2023 62nd Annual Conference of the Society of
Instrument and Control Engineers of Japan (SICE)
September 6-9, 2023. Tsu City, Japan

Distributed Secure State Estimation Against Sparse Attacks: A Sensor Clustering Approach

Xuqiang Lei[1†], Guanghui Wen[1] and Shuai Wang[2]

[1] School of Mathematics, Southeast University, Nanjing, China
(E-mail: xq-lei@seu.edu.cn, ghwen@seu.edu.cn)
[2] School of Computer Science and Engineering, Beihang University, Beijing, China
(E-mail: wangshuai@buaa.edu.cn)

Abstract: This paper investigates the problem of sparse attack isolation and distributed secure state estimation (SSE) for remote sensor network in an attack-prone environment. The attack considered in the present paper is a variant of the sparse sensor attack model, and is characterized by the number of sensors clusters that can be compromised at each moment. By analyzing the worst-case scenario in the presence of the considered attack, an attack detection and isolation mechanism based on the weighted communication design is provided to block the spread of the attack across the observer network. Moreover, based on the attack isolation, a distributed SSE algorithm is constructed to ensure globally consistent state estimation. Finally, a numerical simulation is provided to validate the effectiveness of the proposed conditions and algorithm.

Keywords: Sensor network, distributed secure state estimation, sparse attack, weighted communication.

1. INTRODUCTION

The security and reliability of cyber-physical systems (CPSs) have raised concerns within the control and information communities [1, 2]. One critical issue in this context is the secure state estimation problem for sensor network, which has gained extensive attention from researchers [3–5]. Roughly speaking, research on the SSE problems can be classified into two categories: one that explores how to optimally generate the sparse attacks at the attacker's standpoint; while the other investigates how to resist malicious attacks from the defender's point-of-view.

As a resource-constrained attacker, the priority is to create stealthy attacks that manipulate observed values to diverge from the true ones. The optimal attack strategy is one that is both hard to detect and easy to deploy, and thus far-reaching [6]. Recently, there has been a vast amount of research work that covers multiple scenarios [7–9]. Moreover, a convex relaxation algorithm was proposed in [10] to maximize the trace of the remote estimation error covariance, as well as an optimal sparse attack strategy was given. In [11], the existence and design conditions of the false data injection attack with energy stealthiness were given. Based on the system dynamics model, [12] proposed the distributed optimal algorithm for robust sparse undetectable attacks. Furthermore, to counter the SSE, two regimes of sparse undetectable and unidentifiable sensor attacks were proposed in [13]. As a result, research to explore more efficient and universal anti-attack algorithm for specific sparse attacks is still valuable.

Additionally, for the defender, one focus on developing effective attack isolation and SSE protocols that can function in partially compromised sensor measurements [14–22]. Among them, in [16–18], effective distributed SSE protocols and attack isolation algorithms were proposed for CPSs against the sparse sensor attacks, respectively. A Luenberger-like observer combined with the saturation-innovation update (SIU) idea to resist the location-varying sparse attacks could be found in [20–22]. In addition, [23] proposed a fast state estimation algorithm by reducing the sparse attack mismatch search candidates with the equivalent measurement of the sensors. However, the above foundation, which is built on more benign channels than compromised ones, may face failure as malicious attacks become more sophisticated. This paper namely aims to explore the conditions to relax this restriction and propose algorithms to improve SSE.

To combat the growing threat of attacks, researchers in [24, 25] proposed leveraging the diversity of sensor components. The concept rests on the idea that attackers can only target specific types of components and that large-scale network systems tend to have a variety of hardware and software implementations with varying weaknesses. Again, we incorporated this concept into our paper to develop the findings.

In view of the above research developments, a weaker constrained sparse sensor attack is proposed in this paper by clustering sensors in terms of the resistance to attacks. Based on the weighted interaction design of the sensor network, the convergence of the state estimation error under extreme worst-case attack scenarios is analyzed, and the construction form of the dynamic decay threshold function is proposed, which ensures accurate isolation of malicious attacks and the effective implementation of SSE.

In section 2, a general description of the notations, and the system and observer are provided. Section 3 gives the main results and analysis of this paper. Section 4 shows a numerical simulation to illustrate the effectiveness of this study and the conclusion is presented in Section 5.

† Xuqiang Lei is the presenter of this paper.

2. SYSTEM OVERVIEW

2.1. Notation and graph Description

In this paper, the $\mathbf{R}^n$ denotes the n-dimensional Euclidean space, $\mathbf{0}_n$ represent n-dimensional zero vector. The symbol $[N]$ denotes the set $\{1,\dots,N\}$. For a set of vectors $x_i \in \mathbf{R}^n$ with $i \in [N]$, the default $x = \text{col}\{x_i\} = [x_1^T,\dots,x_N^T]^T$. The notation $\|\cdot\|$ is the Euclidean vector norm, the symbol $\lfloor\cdot\rfloor$ represents the floor function. For an arbitrary set $\mathcal{S}$, $|\mathcal{S}|$ means the cardinality of $\mathcal{S}$.

For a set of matrices $M_i (i \in [N])$, similarly defined by default $M = [M_1^T,\dots,M_N^T]^T$ and $\tilde{M} = \text{Diag}\{M_i\}$ means the diagonal block matrix with M_i as the element. And $M_{\bar{\mathcal{S}}}$ denotes the matrix derived after making $M_i = 0$ with $i \in \mathcal{S}$ in M. Then for any given matrices X and Y, the Kronecker product denoted as $X \otimes Y$, the symbol X^+ is defined as the pseudo-inverse of X, and $He\{X\}$ denote $X + X^T$.

A weighted undirected graph, denoted by $\mathcal{G} = (\mathcal{V},\mathcal{E},\mathcal{A})$, consists of a nodes set $\mathcal{V} = [N]$, an edge set $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$, and a weighted adjacency matrix $\mathcal{A}$ with $a_{ij} > 0$ if $(i,j) \in \mathcal{E}$ and $a_{ij} = 0$ otherwise, in which the $(i,j) \in \mathcal{E} \Leftrightarrow (j,i) \in \mathcal{E}$ indicate that nodes i and j can transmit information to each other. The Laplacian matrix of $\mathcal{G}$ is denoted as $\mathcal{L} = [l_{ij}]_N$ with $l_{ij} = -a_{ij}$ if $i \neq j$ and $l_{ii} = \sum_{j=1}^N a_{ij}$. Beside, if graph $\mathcal{G}$ is connected, it is known that the matrix $\mathcal{L}$ is symmetric and semi-positive definite with its eigenvalues can be arranged as $0 = \lambda_1(\mathcal{L}) < \lambda_2(\mathcal{L}) \leq \cdots \leq \lambda_N(\mathcal{L})$.

2.2. System Description

Consider a sensor network designed to measure a continuous dynamic process in an attack-prone environment, where each node represents a measurement device connected through an undirected connected graph $\mathcal{G}$.

Specifically, the dynamics are modeled as

$$\dot{x}(t) = Ax(t) + Bu(t), \tag{1a}$$

$$\tilde{y}_i(t) = C_i x(t) + a_i(t), \quad i \in \mathcal{V} = [N], \tag{1b}$$

where $x(t) \in \mathbf{R}^n$, $u(t) \in \mathbf{R}^q$ are the system state and the control input, respectively. And $\tilde{y}_i(t) \in \mathbf{R}^{p_i}$ denote the measured output of the i-th sensor that could be compromised by an arbitrary attack injection $a_i(t)$. And A, B are the intra-system matrix, $C_i \in \mathbf{R}^{p_i \times n}$ is the local measurement matrix for sensor i, which satisfies itself to be column full-rank. And the global measurement matrix is expressed in the form $C = [C_1^T,\dots,C_N^T]^T$.

The sensors can be divided into two disjoint subsets, the attacked set $\mathbf{A}_t = \{i | a_i(t) \neq \mathbf{0}_{p_i}\}$ and the normal set $\mathbf{N}_t = \{i | a_i(t) = \mathbf{0}_{p_i}\}$, depending on whether compromised by the attacker. Then, the attack $a(t) = \text{col}\{a_i(t)\}$ is said to be s-sparse if one has $|\mathbf{A}_t| \leq s$, and the subset $\mathbf{A}_t$ can be unknown and time-varying.

To obtain the state estimate of system $x(t)$ securely, the following notion of s-sparse observable usually needs to be introduced.

Definition 1: The sensor network (1b) is said to be s-sparse observable, if for any subset $\Gamma \subset \mathcal{V}$ with $|\Gamma| = s$, the matrix $C_{\bar{\Gamma}}$ is column full-rank.

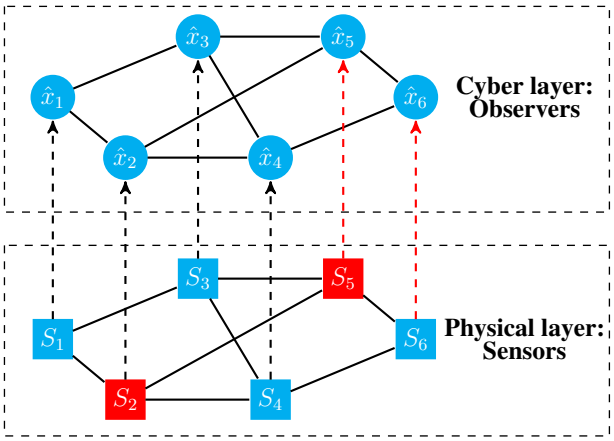

Fig. 1: An illustration example for distributed SSE.

Evidently, the sensor network (1b) is $(N-1)$-sparsely observable here. From the conclusion in [15–22], it is clear that an SSE algorithm can be performed in the case of arbitrary $\tilde{s}$ ($\tilde{s} \leq \lfloor \frac{N-1}{2} \rfloor$)-sparse sensor attacks.

Typically, a column full-rank matrix C_i indicates that the i-th observer can directly access the global system state. However, each local observer is incapable of verifying the accuracy of their sensor measurements independently, since the vulnerability of individual sensors to distortion in an attack-prone environment. It is, therefore, essential to detect and isolate sparse attacks through distributed interactions among neighbors and ensure the overall effectiveness of the global SSE.

2.3. Distributed Observers with Attack Isolation

Drawing inspiration from previous collaborative observer in [20–22], a similar distributed one-by-one observer architecture is constructed here to estimate the original state of the system accurately. Fig.1 provides a simple illustration example, in which the nodes and dotted lines colored in red, indicate the maliciously attacked sensors or wireless channels in the network. Conversely, the rest are normal.

The observer is specifically constructed as follows

$$\dot{\hat{x}}_i(t) = A\hat{x}_i(t) + Bu(t) + \theta v_i(t) + \delta_i(t)\beta L_i z_i(t), \tag{2}$$

where $\hat{x}_i(t)$ is the estimation of $x(t)$ by the i-th observer, and L_i is the observation matrix designed as $L_i = C_i^+$. Also, θ and β are gain parameters to be designed, the control $v_i(t)$ and residual $z_i(t)$ are expressed as follows

$$v_i(t) = -\sum_{j=1}^N l_{ij}\hat{x}_j(t), \quad z_i(t) = \tilde{y}_i - C_i\hat{x}_i(t).$$

Additionally, $\delta_i(t)$ indicates the isolation factor for the attack $a_i(t)$, its design is based on the variation of the SIU algorithm in [20] as follows

$$\delta_i(t) = \begin{cases} 1, & \|QL_i z_i(t)\| \leq \|Q\|\gamma(t), \\ 0, & \text{otherwise}, \end{cases} \tag{3}$$

where the matrix Q and threshold $\gamma(t)$ will be designed later. The $\delta_i(t) = 1$ signals that data $\tilde{y}_i(t)$ can be trusted for innovation, otherwise its is an anomaly that needs

to be isolated. On the basis of the successful isolation of all malicious attacks, the global convergence of the state estimation can be ensured by relying on distributed observer (2).

3. MAIN RESULT

3.1. Attacks Framework

The goal of the attacker is to generate multiple nonzero attack sequences $\{a_i(t)\}$ with limited resources, which would lead to the creation of inaccurate state estimations by the observers (e.g., (2)). Nevertheless, the heterogeneity among sensors, which have different brands and security features, limits the attacker's ability to compromise them. As a result, the attacker can usually only attack a limited selection of sensors, generally only a few types at a time.

Sensor Clusters: Inspired by the above thought, sensors are planned into m distinct clusters based on their vulnerability against attack, and each sensor is uniquely mapped to the index set $\mathcal{M} = \{c_1, \ldots, c_m\}$. Consequently, the entire sensor set can be re-described as $\mathcal{V} = \{\mathcal{V}_{c_1}, \ldots, \mathcal{V}_{c_m}\}$.

Assumption 1: Each observer $i \in \mathcal{V}_{c_l}$ ($l \in [m]$) is aware of the number of sensor types m and the number $r_i = |\mathcal{V}_{c_l}|$ of nodes that are the same cluster as it.

Remark 1: In order to satisfy Assumption 1 holds, which can usually be assigned at the time of initial network deployment, the design can be offline.

Then, with the above assumptions, each observer could broadcasts its message r_i to the neighbors, and thus the communication weight $a_{ij} = 1/r_j$ can be set here if $(j, i) \in \mathcal{E}$. Further, for the Laplacian matrix $\mathcal{L}$ constructed as above, one has $g^T\mathcal{L} = \mathbf{0}_N^T$ with $g^T = \frac{1}{|\mathcal{V}|}[1/r_1, \cdots, 1/r_i, \cdots, 1/r_N]$ established.

Definition 2: The attacks signal $a(t) = \mathrm{col}\{a_i(t)\}$ is weak s-sparse, if the number of clusters destroyed by the attack at any given moment is at most s.

Remark 2: In this setting, all sensors within a cluster are equated to one, thus expecting to improve the tolerance of the SSE algorithm to sparse attacks in response to resource-constrained adversaries.

3.2. Threshold Function Design

In this subsection, a suitable threshold function $\gamma(t)$ is designed based on the worst-case estimation error analysis, such that it could enables the accurate detection of malicious attacks while maintaining the benign operation of state estimation.

Denote the estimation error as $e_i(t) = \hat{x}_i(t) - x(t)$, $i \in \mathcal{V}$, and assume that the initial estimation error is upper bounded by η, i.e., $\max_{i\in\mathcal{V}}\{\|e_i(0)\|\} \leq \eta$. Then, consider the threshold function designed as $\gamma(t) = \gamma_1(t) + \gamma_2(t)$, with $\gamma_1(t)$ and $\gamma_2(t)$ constructed as follows

$$\begin{aligned}\dot{\gamma}_1(t) &= (\alpha - \theta\lambda_2^{\mathcal{L}} + 2\sqrt{N}\beta)\gamma_1(t) + 2\sqrt{N}\beta\gamma_2(t),\\ \dot{\gamma}_2(t) &= [\alpha - (1 - s_f)\beta]\,\gamma_2(t) + s_f\beta\gamma_1(t),\end{aligned} \tag{4}$$

where $s_f = 2s/m$, the initial values are picked to satisfy $\gamma_1(0) = \sqrt{N}\eta$ and $\gamma_2(0) = \eta$. The parameter $\alpha \geq 0$ and matrix $P = Q^TQ > 0$ are the solution to the following generalized eigenvalue minimization problem:

$$\begin{aligned}&\min\ \alpha \\ &\text{s.t. } \frac{1}{2}(PA + A^TP) < \alpha P,\end{aligned} \tag{5}$$

which can be solved by the 'gevp' solver for LMI.

The gains $\theta > 0$ and $\beta > 0$ are designed to satisfy the following LMI holds:

$$He\left\{\begin{bmatrix}\alpha - \lambda_2^{\mathcal{L}}\theta + 2\sqrt{N}\beta & 2\sqrt{N}\beta \\ s_f\beta & \alpha - (1 - s_f)\beta\end{bmatrix}\right\} < 0. \tag{6}$$

Remark 3: From the LMI (6), it is evident that the system (4) is asymptotically stable, i.e., $\lim_{t\to\infty}\gamma(t) = 0$. Following that, with the decreasing of $\gamma(t)$, the undetectable space of malicious attacks will gradually diminish, so that the detection and isolation of all attacks can be ensured.

3.3. Distributed secure state estimation

In this subsection, the upper bound of the attack constraint for reaching an asymptotically estimation of the original state of the system is analyzed and its theoretical proof is briefly given. Finally, based on the synthesis analysis, an algorithm design for distributed SSE under weak s-sparse attacks is provided.

The update of the estimation error $e_i(t)$ can be obtained from (1)–(2) as

$$\dot{e}_i(t) = Ae_i(t) - \theta\sum_{j=1}^{N} l_{ij}e_j(t) + \beta\delta_i(t)L_iz_i(t). \tag{7}$$

Furthermore, the following error variables are introduced to analyze the estimation performance:

$$\tilde{e}_i(t) = e_i(t) - \bar{e}(t), \quad \bar{e}(t) = (g^T \otimes I_n)e(t),$$

where $e(t) = \mathrm{col}_{i\in\mathcal{V}}\{e_i(t)\}$.

Denote $\mathcal{I} = I_N - \mathbf{1}_N g^T$ and $\tilde{\mathcal{I}} = \mathcal{I} \otimes I_n$, one has $\mathcal{I}\mathcal{L} = \mathcal{L}\mathcal{I} = \mathcal{L}$ and $\tilde{e}(t) = \mathrm{col}_{i\in\mathcal{V}}\{\tilde{e}_i(t)\} = \tilde{\mathcal{I}}e(t)$. Hence, the following dynamics can be obtained:

$$\dot{\tilde{e}}(t) = [I_N \otimes A - \theta(\mathcal{L} \otimes I_n)]\,\tilde{e}(t) + \beta\tilde{\mathcal{I}}\tilde{\Delta}_t\tilde{L}\tilde{Z}(t), \tag{8}$$

$$\begin{aligned}\dot{\bar{e}}(t) = (A - \beta I_n)\bar{e}(t) &+ \beta\sum_{i\in\mathbf{N_t}}(1 - \delta_i(t))g_ie_i(t) \\ &+ \beta\sum_{i\in\mathbf{A_t}} g_i(e_i(t) + \delta_i(t)L_iz_i(t)),\end{aligned} \tag{9}$$

where $\tilde{\Delta}_t = \mathrm{diag}\{\delta_i(t)\} \otimes I_n$ and $\tilde{Z}(t) = \mathrm{col}_{i\in\mathcal{V}}\{z_i(t)\}$, and g_i denotes the i-th element in vector g.

Theorem 1: Under Assumption 1, based on the threshold function and observation gains designed in (4)-(6). Then, the distributed observer (2) can reach the SSE for system (1), i.e.,

$$\lim_{t\to\infty}\|\hat{x}_i(t) - x(t)\| = 0, \tag{10}$$

if the malicious sensor attack satisfies is weak s-sparse with $s < m/2$.

Algorithm 1 Clusters-based distributed SSE algorithm

Input: $A,\ B,\ C_i,\ u(t),\ \tilde{y}_i(t),\ \hat{x}_i(0),\ s$;

Initialization: Assigning sensors to m clusters, and obtaining inter-neighborhood weights a_{ij};
Solve (5) to obtain the feasible solution α, P and Q;
Compute the LMI (6) to obtain the gains θ, β;
while $t > 0$ **do**
    Communicate with its neighbors to exchange their respective state estimates $\hat{x}_i(t)$
    Update the threshold function $\gamma(t)$ as (4);
    Calculate the residual $z_i(t) = \tilde{y}_i(t) - C_i\hat{x}_i(t)$;
    if $\|QL_iz_i(t)\| \leq \|Q\|\gamma(t)$ **then**
        $\delta_i(t) = 1$;
    else
        $\delta_i(t) = 0$;
    end if
    Update the state estimation $\hat{x}_i(t)$ as in (2);
end while

Output: State estimate $\hat{x}_i(t)$;

Proof: For the error system (8) and (9), consider the following Lyapunov functions:

$$V_1(t) = \|(I_N \otimes Q)\tilde{e}(t)\|, \quad V_2(t) = \|Q\bar{e}(t)\|.$$

Firstly, from (3), there is $\delta_i(t)\|QL_iz_i(t)\| \leq \|Q\|\gamma(t)$ holds. Hence, take the differential along (8) and (9) as follows:

$$\begin{aligned}\dot{V}_1(t) &= V_1^{-1}(t) \cdot \tilde{e}^T(t)\tilde{P}\dot{\tilde{e}}(t)\\ &\overset{(a)}{\leq} (\alpha - \theta\lambda_2^{\mathcal{L}})V_1(t) + 2\beta\|\tilde{\Delta}_t\tilde{L}\tilde{Z}(t)\|\\ &\leq (\alpha - \theta\lambda_2^{\mathcal{L}})V_1(t) + 2\sqrt{N}\beta\|Q\|\gamma(t),\\ \dot{V}_2(t) &\overset{(b)}{\leq} (\alpha - \beta)V_2(t) + \beta s_f\|Q\|\gamma(t),\end{aligned}$$

where $\tilde{P} = I_N \otimes P$, and the inequation (a) is deduced due to the fact that $\|\mathcal{I}\| < 2$, $x^TPy \leq \|Qx\|\|Qy\|$. Inequation (b) relies on $\delta_i(t) = 1$ holds for all $i \in \mathbf{N}_t$, which will be shown in the subsequent proof.

Combining (4) with the fact that $V_1(0) \leq \sqrt{N}\eta = \gamma_1(0)$ and $V_2(0) \leq \eta = \gamma_2(0)$, it follows $V_1(t) \leq \gamma_1(t)$ and $V_2(t) \leq \gamma_2(t)$ hold for $\forall t \geq 0$. Hence, from $e_i(t) = \tilde{e}_i(t) + \bar{e}(t)$, one has

$$\|Qe_i(t)\| \leq \|(I_N \otimes Q)\tilde{e}(t)\| + \|Q\bar{e}(t)\| \leq \|Q\|\gamma(t).$$

holds for $\forall i \in \mathcal{V}$.

Further, for $\forall i \in \mathbf{N}_t$, it follows that $\|QL_iz_i(t)\| = \|Qe_i(t)\| \leq \|Q\|\gamma(t)$. Hence, one has $\delta_i(t) = 1$ holds for all $i \in \mathbf{N}_t$.

In conclusion, from the fact that $\lim_{t\to\infty}\gamma(t) = 0$, it follow that $\lim_{t\to\infty}\|e_i(t)\| = 0$ hold for any $i \in \mathcal{V}$. ■

Based on the above analysis and proof, it can be concluded that each local observation node needs to perform the operations of the following Algorithm 1 in parallel.

Remark 4: In short, a sensor network with multiple heterogeneous brands can enhance the resistance of SSE algorithm to sparse attacks, which mainly builds on the intuition that an attacker can launch attacks more easily against sensors that have been informed of the vulnerabilities. It is worth noting that Algorithm 1 remains effective in detecting any form of weak s-sparse attack ($s < m/2$), as long as the sensor network satisfies Assumption 1.

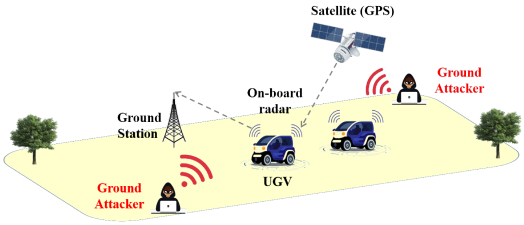

Fig. 2: An example of cyber attacks on UGV.

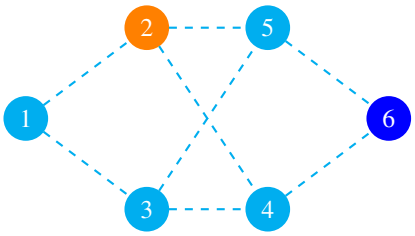

Fig. 3: Network topology with 6 nodes and 3 clusters.

4. SIMULATION EXAMPLE

In this section, to validate the effectiveness of the algorithm, an example is borrowed from [5, 12] regarding the distributed SSE of an Unmanned Ground Vehicle (UGV is seen in Fig.2). Specifically, the dynamics of the UGV is described as follows:

$$\begin{bmatrix}\dot{\chi}\\ \dot{v}\\ \dot{\theta}\\ \dot{\omega}\end{bmatrix} = \begin{bmatrix}0 & 1 & 0 & 0\\ 0 & -\frac{B}{M} & 0 & 0\\ 0 & 0 & 0 & 1\\ 0 & 0 & 0 & \frac{B_r}{J}\end{bmatrix}\begin{bmatrix}\chi\\ v\\ \theta\\ \omega\end{bmatrix} + \begin{bmatrix}0 & 0\\ \frac{1}{M} & 0\\ 0 & 0\\ 0 & \frac{1}{J}\end{bmatrix}\begin{bmatrix}F\\ T\end{bmatrix},$$

where $x = [\chi,\ v,\ \theta,\ \omega]^T$ represent the state of the position, linear velocity, angular position and angular velocity, respectively. The parameters $M = 2,\ J = 2,\ B = 1$, and $B_r = 1$ shows the mechanical mass, inertia, translational friction coefficient and the rotational friction coefficient. The inputs are the force $F = 2sin(2t)$ and the torque $T = 1$. Then, six sensors from each of the 3 different sensor clusters are randomly selected to satisfy the C_i column full rank, and communicate based on the topology shown in Fig.3. Consider the initial values picked as $x(0) = [0.5,\ 4,\ 0.5,\ 7]^T$ in the simulation.

In order to overcome the weak s-sparse attack described in 3.1, we performed simulations using the proposed Algorithm 1 in this paper with initial estimates randomly selected. In particular, during the runs of the algorithm, a malicious adversary is considered to inject a randomly generated malicious attack vector into a randomly selected cluster in the sensors every 1 seconds, and then the simulation results are obtained as shown below.

Taking advantage of the 'gevp' solver for LMI in the MATLAB robust toolbox, one obtain $\alpha = 5.4887 \times 10^{-8}$

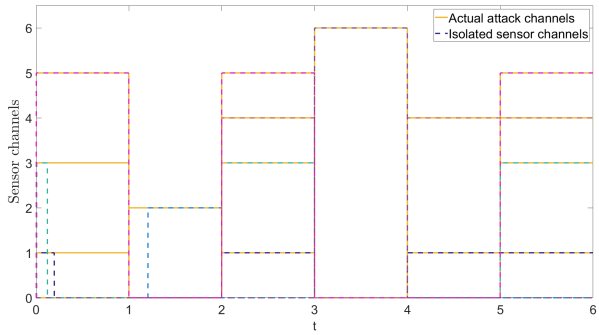

Fig. 4: The actual attack injection channels and the sensor channels isolated by the observer.

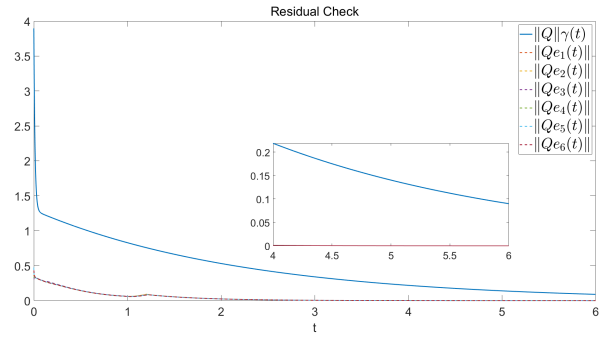

Fig. 5: The norms $\|Qe_i(t)\|$ and the threshold $\|Q\|\gamma(t)$.

and the following parameters:

$$Q = \begin{bmatrix} 0.0173 & 0.0010 & 0 & 0 \\ 0.0001 & 0.5736 & 0 & 0 \\ 0 & 0 & 0.0173 & 0.0010 \\ 0 & 0 & 0.0010 & 0.5736 \end{bmatrix}.$$

Next, the gain $\theta = 194.013$ and $\beta = 1.694$ are obtained by solving the LMI (6). The specific observation estimation and attack detection and isolation results can be found in Fig.4–7. Specifically, to illustrate the effectiveness of detection and isolation against weak f-sparse attacks, the sensor channels injected by the attack at each moment and the channels isolated by the defender are presented in Fig.4. Clearly, it can be seen that Algorithm 1 can effectively handle weak f-sparse attacks regardless of any target modification made by the attacker. Thus, in an attack-prone environment, the defender can enhance the anti-attack capability of the algorithm 1 by adding new sensor types.

The relationship regarding the threshold function $\|Q\|\gamma(t)$ and the detection term $\|Qe_i(t)\|$ are shown in Fig.5. It can be seen that the benign residual term never exceeds the threshold function, and thus the measurement data over the threshold can be identified as corrupted by the attacker. Moreover, as the threshold function converges to zero asymptotically, the survival space of malicious attacks will be continuously compressed, so that all non-zero attacks will eventually be detected and isolated.

Finally, the real-time path trajectory of the UGV and the trajectory generated by the estimation of each observer are given in Fig.6, and the estimation error corresponding to each state is shown in Fig.7. It can be seen from this that the proposed Algorithm 1 is effective against weak f-sparse attacks. At this point, the number of malicious attacks may exceed half of the number of sensors (e.g, in this example, the maximum number of attacks could be $4 > (6-1)/2$).

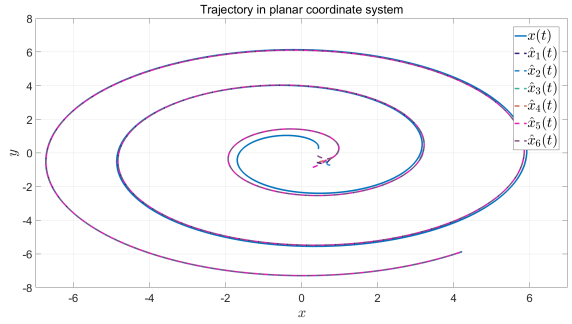

Fig. 6: System dynamics trajectory and its estimation.

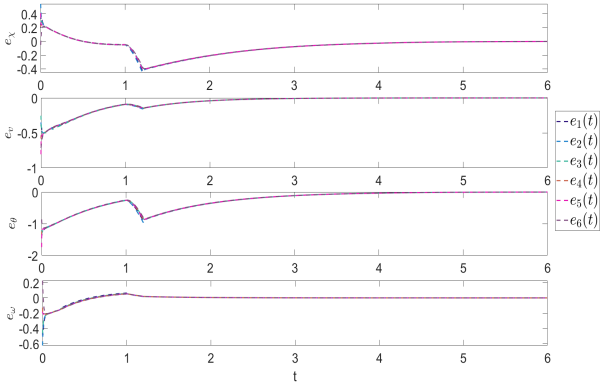

Fig. 7: State estimation error $e_i(t)$.

5. CONCLUSION

In this paper, a distributed SSE algorithm that relies on sensor clustering approach has been presented for an attack-prone sensor network. A weak s-sparse attack has been proposed by classifying the resistance of sensors to attacks. Then, utilizing the above classification, a weighted communication topology has been constructed and an attack detection threshold has been designed based on it. Finally, a distributed SSE algorithm design has been proposed under the worst-case attack consideration. In future work, the malicious attack model and the sensor model constraints will be further improved in order to expect the secure state estimation algorithm design to be completed under global observable condition.

ACKNOWLEDGEMENT

The authors express their gratitude to Dr. Dan Zhao (School of Cyber Science and Engineering) from Southeast University for helpful discussions. This work was supported in part by the National Natural Science Foundation of China through Grant Nos. U22B2046, 62073079 and 62088101, and in part by the General Join-

t Fund of the Equipment Advance Research Program of Ministry of Education under Grant No. 8091B022114.

REFERENCES

[1] S. Sridhar, A. Hahn, and M. Govindarasu, "Cyber–physical system security for the electric power grid," *Proceedings of the IEEE*, Vol. 100, No. 1, pp. 210–224, 2012.

[2] G. Wen, P. Wang, Y. Lv, G. Chen, and J. Zhou, "Secure consensus of multiagent systems under denialofservice attacks," *Asian Journal of Control*, Vol. 25, No. 2, pp. 695-709, 2023.

[3] H. Fawzi, P. Tabuada, and S. Diggavi, "Secure estimation and control for cyber-physical systems under adversarial attacks," *IEEE Transactions on Automatic control*, Vol. 59, No. 6, pp. 1454–1467, 2014.

[4] F. Pasqualetti, F. Dorfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Transactions on Automatic Control*, Vol. 58, No. 11, pp. 2715–2729, 2013.

[5] Y. Shoukry and P. Tabuada, "Event-triggered state observers for sparse sensor noise/attacks," *IEEE Transactions on Automatic Control*, Vol. 61, No. 8, pp. 2079–2091, 2016.

[6] Y. Zhao, A. Goldsmith, and H. V. Poor, "Minimum sparsity of unobservable power network attacks," *IEEE Transactions on Automatic Control*, Vol. 62, No. 7, pp. 3354–3368, 2017.

[7] Y. Mo and B. Sinopoli, "On the performance degradation of cyber-physical systems under stealthy integrity attacks," *IEEE Transactions on Automatic Control*, Vol. 61, No. 9, pp. 2618–2624, 2016.

[8] Z. Zhao, Y. Yang, Y. Li, and R. Liu, "Security analysis for cyber-physical systems under undetectable attacks: A geometric approach," *International Journal of Robust and Nonlinear Control*, Vol. 30, No. 11, pp. 4359–4370, 2020.

[9] L. Zhang, Y. Chen, and M. Li, "ADP-based remote secure control for networked control systems under unknown nonlinear attacks in sensors and actuators," *IEEE Transactions on Industrial Informatics*, Vol. 18, No. 9, pp.6003–6014, 2022.

[10] F. Li and Y. Tang, "False data injection attack for cyber-physical systems with resource constraint," *IEEE Transactions on Cybernetics*, Vol. 50, No. 2, pp. 729–738, 2020.

[11] T.-Y. Zhang and D. Ye, "False data injection attacks with complete stealthiness in cyber–physical systems: A self-generated approach," *Automatica*, Vol. 120, p. 109117, 2020.

[12] L. An and G.-H. Yang, "Distributed sparse undetectable attacks against state estimation," *IEEE Transactions on Control of Network Systems*, Vol. 9, No. 1, pp. 463–473, 2021.

[13] A. Lu and G.-H. Yang, "Malicious adversaries against secure state estimation: Sparse sensor attack design," *Automatica*, Vol. 136, p.110037, 2022.

[14] Y. Jeong and Y. Eun, "A robust and resilient state estimation for linear systems," *IEEE Transactions on Automatic Control*, Vol. 67, No. 5, pp. 2626–2632, 2022.

[15] N. R. Chowdhury, J. Belikov, D. Baimel, and Y. Levron, "Observer-based detection and identification of sensor attacks in networked CPSs," *Automatica*, Vol. 121, p. 109166, 2020.

[16] A. Lu and G.-H. Yang, "Secure switched observers for cyber-physical systems under sparse sensor attacks: A set cover approach," *IEEE Transactions on Automatic Control*, Vol. 64, No. 9, pp. 3949–3955, 2019.

[17] Y. Shoukry, P. Nuzzo, A. Puggelli, A. Sangiovannivincentelli, S. A. Seshia, and P. Tabuada, "Secure state estimation for cyber-physical systems under sensor attacks: A satisfiability modulo theory approach," *IEEE Transactions on Automatic Control*, Vol. 62, No. 10, pp. 4917–4932, 2017.

[18] L. An and G.-H. Yang, "Distributed secure state estimation for cyber-physical systems under sensor attacks," *Automatica*, Vol. 107, pp. 526–538, 2019.

[19] T. Shinohara, T. Namerikawa, and Z. Qu, "Resilient reinforcement in secure state estimation against sensor attacks with a priori information," *IEEE Transactions on Automatic Control*, Vol. 64, No. 12, pp. 5024–5038, 2019.

[20] Y. Chen, S. Kar, and J. M. Moura, "Resilient distributed estimation: Sensor attacks," *IEEE Transactions on Automatic Control*, Vol. 64, No. 9, pp. 3772–3779, 2019.

[21] X. He, X. Ren, H. Sandberg, and K. H. Johansson, "How to secure distributed filters under sensor attacks," *IEEE Transactions on Automatic Control*, Vol. 67, No. 6, pp. 2843–2856, 2022.

[22] X. He, E. Hashemi, and K. H. Johansson, "Distributed control under compromised measurements: Resilient estimation, attack detection, and vehicle platooning," *Automatica*, Vol. 134, p. 109953, 2021.

[23] L. An and G.-H. Yang, "Fast state estimation under sensor attacks: A sensor categorization approach," *Automatica*, Vol. 142, p. 110395, 2022.

[24] F. Ghawash and W. Abbas, "Leveraging diversity for achieving resilient consensus in sparse networks," *IFAC-PapersOnLine*, Vol. 52, No. 20, pp. 339–344, 2019.

[25] A. Mitra, F. Ghawash, S. Sundaram, and W. Abbas, "On the impacts of redundancy, diversity, and trust in resilient distributed state estimation," *IEEE Transactions on Control of Network Systems*, Vol. 8, No. 2, pp.713–724, 2021.