \end{bmatrix} \right]^T = \begin{bmatrix} \mathbf{Q}_i & 0 \\ 0 & \mathbf{R}_i \end{bmatrix} \delta_{ij}, \tag{4}
\]

where the symbols \( T \) and \( \delta_{ij} \) represent transpose operator and Kronecker delta, respectively. We use the vector \( \mathbf{x}_0 \) satisfying \( \mathbf{x}_0 \sim \mathcal{CN}(0, \mathbf{P}_0) \) to denote the initial state of linear system with covariance matrix \( \mathbf{P}_0 > 0 \), and \( \mathbf{x}_0 \) is independent of \( \mathbf{w}_i \) and \( \mathbf{v}_i \) for all \( i \). According to (3), at time \( i \), the measurement \( \mathbf{y}_{k,i} \in \mathbb{C}^b \) collected by the \( k \)th sensor node is given by

\[
\mathbf{y}_{k,i} = H_{k,i} \mathbf{x}_{k,i} + \mathbf{v}_{k,i}, \quad k = 1, \ldots, N. \tag{5}
\]

Considering all sensor nodes together, the global measurement for this WSN at time \( i \) can be expressed as

\[
\mathbf{y}_i = \begin{bmatrix} \mathbf{y}_{1,i}^T \\ \vdots \\ \mathbf{y}_{N,i}^T \end{bmatrix}^T, \tag{6}
\]

\[
H_i = \begin{bmatrix} H_{1,i}^T \\ \vdots \\ H_{N,i}^T \end{bmatrix}^T, \tag{7}
\]

\[
\mathbf{v}_i = \begin{bmatrix} \mathbf{v}_{1,i}^T \\ \vdots \\ \mathbf{v}_{N,i}^T \end{bmatrix}^T. \tag{8}
\]

The measurement noise \( \mathbf{v}_{k,i} \) is assumed to be spatially uncorrelated, thus we have

\[
\mathbb{E} \mathbf{v}_{k,i} \mathbf{v}_{l,i}^T = R_{k,i} \delta_{k,l} \delta_{i,j}, \tag{9}
\]

where \( R_{k,i} > 0 \) for all \( k \) and \( i \).

Here, distributed secure estimation is considered in the WSN, where an invader sends false data to some sensor nodes for imposing attacks on sensing and malicious control. For the unattacked sensor nodes, they follow the discrete-time linear system given by (2) and (3). For the compromised sensor nodes,
fictitious data is injected into them by the attacker. As a result, once the fictitious data is collected by the sensor nodes, the attacked sensor nodes may miss the true measurements. It is worth stressing that the obtained L2-norm measurement residual for one sensor node has nothing to do with whether or not this sensor node is compromised [24]. That is, the injection attack is immune to the detection method of measurement residual.

To simplify notations, the measurement model for either of unattacked and compromised sensor nodes can be expressed as

\[ y_{k,i} = H_{k,i}(x_{k,i} + q_{k,i}) + v_{k,i}. \] (10)

The \( M \)-dimensional error vector \( q_{k,i} \) is a null vector for an unattacked sensor node but non-zero for any compromised sensor node.

3 DISTRIBUTION SECURE ESTIMATION AGAINST ATTACKS

For reliable distributed estimation in the attacked WSN, this section proposes a secure dKF algorithm to reduce and prevent the negative effects resulted from false data injection on some sensor nodes. The proposed secure dKF algorithm is implemented via four steps: forecast update, transmission of measurement information then calculation of local estimates, transmission of local estimates, then detection of trusted neighbours, and combination.

First, all sensor nodes, whether compromised or not by an attacker, separately update their forecast.

\[ \hat{x}_{k|1}^{loc} = F_{r-1} \hat{x}_{k|1-1}^{loc}, \] (11)

\[ P_{k|1}^{loc} = F_{r-1} P_{k|1-1}^{loc} F_{r-1}^T + G_r \mid Q_r \mid G_r^T, \] (12)

where \( P_{k|1}^{loc} \) is the state error covariance matrix.

Second, any sensor node exchanges measurement information with its neighbours, and then incremental updates are made to obtain local estimates by combining the information of this sensor node and its neighbours.

\[ \hat{x}_{k|1}^{loc} = x_{k|1}^{loc} + P_{k|1}^{loc} \sum_{l \in N_k} H_{l,i}^T R_{l,i}^{-1} y_{l,i}, \]

\[- P_{k|1}^{loc} \sum_{l \in N_k} H_{l,i}^T R_{l,i}^{-1} H_{l,i} \hat{x}_{l|1-1}^{loc} \]

\[ (P_{k|1}^{loc})^{-1} = (P_{k|1-1}^{loc})^{-1} + \sum_{l \in N_k} H_{l,i}^T R_{l,i}^{-1} H_{l,i}, \] (13)

Third, local estimates are exchanged between any sensor node and its neighbours. In the existing conventional dKF algorithm [8], each sensor node directly combines the local estimates of itself and its neighbours. However, the existing dKF algorithm is inappropriate in the considered attacked WSN. That is primarily because it is possible that the estimate of the whole WSN is ruined by the compromised sensor nodes through distributed cooperation. In view of this, the trusted neighbours of each sensor node are detected before local estimate fusion. Here, we adopt the method proposed in [24] to detect the compromised sensor nodes and form secure network topology.

To obtain a secure network topology, we need to get the reliable reference estimate of the \( k \)th sensor node at time \( t \). Once any sensor node \( k \) receives the neighbours’ estimates \( \hat{x}_{l|1}^{loc} \), it will sort all the \( n \)th elements from the neighbours by size including itself

\[ W_{k,i}^{(m)} = \left\{ \hat{x}_{i|1}^{loc}, \hat{x}_{j|1}^{loc}, \ldots, \hat{x}_{l|1}^{loc}, \hat{x}_{m|1}^{loc}, \hat{x}_{n|1}^{loc} \right\}^T, \] (15)

where \( l_1, l_2, \ldots, n_k \in N_k \) and \( \hat{x}_{i|1}^{loc} < \hat{x}_{j|1}^{loc} < \hat{x}_{k|1}^{loc} < \hat{x}_{l|1}^{loc} \).

From the perspective of attacker, the error between the local estimate \( \hat{x}_{l|1}^{loc} \) and the true value \( x \) is large if an attack is effective. In other words, the probability that the attack appears in the left or right hand of the set \( W_{k,i}^{(m)} \) is very high. It is assumed that the number of attacked neighbours is less than \( n_k/2 \) for the \( k \)th sensor node, then the number of trusted neighbours is greater than \( n_k/2 \). It is obtained that the \( n_k/2 \)th estimate of \( W_{k,i}^{(m)} \) is trusted with a very high probability. The detailed introduction can be referred in [24]. Then the \( n_k/2 \)th estimate of \( \hat{W}_{k,i}^{(m)} \) is set as

\[ \hat{W}_{k,i}^{(m)} = W_{k,i}^{(m)} \] (16)

Going through all elements among \( M \), the obtained reference estimate is given by

\[ \hat{W}_{k,i} = \left\{ \hat{W}_{k,i}^{(1)}, \hat{W}_{k,i}^{(2)}, \ldots, \hat{W}_{k,i}^{(M)} \right\}^T. \] (17)

Since a common vector \( x \) needs to be estimated through a distributed estimate algorithm, all local estimates \( \hat{x}_{l|1}^{loc} \) collected from the neighbours of the \( k \)th sensor node approximate to \( \hat{W}_{k,i} \). In view of this, the “consistency” principle is employed to confirm whether a sensor node is attacked or not, based on the distance between \( \hat{x}_{l|1}^{loc} \) and \( \hat{W}_{k,i} \). Defining a random variable characterizing the error between \( \hat{x}_{l|1}^{loc} \) and \( \hat{W}_{k,i} \), which is given by

\[ s_{k,i} \triangleq \hat{x}_{l|1}^{loc} - \hat{W}_{k,i}, \] (18)

where \( m_1 \) is the index of the entity whose element-wise \( L_1 \)-norm distance between \( \hat{x}_{l|1}^{loc} \) and \( \hat{W}_{k,i}^{(m_1)} \) is the largest among \( M \) elements.

Clearly, if an attack is effective, the error between the local estimator \( \hat{x}_{l|1}^{loc} \) and the true value \( x \) is large, resulting in a large
Then, a compromised sensor node can be detected using the threshold test, which is given by

$$\hat{s}_{k,i}^2 \geq \gamma_{k,i}, \quad l \in \mathcal{N}'_{k,i}$$

where $H_0$ and $H_1$ represent the hypotheses that the sensor node $l$ is reliable and unreliable, respectively. $\gamma_{k,i}$ is a pre-defined threshold

$$\gamma_{k,i} = 4\hat{\sigma}_k^2$$

where $\hat{\sigma}_k^2$ is $\hat{\mathbf{W}}_{k,i}$'s averaged element-wise sample variance, which are detailed in [24]. By derivation, it can be found that $\hat{\sigma}_k^2$ is a function of $\hat{\mathbf{W}}_{k,i-1}, \ldots, \hat{\mathbf{W}}_{k,1}, \hat{\sigma}_k^2$ and the reference estimate mean $\bar{\mathbf{\mu}}_{k,1}$, where $\hat{\mathbf{W}}_{k,1}$, $\hat{\sigma}_k^2$ and $\bar{\mathbf{\mu}}_{k,1}$ can be initialized. Moreover, the $s$th element $\hat{\mathbf{W}}_{k,i}^{(s)}$ of $\hat{\mathbf{W}}_{k,i}$ is obtained by sorting all the $s$th elements from the neighbours and finding the middle value. Obviously, $\hat{\mathbf{W}}_{k,i}^{(s)}$ is entirely unrelated to the noise covariances, and the same as $\hat{\mathbf{W}}_{k,i}$. Clearly, the variance $\hat{\sigma}_k^2$ is independent of noise covariances.

Through the threshold test, the set of trusted neighbours of the $k$th sensor node in an instant can be obtained as follows

$$\mathcal{N}_{k,i} = \left\{ l \in \mathcal{N}_{k,i}' : \hat{s}_{l,k,i}^2 < \gamma_{k,i} \right\}$$

Finally, based on the coefficient $c_{k,i,l}$ and the local estimates $\hat{s}_{l,k,i}^2$ in (13) received from the trusted neighbours of the $k$th sensor node, the combination step of the secure dKF algorithm is

$$\hat{\mathbf{x}}_{k,i} = \sum_{l \in \mathcal{N}_{k,i}} c_{k,i,l} \hat{\mathbf{x}}_{l,k,i}$$

where the convex combination coefficient $c_{k,i,l}$ satisfies

$$\sum_{l \in \mathcal{N}_{k,i}} c_{k,i,l} = 1, \quad 0 \leq c_{k,i,l} \leq 1.$$  

### 4 PERFORMANCE ANALYSIS

This section will analyse the performance of the proposed secure dKF algorithm. In the attacked WSN, distributed estimation suffers significantly from false alarm as well as missing detection. Particularly, once false alarm occurs, the performance of distributed estimation may be degraded since the trusted neighbours of a sensor node are reduced. Unfortunately, malicious data does not spread out. By contrast, accompanied by missing detection, the estimate of the whole WSN may be ruined by the compromised nodes due to improper data fusion. In view of this, more attention should be paid to the error of missing detection. Lacking prior information of false data injection, it is hard to exactly derive the probability of false alarm as well as missing detection in the distributed estimation. As the purpose of detection in this work is to provide a secure distributed estimation against the attacks of false data injection, the performance is analysed for distributed estimation.

In the proposed secure dKF algorithm, a sensor node, whose index is represented by $l$ without loss of generality, is regarded as a reliable and trustworthy neighbour of the $k$th sensor node if the following relationship holds

$$\left| \hat{x}_{l,k,i}^{bc}(m) - \hat{\mathbf{W}}_{k,i}^{(m)} \right| < \sqrt{\gamma_{k,i}},$$

where $m$ denotes the element index of the vector $\hat{x}$. Further, the inequality (24) is rewritten as

$$\hat{\mathbf{W}}_{k,i} - \sqrt{\gamma_{k,i}} \cdot \| l < \hat{x}_{l,k,i}^{bc} < \hat{\mathbf{W}}_{k,i} + \sqrt{\gamma_{k,i}} \cdot \|,$$

where $\|l$ denotes an $M \times 1$ unit-entry vector. Therefore, the state $\hat{x}_{l,k,i}$ is given by

$$\hat{x}_{l,k,i}^{bc} = \hat{\mathbf{W}}_{k,i} + \theta_{l,i},$$

where $\theta_{l,i} = [\theta_{l,i}^{(1)}, \theta_{l,i}^{(2)}, \ldots, \theta_{l,i}^{(M)}]^T$ is a random vector, each entry of which is under the bound constraint $\| \theta_{l,i}^{(m)} < \sqrt{\gamma_{k,i}}$. It is easy to find that $\theta_{l,i}$ is independent of $\hat{\mathbf{W}}_{k,i}$ and the noise covariances.

Depending on whether the detection of reliable neighbours in the WSN is missing or not, the set $\mathcal{N}_{k,i}'$ of the $k$th sensor node is divided into two disjoint subsets, which are denoted by $\mathcal{N}_{k,i}'^+$ and $\mathcal{N}_{k,i}'^-$. To be specific, the subset $\mathcal{N}_{k,i}'^+$ consists of unattacked and detected reliable neighbours, and the subset $\mathcal{N}_{k,i}'^-$ is made up of compromised but detected reliable neighbours. Then the state $\hat{\mathbf{x}}_{k,i}$ in (22) is given by

$$\hat{\mathbf{x}}_{k,i} = \sum_{l \in \mathcal{N}_{k,i}'^+} c_{k,i,l} \hat{x}_{l,k,i}^{bc} + \sum_{l \in \mathcal{N}_{k,i}'^-} c_{k,i,l}(\hat{\mathbf{W}}_{k,i} + \theta_{l,i}).$$

Combining (13) and (27), we have

$$\hat{\mathbf{x}}_{k,i} = \sum_{l \in \mathcal{N}_{k,i}'^+} c_{k,i,l} \hat{x}_{l,k,i}^{bc} + \sum_{l \in \mathcal{N}_{k,i}'^-} c_{k,i,l}(\hat{\mathbf{W}}_{k,i} + \theta_{l,i}).$$

$$\hat{\mathbf{x}}_{k,i} = \sum_{l \in \mathcal{N}_{k,i}'^+} c_{k,i,l} \hat{x}_{l,k,i}^{bc} + \sum_{l \in \mathcal{N}_{k,i}'^-} c_{k,i,l}(\hat{\mathbf{W}}_{k,i} + \theta_{l,i}).$$
Recall that
\[
\sum_{l \in \mathcal{N}_{h,i}^+} \epsilon_{kl,i} = \sum_{l \in \mathcal{N}_{h,i}^-} \epsilon_{kl,i} + \sum_{l \in \mathcal{N}_{h,i}^0} \epsilon_{kl,i} = 1.
\]

Let the true value \( \mathbf{x} \), subtract both sides of (28). Then we have
\[
\tilde{\mathbf{x}}_{k,i} = \sum_{l \in \mathcal{N}_{h,i}^+} \epsilon_{kl,i} \tilde{\mathbf{x}}_{l,i} - \sum_{l \in \mathcal{N}_{h,i}^+} \epsilon_{kl,i} \mathbf{P}_{l,i}^{(k)} \mathbf{H}_{m,i}^T \mathbf{R}_{m,i}^{-1} (\mathbf{H}_{m,i} \mathbf{x} + \mathbf{v}_{m,i}) + \sum_{l \in \mathcal{N}_{h,i}^+} \epsilon_{kl,i} \mathbf{P}_{l,i}^{(k)} \mathbf{H}_{m,i}^T \mathbf{R}_{m,i}^{-1} \mathbf{H}_{m,i} \tilde{\mathbf{x}}_{l,i} - \sum_{l \in \mathcal{N}_{h,i}^0} \epsilon_{kl,i} \mathbf{R}_{l,i} \mathbf{\theta}_{l,i},
\]
where \( \tilde{\mathbf{x}}_{l,i} = \mathbf{x}_i - \hat{\mathbf{x}}_{l,i} \) and \( \hat{\mathbf{x}}_{k,i} = \mathbf{x}_i - \hat{\mathbf{x}}_{k,i} \) denote the local estimate error and estimate error, respectively, and \( \mathbf{W}_{k,i} = \mathbf{x}_i - \mathbf{W}_{k,i} \) is the reference estimate error.

To analyse the performance of the whole network, some global quantities are defined. The augmented state error \( \tilde{\mathbf{X}}_{i,j} \), reference error estimate vector \( \mathbf{\tilde{W}}_{i,j} \), random vector \( \mathbf{\theta}_{i,j} \), the block diagonal matrices \( \mathbf{H}_{i,j} \), \( \mathbf{P}_{i,j}^{(k)} \), and \( \mathbf{R}_{i,j} \) are defined respectively as follows:

\[
\tilde{\mathbf{X}}_{i,j} \triangleq \text{col}\{\tilde{\mathbf{x}}_{1,i,j}, \ldots, \tilde{\mathbf{x}}_{N_{h,i,j}}\},
\]
\[
\tilde{\mathbf{W}}_{i,j} \triangleq \text{col}\{\tilde{\mathbf{W}}_{1,i,j}, \ldots, \tilde{\mathbf{W}}_{N_{h,i,j}}\},
\]
\[
\mathbf{\theta}_{i,j} \triangleq \text{col}\{\mathbf{\theta}_{1,i,j}, \ldots, \mathbf{\theta}_{N_{h,i,j}}\},
\]
\[
\mathbf{H}_{i,j} \triangleq \text{diag}\{\mathbf{H}_{1,i,j}, \ldots, \mathbf{H}_{N_{h,i,j}}\},
\]
\[
\mathbf{P}_{i,j}^{(k)} \triangleq \text{diag}\{\mathbf{P}_{1,i,j}^{(k)}, \ldots, \mathbf{P}_{N_{h,i,j}}^{(k)}\},
\]
\[
\mathbf{R}_{i,j} \triangleq \text{diag}\{\mathbf{R}_{1,i,j}, \ldots, \mathbf{R}_{N_{h,i,j}}\}.
\]

And two combination matrices \( \mathbf{C}_i^+ \) and \( \mathbf{C}_i^- \) are defined, respectively,
\[
[C_i^+]_{kl} = \begin{cases} 
\epsilon_{kl,i} & \text{if } l \in \mathcal{N}_{h,i}^+,
0 & \text{otherwise},
\end{cases}
\]
\[
[C_i^-]_{kl} = \begin{cases} 
\epsilon_{kl,i} & \text{if } l \in \mathcal{N}_{h,i}^-,
0 & \text{otherwise},
\end{cases}
\]

Their extended matrices are defined as follows
\[
\mathbf{C}_i^+ \triangleq \mathbf{C}_i^+ \otimes \mathbf{I}_M,
\]
\[
\mathbf{C}_i^- \triangleq \mathbf{C}_i^- \otimes \mathbf{I}_M,
\]
where \( \otimes \) denotes Kronecker product. It is clear that \( \|\mathbf{C}_i^+\|_2 \), \( \|\mathbf{C}_i^-\|_2 \leq 1 \). And these two extended matrices are related to the errors of false alarm and missing detection, respectively.

Combining (6) and the above global quantities, (30) can be extended to a global form
\[
\mathbf{X}_{i,j} = (\mathbf{C}_i^+)^T (\mathbf{I}_N \otimes \mathbf{F}_{i,j}) \mathbf{X}_{i-1,j-1} + (\mathbf{C}_i^+)^T \mathbf{P}_{i,j}^{(k)} \mathbf{H}_{j,i}^T \mathbf{R}_{j,i}^{-1} \mathbf{H}_{j,i} (\mathbf{I}_N \otimes \mathbf{G}_{i,j}) (\mathbf{I}_N \otimes \mathbf{G}_{i,j})^T (\mathbf{I}_N \otimes \mathbf{F}_{i,j}) \mathbf{X}_{i-1,j-1} + (\mathbf{C}_i^+)^T \mathbf{P}_{i,j}^{(k)} \mathbf{H}_{j,i}^T \mathbf{R}_{j,i}^{-1} \mathbf{H}_{j,i} (\mathbf{I}_N \otimes \mathbf{G}_{i,j}) (\mathbf{I}_N \otimes \mathbf{G}_{i,j})^T (\mathbf{I}_N \otimes \mathbf{F}_{i,j}) \mathbf{X}_{i-1,j-1} + (\mathbf{C}_i^-)^T \mathbf{P}_{i,j}^{(k)} \mathbf{H}_{j,i}^T \mathbf{R}_{j,i}^{-1} \mathbf{\theta}_{j,i}.
\]

To simplify notations, (41) can be further expressed as
\[
\tilde{\mathbf{X}}_{i,j} = \mathbf{F}_{i,j} \tilde{\mathbf{X}}_{i-1,j-1} + \mathbf{G}_{i,j} (\mathbf{I}_N \otimes \mathbf{\omega}_{i-1}) - \mathbf{D}_{i} \mathbf{\nu}_{i} + (\mathbf{C}_i^+)^T \mathbf{P}_{i,j}^{(k)} \mathbf{H}_{j,i}^T \mathbf{R}_{j,i}^{-1} \mathbf{\theta}_{j,i}.
\]

where
\[
\mathbf{F}_{i,j} \triangleq (\mathbf{C}_i^+)^T (\mathbf{I}_N \otimes \mathbf{P}_{i,j}^{(k)} \mathbf{S}_{i,j}^{(k)} \mathbf{I}_N \otimes \mathbf{F}_{i,j}),
\]
\[
\mathbf{G}_{i,j} \triangleq (\mathbf{C}_i^-)^T (\mathbf{I}_N \otimes \mathbf{P}_{i,j}^{(k)} \mathbf{S}_{i,j}^{(k)} \mathbf{I}_N \otimes \mathbf{G}_{i,j}),
\]
\[
\mathbf{D}_{i} \triangleq (\mathbf{C}_i^-)^T \mathbf{P}_{i,j}^{(k)} \mathbf{H}_{j,i}^T \mathbf{R}_{j,i}^{-1},
\]
\[
\mathbf{S}_{i,j} \triangleq \mathbf{H}_{j,i}^T \mathbf{R}_{j,i}^{-1} \mathbf{H}_{j,i}.
\]

It is not difficult to find that (42a) is similar to those in [8]. The only difference between them is the matrix \( \mathbf{C}_i^+ \) in (42) depends on the subset \( \mathcal{N}_{h,i}^+ \) consisting of unattacked and detected reliable neighbours.

### 4.1 Mean performance

For (42), its expectation is given by
\[
\mathbb{E}[\tilde{\mathbf{X}}_{i,j}] = \mathbb{E}[\mathbf{F}_{i,j} \tilde{\mathbf{X}}_{i-1,j-1}] + \mathbb{E}[\mathbf{G}_{i,j} (\mathbf{I}_N \otimes \mathbf{\omega}_{i-1})] - \mathbb{E}[\mathbf{D}_{i} \mathbf{\nu}_{i}] + \mathbb{E}[(\mathbf{C}_i^+)^T \mathbf{P}_{i,j}^{(k)} \mathbf{H}_{j,i}^T \mathbf{R}_{j,i}^{-1} \mathbf{\theta}_{j,i}].
\]
\begin{align}
&= E[F_i, \hat{X}_{i-1| i-1}] \\
&+ E[(C_{i-1}^T)^T \hat{W}_{ki} - E[(C_{i-1}^T)^T \Theta_i]]
\end{align}

where (47d) and (47e) hold as $\omega_i$ and $\nu_i$ are both zero mean.

Equation (47d) is similar to those in the conventional dKF algorithm. The only difference is the matrix $C_{i-1}^+$ depends on the subset $\mathcal{N}_{ki}^+$ consisting of unattacked and detected reliable neighbours. Therefore, it is determined by the probability of false alarm. Considering that $\|C_{i-1}^+\|_2 \leq 1$, it always converges as $i \to \infty$.

Equation (47e) depends on the cooperation matrix $C_{i-1}^+$, reference estimate error $\hat{W}_{ki}$ and $\Theta_i$. Similar to $C_{i-1}^+$, $\|C_{i-1}^+\|_2 \leq 1$ holds as it is bounded. By adopting the proving method of convergence of reference estimate in [24], we can prove that the reference estimate $\hat{W}_{ki}$ always converges to an unbiased estimate, therefore $E[\hat{W}_{ki}] = 0$ as $i \to \infty$ can be concluded. Taking $\|C_{i-1}^+\|_2 \leq 1$ into account, we can derive that $E[(C_{i-1}^+)^T \hat{W}_{ki}] = 0$ as $i \to \infty$. Since $C_{i-1}^+$ and each element $\Theta_i$ in the global vector $\Theta_i$ are both bounded, the last term is also bounded by a small vector. Extremely, if all attacked sensors are detected to be unreliable, $C_{i-1}^+$ must converge to a null matrix. Thus $\hat{X}_{i| i}$ converges to an unbiased estimate of $X_i$.

### 4.2 Mean-square performance

The mean-square deviation (MSD) of the proposed secure dKF algorithm is analysed in this subsection. Considering that it is hard to access the exact performance of detection, the MSD boundness of the proposed secure dKF algorithm will be analysed. The MSD for the $k$th sensor node at time $i$ is given by

\begin{align}
\text{MSD}_{ki} &= E\|X_{i| i} - \hat{X}_{i| i}^T\|^2
\end{align}

According to (42) and (48), using the identity of Kronecker products $(A \otimes B)(C \otimes D) = (AC \otimes BD)$, the networked MSD is given by

\begin{align}
E\|X_{i| i}\|^2 &= E[F_i, \hat{X}_{i-1| i-1}^T \hat{X}_{i-1| i-1}^T F_i^T] \\
&+ G_i(\|\Theta_i\|_2 \otimes Q_{i-1}) G_i^T + D_i R_i D_i^T
\end{align}

where (49a) and (49b) are dependent on the probability of false alarm since its combination matrix $C^+$ is determined by the sub-set $\mathcal{N}_{ki}^+$. Moreover, they always converge as $i \to \infty$ considering that $\|C_{i-1}^+\|_2 \leq 1$. The convergence proof is similar to equation (32) in [8]. Similar to the analysis of mean stability in Section 4.1, (49e)–(49f) are the functions of the cooperation matrix $C_{i-1}^+$, reference error $\hat{W}_{ki}$ and $\Theta_i$. $\|C_{i-1}^+\|_2 \leq 1$ holds, as it is bounded. According to [24], the reference estimate $\hat{W}_{ki}$ always converges to an unbiased estimate of the true value $x_i$. Correspondingly, it is derived that $E[\hat{W}_{ki}] = 0$ as $i \to \infty$, then $E[\hat{W}_{ki}\hat{W}_{ki}^T] = 0$ as $i \to \infty$. In view of the fact that $\|C_{i-1}^+\|_2 \leq 1$, it can be concluded that $E[(C_{i-1}^+)^T \hat{W}_{ki}\hat{W}_{ki}^T C_{i-1}^+] = 0$ as $i \to \infty$. Owing to the boundness of $C_{i-1}^+$ and each element $\Theta_i$ in $\Theta_i$, $E[(C_{i-1}^+)^T \Theta_i\Theta_i^T C_{i-1}^+] = 0$ and $E[(C_{i-1}^+)^T \Theta_i\Theta_i^T C_{i-1}^+]$ is bounded by a small vector.

Extremely, if all attacked sensors are detected to be unreliable, $C_{i-1}^+$ must converge to a null matrix, and the sum of (49e)–(49f) converges to a null matrix. In this case, the MSD in (49) is transformed into a form similar to equation (32) in [8]. Moreover, if no false alarm occurs, $C_{i-1}^+$ converges to a stable combination weight matrix. Then, (49) is given by

\begin{align}
E\|X_{i| i}\|^2 &= E[F_i, X_{i-1| i-1}^T - X_{i-1| i-1}^T F_i^T] \\
&+ G_i(\|\Theta_i\|_2 \otimes Q_{i-1}) G_i^T + D_i R_i D_i^T
\end{align}

Let $P_{\hat{X}_{i| i}} = E[\hat{X}_{i| i}, \hat{X}_{i| i}^T]$, then (50) is written as

\begin{align}
P_{\hat{X}_{i| i}} &= F_i P_{X_{i-1| i-1}} F_i^T \\
&+ G_i(\|\Theta_i\|_2 \otimes Q_{i-1}) G_i^T + D_i R_i D_i^T
\end{align}

It is assumed that $F_i$, $G_{ki}$, $H_{ki}$, $Q_i$ and $R_{ki}$ in model (2) and (3) are independent of $i$. Then the convergence of the secure dKF algorithm can be completely guaranteed. Thus, (51) converges to the unique solution of the Lyapunov equation

\begin{align}
P_{\hat{X}} &= FP_{\hat{X}} F^T + G(\|\Theta\|_2 \otimes Q) G^T + DRD^T.
\end{align}

The steady-state MSD at sensor node $k$ now can be given by [8]

\begin{align}
\text{MSD}_{ki} &= \lim_{i \to \infty} E\|X_i - \hat{X}_{ki}\|^2 = \text{Tr}(P_{\hat{X}} I_k),
\end{align}

where $I_k$ is an $NM \times NM$ block matrix, each block of which has a size of $M \times M$. Moreover, the $(k, k)$th block in $I_k$ is an identity matrix and the others are null matrices. Finally, the average steady-state MSD for the network is given by

\begin{align}
\text{MSD}_{\text{ave}} &= \frac{1}{N} \text{Tr}(P_{\hat{X}}).
\end{align}
5 | SIMULATIONS

This section will show how effective the proposed secure dKF algorithm goes against the attacks due to false data injection by numerical simulations. In simulations, the problem for estimating and tracking projectile position is considered, in which each sensor node is able to obtain its noisy measurement of the projectile position. In the proposed secure dKF algorithm, our goal is to estimate the exact projectile location at each time.

As shown in Figure 2, a WSN consisting of \( N = 20 \) sensor nodes is attacked by false data injection. Here, we assume that sensor nodes \( k = 6 \) and \( k = 16 \) are attacked, and all elements of the corresponding error vectors are set as 5. For simplicity, it is also assumed that each sensor node only measures the vertical position and one horizontal position of the projectile like a two-dimensional plane. Additionally, the estimated projectile state is represented by a vector having \( M = 6 \) elements, which is formed by stacking its position, velocity, and acceleration. That is, \( \mathbf{x}_i = [x_i, y_i, v_{x,i}, v_{y,i}, a_{x,i}, a_{y,i}]^T \). The initial state can be set as \( \mathbf{x}_0 = [13, 0, V \cos \theta, V \sin \theta, 0, 0]^T \), where \( \theta = \frac{\pi}{3} \), \( V = 15 \).

Besides, \( a_{x,i} = 0 \) and \( a_{y,i} = 5 \) are accelerations of projectile in two orthogonal directions, the constant \( g = 10 \) is gravitational acceleration, and \( T = 0.2r \) is time interval. The covariances of states and measurement noises are \( \mathbf{Q}_i = 0.01 \mathbf{I} \) and \( \mathbf{R}_i = 0.01 \mathbf{I} \), respectively. The process noise matrix is \( \mathbf{G}_i = \mathbf{I} \). The covariance of the initial state is \( \mathbf{P}_0 = 0.5 \mathbf{I} \). For all sensor nodes, their state and measurement matrices are the same, which are given by

\[
\mathbf{F} = \begin{bmatrix}
1 & 0 & T & 0 & 0 & 0 \\
0 & 1 & 0 & T & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & -T \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 \\
\end{bmatrix}
\]

and

\[
\mathbf{H} = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
\end{bmatrix}
\]

Based on the same network, the results are averaged over 500 independent and random observations.

In Figure 3, the projectile trajectory is shown in five cases including true trajectory as well as the trajectories obtained by the conventional dKF algorithm [8] with and without attack, the proposed secure dKF algorithm and the fifth sensor node’s noisy measurements. These cases correspond to green square, blue rhombus, black triangle, red pentagon and purple inverse triangle curves, respectively. It is obvious that the estimate produced by the proposed secure dKF algorithm is close to the true trajectory.

Figure 4 depicts the transient MSD for different cases, which is the function of time. The results of the conventional dKF with and without attacks are given for comparison. It is clear that the conventional dKF with attack (blue, rhombus) indicates the worst estimation performance because of the propagation of measurements compromised by attackers. However, the conventional dKF algorithm without attack (black, triangle) achieves the best estimation performance. Obviously, the transient MSD of the proposed secure dKF algorithm (red, pentagon) is close to the transient MSD of the conventional dKF without attack, showing good performance. The reason for this subtle difference between them is that the information produced by compromised sensors is removed.
The steady-state MSD of different algorithms is investigated in Figure 5. Because the propagation of measurements compromised by attackers over the network, the conventional dKF algorithm with attack (blue, rhombus) shows the worst estimation performance at each node. Compared to the case of the conventional dKF algorithm with attack, the proposed secure dKF algorithm (red, pentagon), which shows good performance at all nodes, is more robust to attacks.

6 | CONCLUSIONS

The problem of distributed secure estimation over WSN crippled by false data injection attacks is considered here. To address this problem, a secure dKF algorithm is proposed, in which the trusted neighbours of each sensor node are detected before local estimate fusion to form a secure network topology and then the combination step is performed by combining its own local estimate and the received estimates from the trusted neighbours. The mean and mean-square analysis of the proposed secure dKF algorithm is also provided. Based on this, the convergence is analysed. Finally, the good performance of the proposed secure dKF algorithm is demonstrated by numerical simulations. On the other hand, there exists energy waste in detecting some compromised sensors in the proposed method, which will be investigated in the future work.

REFERENCES

1. Li, W., Jia, Y.: Consensus-based distributed multiple model UKF for jump Markov nonlinear systems. IEEE Trans. Autom. Control. 57(1), 227–235 (2012)
2. Sayed, A., et al.: Diffusion strategies for adaptation and learning over networks: an examination of distributed strategies and network behavior. IEEE Signal Process. Mag. 30(3), 155–171 (2013)
3. He, S., et al.: Distributed estimation over a low-cost sensor network: A review of state-of-the-art. Inf. Fusion 54, 21–43 (2020)
4. Li, W., et al.: A survey on multisensor fusion and consensus filtering for sensor networks. Discrete Dyn. Nat. Soc. 39, 1–12 (2015)
5. Hu, J., et al.: Estimation, filtering and fusion for networked systems with network-induced phenomena: New progress and prospects. Inf. Fusion 31, 65–75, (2016)
6. Mao, J., et al.: Recursive filtering of networked nonlinear systems: A survey. Int. J. Syst. Sci. 52(6), 1110–1128 (2021)
7. Takahashi, N., et al.: Diffusion least-mean squares with adaptive combiners: Formulation and performance analysis. IEEE Trans. Signal Process. 58(9) 4795–4810 (2010)
8. Cattivelli, F., Sayed, A.: Diffusion strategies for distributed Kalman filtering and smoothing. IEEE Trans. Autom. Control. 55(9), 2069–2084 (2010)
9. Wang, G., et al.: Diffusion distributed Kalman filter over sensor networks without exchanging raw measurements. Signal Process. 132, 1–7 (2017)
10. Wang, G., et al.: Diffusion nonlinear Kalman filter with intermittent observations. Proc. Inst. Mech. Eng. Part G: J. Aerosp. Eng. 232(15), 2775–2783 (2018)
11. Hua, J., Li, C.: Distributed robust Bayesian filtering for state estimation. IEEE Trans. Signal Inf. Proc. Netw. 5(3), 428–441 (2019)
12. Carli, R., et al.: Distributed Kalman filtering based on consensus strategies. IEEE J. Sel. Areas Commun. 26(4), 622–633 (2008)
13. Zhu, M., Martinez, S.: On the performance analysis of resilient networked control systems under replay attacks. IEEE Trans. Autom. Control 59(3), 804–808 (2014)
14. Liu, Y., et al.: False data injection attacks against state estimation in electric power grids. In: Proceedings of the 16th ACM Conference on Computer and Communications Security, pp. 21–32 (2009)
15. Liu, J., et al.: Distributed event-triggered state estimators design for sensor networked systems with deception attacks. IET Contr. Theory Appl. 13(17), 2783–2791 (2019)
16. Xu, S., et al.: Intelligent reflecting surface enabled secure cooperative transmission for satellite-terrestrial integrated networks. IEEE Trans. Veh. Technol. 70(2), 2007–2011 (2021)
17. Kannan, C., et al.: Cryptography-based deep artificial structure for secure communication using IoT-enabled cyber-physical system. IET Commun. 15(6), 771–779 (2021)
18. Humayed, A., et al.: Cyber-physical systems security-A survey. IEEE Internet Things J. 4(6), 1802–1831 (2017)
19. Ding, D., et al.: Secure state estimation and control of cyber-physical systems: A survey. IEEE Trans. Syst. Man Cybern. Syst. 51(1), 176–190 (2021)
20. Liu, H., et al.: False-data-injection attacks on remote distributed consensus estimation. IEEE Trans. Cybern. (2020). https://doi.org/10.1109/TCYB.2020.2977056
21. Ye, D., et al.: Stochastic coding detection scheme in cyber-physical systems against replay attack. Inf. Sci. 481, 432–444 (2019)
22. Alanwar, A., et al.: Distributed secure state estimation using diffusion Kalman filters and reachability analysis. In: Proceedings of the 58th IEEE Conference on Decision and Control, pp. 4133–4139 (2019)
23. Lu, G., et al.: Distributed diffusion least mean square algorithm based on the reputation mechanism (in Chinese). J. Electron. Inf. Technol. 37(5), 1234–1240 (2015)
24. Liu, Y., Li, C.: Secure distributed estimation over wireless sensor networks under attacks. IEEE Trans. Aerosp. Electron. Syst. 55(4), 1815–1831 (2019)
25. Gao, Y., Ge, X.: Distributed attack detection and secure estimation of networked cyber-physical systems against false data injection attacks and jamming attacks. IEEE Trans. Signal Inf. Proc. Netw. 4(1), 48–59 (2018)
26. Chen, B., et al.: Secure fusion estimation for bandwidth constrained cyber-physical systems under replay attacks. IEEE Trans. Cybern. 48(6), 1862–1876 (2018)