Probability-Guaranteed Distributed Secure Estimation for Nonlinear Systems over Sensor Networks under Deception Attacks on Innovations

Lifeng Ma, Zidong Wang, Yun Chen, and Xiaojian Yi

Abstract—This paper studies the distributed secure state estimation problem for a class of general nonlinear systems over sensor networks under unknown deception attacks on innovations. At each sensing node, an estimator is designed to generate the state estimate by making use of the local measurements in combination with the neighbours' information shared via the communication network. During the transmission of innovations among nodes, the data are maliciously falsified by adversaries in a random way. A neural-network-based mechanism is put forward to approximate the unknown falsified innovations with the aim to mitigate the effects on the estimation performance. The objective of the addressed problem is to develop a distributed estimation approach to jointly estimate the system states as well as the unknow deception attacks, ensuring that the state estimation errors at each sensing node reside within required ellipsoidal regions with a pre-specified probability. With the help of certain convex optimization methods, we obtain sufficient conditions for the solvability of the addressed problem and the desired estimator gains can be iteratively computed by solving a series of matrix inequalities. On basis of the proposed framework, some optimization problems are presented to determine sub-optimal estimator parameters from different perspectives. Finally, the applicability of the developed algorithms is validated via a numerical simulation example.

Index Terms—Distributed estimation, set-membership state estimation, deception attack, falsified innovations, probability-guaranteed estimation, neural networks

I. INTRODUCTION

The past decades have witnessed an exponential growth of the deployments of sensor networks in various fields such as military surveillance and environment detection [6], [13], [29], [31], [32], [45], [46], [48]. A typical sensor network comprises a group of spatially dispersed sensing nodes with basic capabilities of sensing, computing and communicating. To date, considerable research effort has been devoted to sensor networks from many different perspectives, and the

L. Ma is with the School of Automation, Nanjing University of Science and Technology, Nanjing 210094, China. (Email: `malifeng@njust.edu.cn`)

Z. Wang is with the Department of Computer Science, Brunel University London, Uxbridge, Middlesex, UB8 3PH, United Kingdom. (Email: `Zidong.Wang@brunel.ac.uk`)

Y. Chen is with the School of Automation, Hangzhou Dianzi University, Hangzhou 310018, China. (Email: `yunchen@hdu.edu.cn`)

X. Yi is with the School of Mechatronical Engineering, Beijing Institute of Technology, Beijing 100081, China. (Email: `yixiaojianbit@sina.cn`)

distributed state estimation problem has stirred particular interest, see e.g. [10]–[12], [19], [24], [34], [41] for some recent works. Compared with the traditional single-sensor systems, the sensor networks could collect more information in a comprehensive and complementary way via the cooperation among individual nodes, thereby making the corresponding estimation algorithms more robust, accurate and flexible. As is widely recognized, in the context of sensor networks, the scalability issue and the computational complexity are two major concerns for the conventional *centralized* estimation algorithms, and this gives rise to the necessity of developing the so-called distributed state estimation/filtering problem where each individual sensing node provides the state estimates for the targets of interest through making use of the local measurements in combination with the information shared by neighbours [4], [5], [15], [25], [26], [30], [37], [39].

So far, a great number of research results concerning distributed state estimation have been available in the literature, among which the most popular techniques are Kalman filtering [14], [18], [20], [23], [30] and H_∞ approach [7], [33], [37]. Note that the performance of Kalman filtering largely depends on whether the noises meet the Gaussian assumption [28], [35], while the H_∞ approach is only applicable for handling the so-called energy-bounded disturbances [21]. In many practical scenarios, unfortunately, neither Gaussian nor energy-bounded assumptions could adequately reflect the noise properties. For instance, in electronics and electrical engineering, the disturbance caused by the man-made electromagnetic interference is more suitable to be modeled by the so-called unknown-but-bound (UBB) noise [1], [8]. By now, several approaches have been exploited to study the state estimation issue against UBB noises (see e.g. [1], [3], [8], [16], [25], [47]), most of which are based on the linear matrix inequality approach and its variants in a *set-membership* framework.

Aside from the exogenous noises, another factor that comes with the use of open yet unprotected communication networks is the malicious threats from adversaries. It has now been widely recognized that cyber-attacks against the networks are prevalent that behave as a major source of the performance deterioration or even the system crash [9], [43], [44]. In the context of sensor networks, it is worth noting that a large part of malicious attacks against the target systems are performed by injecting specific signals into the intercepted data during the information transmission. By utilizing the falsified information, the adversaries are able to deceive the receivers, thereby deteriorating or manipulating system performances

[17]. As the reliability and safety are among the fundamental requirements of communication networks, the security-relevant issues have provoked an ever-growing attention, see, e.g. [17], [22], [42] for some recent publications. Most existing results, however, have been based on the assumption that the falsified signals are known and linear, and such an assumption is often unrealistic in practical scenarios. As such, in this paper, we make dedicated efforts to handle *unknown* deception attacks on the innovations by proposing a novel algorithm for the distributed state estimation problem with the aid of artificial neural networks (NN).

In practical engineering, it is often unnecessary (and unattainable) to design controllers/estimators which persistently achieve the desired performance indices with a 100% confidence level. Instead, one would prefer to follow the so-called probability-guaranteed design principle which aims at reaching the desired indices with a satisfactory chance (less than 1) [36], [40]. Such a concept could find wide applications in various engineering branches, for example, the target tracking issue and weapon shooting tests [27]. The advantage of such a concept lies in its capability of ensuring satisfactory performance while i) avoiding unnecessarily stringent requirements and ii) leaving additional design freedom for other system specifications.

In response to the above discussions, this paper aims to develop a probability-guaranteed distributed state estimation approach over sensor networks where the communications among nodes are under the threat of malicious manipulations. Artificial neural networks are utilized to approximate the unknown falsified innovations during data propagation among nodes. This would be a very challenging problem due to some essential difficulties summarized as follows. 1) For the *unknown* falsified innovations, it is invariably arduous to model the dynamical characteristics via a neural network and yet facilitate the subsequent development. 2) The cross coupling between the desired estimator parameters and the NN weight matrices brings considerable difficulties in analysis and design. 3) The joint estimation of the system state and falsified innovations brings extra challenges when determining the estimator gains and NN weight matrices. Consequently, it is our main purpose to tackle these challenges by investigating the addressed probability-guaranteed secure distributed state estimation issue.

The contributions of this paper can be identified as follows:*i) the system under consideration is modeled by a nonlinear difference equation, which is much more general than those studied in literature and could reflect engineering practice in a more comprehensive way; ii) in order to depict appropriately dynamical behaviors over a finite time interval of interest, a new transient performance index is defined for system at each time step; iii) with the purpose of avoiding unnecessarily stringent requirements, a probabilistic design principle is put forward aiming to achieve desired performances with a satisfactory chance; and iv) a joint estimation algorithm is provided to simultaneously estimate the system state and falsified innovations.*

Notation $\mathbb{R}^n$ denotes the n-dimensional Euclidean space and $\mathbf{1}_n$ denotes an n-dimensional column vector with all

ones. I_n denotes the identity matrix of n dimensions. The notation $A \geq B$ (respectively $A > B$), where A and B are symmetric matrices, means that $A - B$ is positive semi-definite (respectively positive definite). For matrices X and Y, their Kronecker product is denoted as $X \otimes Y$. For a vector x, $\|x\| = x^\mathrm{T}x$. For a matrix Z, $\mathrm{tr}[Z]$ means the trace of matrix Z, and $\|Z\|_F = \sqrt{\mathrm{tr}[Z^\mathrm{T}Z]}$ is the Frobenius norm of Z. $\mathrm{diag}\{X_1, X_2, \ldots, X_n\}$ denotes a block diagonal matrix whose diagonal blocks are given by $X_1, X_2, \ldots, X_n$. The notation $\mathrm{diag}_n\{X_i\}$ represents the block diagonal matrix $\mathrm{diag}\{X_1, X_2, \ldots, X_n\}$ and $\mathrm{col}_n\{x_i\}$ denotes the column vector $[x_1^\mathrm{T} \ x_2^\mathrm{T} \ \ldots \ x_n^\mathrm{T}]^\mathrm{T}$. $\mathbb{P}\{X\}$ means the occurrence probability of the event 'X'.

II. PROBLEM FORMULATION

The sensor network under consideration consists of N sensors with the communication topology described by a directed graph $\mathscr{G} = (\mathscr{V}, \mathscr{E}, \mathscr{L})$, where $\mathscr{V} = \{1, 2, ..., N\}$ represents the set of sensing nodes, $\mathscr{E} \subseteq \mathscr{V} \times \mathscr{V}$ represents the set of edges, and $\mathscr{L} = [\theta_{ij}]_{N \times N}$ represents the nonnegative adjacency matrix. Specifically, $\theta_{ij} > 0$ means that sensor j can send information to sensor i (in such a case, j is called a neighbour of i), while $\theta_{ij} = 0$ indicates there are no data transmitted from node j to node i. Denote by $\mathscr{N}_i \triangleq \{j \in \mathscr{V} | (i, j) \in \mathscr{E}\}$ the set of neighbours of node i. Moreover, we assume that $\theta_{ii} = 0$ for all $i \in \mathscr{V}$.

Consider the following nonlinear system defined on the horizon $[0, T]$:

$$\begin{cases} x_{k+1} = f(x_k) + B_k \nu_k \\ y_{i,k} = g_i(x_k) + E_{i,k} \mu_k \end{cases} \quad (1)$$

where $x_k \in \mathbb{R}^{n_x}$ is the system state; $y_{i,k} \in \mathbb{R}^{n_y}$ is the measurement output of sensor node i; $\nu_k \in \mathbb{R}^{n_\nu}$ and $\mu_k \in \mathbb{R}^{n_\mu}$ are the process and measurement disturbances; B_k and $E_{i,k}$ are known real-valued matrices of compatible dimensions; $f(x_k) : \mathbb{R}^{n_x} \mapsto \mathbb{R}^{n_x}$ and $g_i(x_k) : \mathbb{R}^{n_x} \mapsto \mathbb{R}^{n_y}$ are smooth nonlinear functions.

Assumption 1: The noises ν_k and μ_k satisfy the following conditions:

$$\begin{cases} \nu_k \in \mathcal{V}_k \triangleq \{\nu_k : \nu_k^\mathrm{T} V_k^{-1} \nu_k \leq 1\} \\ \mu_k \in \mathcal{U}_k \triangleq \{\mu_k : \mu_k^\mathrm{T} U_k^{-1} \mu_k \leq 1\} \end{cases} \quad (2)$$

where $V_k > 0$ and $U_k > 0$ are known matrices with suitable dimensions.

Before giving the structure of distributed NN-based state estimator to be designed, we first recall the form of the *traditional* distributed estimators. For each node i ($i = 1, 2, \ldots, N$), the local state estimator has the following form:

$$\hat{x}_{i,k+1} = F_{i,k}\hat{x}_{i,k} + \sum_{j \in \mathscr{N}_i} \theta_{ij} H_{ij,k}(y_{j,k} - g_j(\hat{x}_{j,k}))$$

$$\triangleq F_{i,k}\hat{x}_{i,k} + \sum_{j \in \mathscr{N}_i} \theta_{ij} H_{ij,k} z_{j,k} \quad (3)$$

where $\hat{x}_{i,k}$ represents the state estimate of node i; $z_{i,k} \triangleq y_{i,k} - \hat{y}_{i,k}$ is the innovation information with $\hat{y}_{i,k} \triangleq g_i(\hat{x}_{i,k})$ being the estimated output; and the matrices $F_{i,k}$ and $H_{ij,k}$ are estimator gains.

The most essential feature of a distributed estimation algorithm is that each local estimator generates state estimates by utilizing its own measurements in combination with neighbors' information shared via communication networks according to the interaction topology. Such an appealing feature would be jeopardized with deteriorated estimation performance in the case that the networks are under the threat of malicious manipulations, and this gives rise to the motivation for us to consider the secure estimation issue in this paper.

Consider the information propagation process from node j to node i. The adversary first intercepts the innovation information sent by node j (i.e., $z_{j,k}$). Then, by using the intercepted $z_{j,k}$, the adversary generates a signal $z_{j,k}^{(a)}$ which will be utilized to inject into the original innovation. The injected signal $z_{j,k}^{(a)}$ is assumed to have the following form:

$$z_{j,k}^{(a)} = \chi(z_{j,k}) \tag{4}$$

where $\chi(\cdot) : \mathbb{R}^{n_y} \mapsto \mathbb{R}^{n_y}$ is an unknown nonlinear function defined on a compact set.

Taking (4) into account, the falsified innovation sent from node j (after being manipulated) is described as follows:

$$\tilde{z}_{j,k} = z_{j,k} + \alpha_k z_{j,k}^{(a)} = z_{j,k} + \alpha_k \chi(z_{j,k}) \tag{5}$$

where α_k is a Bernoulli distributed random sequence satisfying $\text{Prob}\{\alpha_k = 1\} = \bar{\alpha}$. Subsequently, the adversary will continue to send such a manipulated message (5) to node i with the aim to deteriorate the system performance.

Remark 1: It should be mentioned that, due to the complexity of sensor networks, the adversary cannot perform the attacks successfully all the time but has certain limitations on ability. The Bernoulli distributed random variable α_k is introduced to describe the constraints or limitations imposed on the ability of the adversary. In other words, the innovation will be falsified with a certain success rate.

On the other hand, at node i, the local estimator tries to provide the state estimate by using local measurements as well as the innovation information sent by its neighbours $j \in \mathcal{N}_i$. Taking (5) into account, we shall use the neural networks to approximate the unknown function $\chi(z_{j,k})$ with the hope to mitigate the effects from the falsified innovations as follows:

$$\chi(z_{j,k}) = W_j \phi(z_{j,k}) + \delta_{j,k} \tag{6}$$

where W_j is the ideal weight matrix of the neural network, $\phi(\cdot)$ is the activation function and $\delta_{j,k}$ is the approximation error.

Assumption 2: [38] The ideal weight matrix W_j, the activation function $\phi(\cdot)$ and the approximation error $\delta_{j,k}$ satisfy:

$$\|W_j\|_F \leq \epsilon_{1j}, \quad \|\phi(\cdot)\| \leq \epsilon_2, \quad \|\delta_{j,k}\| \leq \epsilon_{3j} \tag{7}$$

where ϵ_{1j}, ϵ_2 and ϵ_{3j} are known positive scalars.

We are now in the position to present the structure of the local estimator by taking into account the falsified innovations. In this paper, the local estimator on node i is reconstructed as follows:

$$\hat{x}_{i,k+1} = F_{i,k}\hat{x}_{i,k} + G_{i,k}(y_{i,k} - \hat{y}_{i,k})$$

$$+ \sum_{j \in \mathcal{N}_i} \theta_{ij} H_{ij,k}\left(\tilde{z}_{j,k} - \bar{\alpha}\hat{W}_{ij,k}\phi(z_{i,k})\right) \tag{8}$$

where $F_{i,k}$, $G_{i,k}$ and $H_{ij,k}$ are estimator parameters to be designed; $\hat{W}_{ij,k}$ is the estimate of the weight matrix W_{ij} which is defined as the ideal weight matrix for W_j at sensing node i.

Remark 2: It is worth noting that the innovations received by node i are not the original ones but containing the injected signals, which implies that the true value of $z_{j,k}$ are not available on the side of node i. Therefore, in (8), instead of $z_{j,k}$, we shall utilize the local innovation $z_{i,k}$ as the input of the neural networks to approximate the injected signal $\chi(z_{j,k})$. On the other hand, it is obvious that, on the side of node i, we cannot obtain the exact value of the ideal weight matrix W_{ij} due to the lack of original innovations sent from j. Therefore, in (8), we use the estimated $\hat{W}_{ij,k}$ to replace the ideal value of W_{ij}, and an associated updating law for $\hat{W}_{ij,k}$ will be designed later in Subsection III-A for tuning $\hat{W}_{ij,k}$ at each time step according to the state estimation performance adaptively.

Definition 1: An ellipsoid $\mathfrak{X}(a, Y)$ in $\mathbb{R}^n$ is defined by

$$\mathfrak{X}(a, Y) \triangleq \{x \in \mathbb{R}^n : (x-a)^{\mathrm{T}}Y^{-1}(x-a) \leq 1\} \tag{9}$$

where $a \in \mathbb{R}^n$ represents the center and $Y > 0$ is a positive definite matrix with appropriate dimension.

To give the design objective, we define the system state estimation error by $\tilde{x}_{i,k} \triangleq x_k - \hat{x}_{i,k}$ and the NN weights tuning error by $\tilde{W}_{ij,k} \triangleq W_{ij} - \hat{W}_{ij,k}$.

The following assumption concerning the initial conditions on the estimation errors will be needed for our further development.

Assumption 3: The initial condition of $\tilde{W}_{ij,0}$ and $\tilde{x}_{i,0}$ satisfy

$$\text{tr}\left[\tilde{W}_{ij,0}^{\mathrm{T}}Q_{ij,0}^{-1}\tilde{W}_{ij,0}\right] \leq 1$$
$$(x_0 - \hat{x}_{i,0})^{\mathrm{T}}P_0^{-1}(x_0 - \hat{x}_{i,0}) \leq 1$$

where $Q_{ij,0}$ and P_0 are known positive definite real-valued matrices.

It is the aim of this paper to design the estimator gains $F_{i,k}$, $G_{i,k}$ and $H_{ij,k}$ in (8) such that the state estimation errors at each sensing node satisfy the following probabilistic ellipsoidal constraint:

$$\mathbb{P}\{x_k \in \mathfrak{X}(\hat{x}_{i,k}, \mathfrak{P}_k)\} \geq \mathbf{p} \tag{10}$$

or, equivalently,

$$\mathbb{P}\{\tilde{x}_{i,k}^{\mathrm{T}}\mathfrak{P}_k^{-1}\tilde{x}_{i,k} \leq 1\} \geq \mathbf{p} \tag{11}$$

where $\mathfrak{P}_k > 0$ is a pre-specified matrix and $\mathbf{p}$ is a pre-specified scalar satisfying $0 < \mathbf{p} < 1$.

III. MAIN RESULTS

In this section, we will discuss the distributed state estimation issue subject to falsified innovations. First, two useful lemmas are introduced.

Lemma 1: (S-procedure [2]) Define a series of quadratic functions $\kappa_0(\cdot), \kappa_1(\cdot), \ldots, \kappa_\iota(\cdot)$ with respect to the variable $a \in \mathbb{R}^n$ as $\kappa_j(a) \triangleq a^{\mathrm{T}}Y_j a$ where $Y_j^{\mathrm{T}} = Y_j$ ($j = 0, \ldots, \iota$). If

there exist a sequence of non-negative scalars $\{\epsilon_1, \epsilon_2, \ldots, \epsilon_\iota\}$ satisfying $Y_0 - \sum_{j=1}^{\iota} \epsilon_j Y_j \leq 0$, then we have the following derivation:

$$\kappa_1(a) \leq 0, \ldots, \kappa_\iota(a) \leq 0 \to \kappa_0(a) \leq 0.$$

Lemma 2: (Schur Complement Equivalence) For the matrices $\mathcal{Y}_1, \mathcal{Y}_2, \mathcal{Y}_3$ where $\mathcal{Y}_1 = \mathcal{Y}_1^{\mathrm{T}}$ and $0 < \mathcal{Y}_2 = \mathcal{Y}_2^{\mathrm{T}}$, the matrix inequality $\mathcal{Y}_1 + \mathcal{Y}_3^{\mathrm{T}} \mathcal{Y}_2^{-1} \mathcal{Y}_3 < 0$ if and only if

$$\begin{bmatrix} \mathcal{Y}_1 & \mathcal{Y}_3^{\mathrm{T}} \\ \mathcal{Y}_3 & -\mathcal{Y}_2 \end{bmatrix} < 0 \quad \text{or} \quad \begin{bmatrix} -\mathcal{Y}_2 & \mathcal{Y}_3 \\ \mathcal{Y}_3^{\mathrm{T}} & \mathcal{Y}_1 \end{bmatrix} < 0.$$

By using Taylor expansion technique, we describe $f(x_k)$ and $g_i(x_k)$ as follows:

$$f(x_k) = f(\hat{x}_{i,k}) + \Phi_{i,k} \tilde{x}_{i,k} + L_i \Delta_{1i} \tilde{x}_{i,k} \quad (12)$$
$$g_i(x_k) = g_i(\hat{x}_{i,k}) + \Psi_{i,k} \tilde{x}_{i,k} + \Sigma_i \Delta_{2i} \tilde{x}_{i,k} \quad (13)$$

where $L_i \in \mathbb{R}^{n_l}$ and $\Sigma_i \in \mathbb{R}^{n_\Sigma}$ are known matrices; $\Delta_{1i} \in \mathbb{R}^{n_l \times n_x}$ and $\Delta_{2i} \in \mathbb{R}^{n_\Sigma \times n_x}$ are unknown matrices satisfying $\|\Delta_{1i}\| \leq 1$ and $\|\Delta_{2i}\| \leq 1$; $\Phi_{i,k}$ and $\Psi_{i,k}$ are obtained as follows:

$$\Phi_{i,k} \triangleq \left. \frac{\partial f(x)}{\partial x} \right|_{x=\hat{x}_{i,k}}, \quad \Psi_{i,k} \triangleq \left. \frac{\partial g_i(x)}{\partial x} \right|_{x=\hat{x}_{i,k}}. \quad (14)$$

Remark 3: In (12) and (13), the terms $L_i \Delta_{1i} \tilde{x}_{i,k}$ and $\Sigma_i \Delta_{2i} \tilde{x}_{i,k}$ account for the truncation errors due to the neglected higher order terms in the Taylor series expansion of the corresponding nonlinear functions. Alternatively, Δ_{1i} and Δ_{2i} can also be illustrated as uncertain terms that are employed to describe the modeling errors in the matrices $\Phi_{i,k}$ and $\Psi_{i,k}$, respectively. The readers are referred to [3] for a rigorous justification of the validity of such a representation for the linearization error.

Then, the one-step estimation error dynamics are acquired as follows:

$$\begin{aligned} \tilde{x}_{i,k+1} =& x_{k+1} - \hat{x}_{i,k+1} \\ =& f(x_k) + B_k \nu_k \\ & - \Big(F_{i,k} \hat{x}_{i,k} + G_{i,k}(y_{i,k} - \hat{y}_{i,k}) \\ & + \sum_{j \in \mathcal{N}_i} \theta_{ij} H_{ij,k} \big(\tilde{z}_{j,k} - \bar{\alpha} \hat{W}_{ij,k} \phi(z_{i,k}) \big) \Big) \\ =& f(\hat{x}_{i,k}) + \Phi_{i,k} \tilde{x}_{i,k} + L_i \Delta_{1i} \tilde{x}_{i,k} + B_k \nu_k \\ & - \Big(F_{i,k} \hat{x}_{i,k} + G_{i,k} \big(\Psi_{i,k} \tilde{x}_{i,k} + \Sigma_i \Delta_{2i} \tilde{x}_{i,k} + E_{i,k} \mu_k \big) \\ & + \sum_{j \in \mathcal{N}_i} \theta_{ij} H_{ij,k} \big(\Psi_{j,k} \tilde{x}_{j,k} + \Sigma_j \Delta_{2j} \tilde{x}_{j,k} + E_{j,k} \mu_k \\ & + \alpha_k W_{ij} \phi(z_{j,k}) + \alpha_k \delta_{j,k} - \bar{\alpha} \hat{W}_{ij,k} \phi(z_{i,k}) \big) \Big). \end{aligned} \quad (15)$$

For brevity of development, we define

$$\xi_k \triangleq \mathrm{col}_N\{x_k\}, \ \hat{x}_k \triangleq \mathrm{col}_N\{\hat{x}_{i,k}\}, \ \tilde{x}_k \triangleq \mathrm{col}_N\{\tilde{x}_{i,k}\},$$
$$\hat{f}_k \triangleq \mathrm{col}_N\{f(\hat{x}_{i,k})\}, \hat{g}_k \triangleq \mathrm{col}_N\{g(\hat{x}_{i,k})\}, \ \delta_k \triangleq \mathrm{col}_N\{\delta_{i,k}\},$$
$$\phi_k \triangleq \mathrm{col}_N\{\phi(z_{i,k})\}, \ \eta_k \triangleq \mathrm{col}_N\{y_{i,k}\}, \ \mathcal{B}_k \triangleq \mathbf{1}_N \otimes B_k,$$
$$\mathcal{G}_k \triangleq \mathrm{diag}_N\{G_{i,k}\}, \ \mathcal{F}_k \triangleq \mathrm{diag}_N\{F_{i,k}\}, \ \mathcal{E}_k \triangleq \mathrm{diag}_N\{E_{i,k}\},$$
$$\mathcal{W} \triangleq [W_{ij}]_{N \times N}, \ \hat{\mathcal{W}}_k \triangleq [\hat{W}_{ij,k}]_{N \times N},$$

$$\mathcal{L} \triangleq \mathrm{diag}_N\{L_i\}, \ \Phi_k \triangleq \mathrm{diag}_N\{\Phi_{i,k}\},$$
$$\Psi_k \triangleq \mathrm{diag}_N\{\Psi_{i,k}\}, \ \Sigma \triangleq \mathrm{diag}_N\{\Sigma_i\},$$
$$\Delta_1 \triangleq \mathrm{diag}_N\{\Delta_{1i}\}, \ \Delta_2 \triangleq \mathrm{diag}_N\{\Delta_{2i}\},$$
$$\Theta_{\iota,i} \triangleq \mathrm{diag}\{\underbrace{0, \ldots, 0}_{i-1}, I_\iota, \underbrace{0, \ldots, 0}_{N-i}\},$$
$$\mathcal{R}_{\iota,i} \triangleq (\mathbf{1}_N^{\mathrm{T}} \otimes I_\iota)\Theta_{\iota,i}, \quad \iota = \{n_x, n_s, n_y, n_\Sigma\}.$$

Then, we rewrite the estimation error dynamics in the following form:

$$\begin{aligned} \tilde{x}_{k+1} =& \hat{f}_k + \Phi_k \tilde{x}_k + \mathcal{L} \Delta_1 \tilde{x}_k + \mathcal{B}_k \nu_k \\ & - \mathcal{F}_k \hat{x}_k - \mathcal{G}_k \Psi_k \tilde{x}_k - \mathcal{G}_k \Sigma \Delta_2 \tilde{x}_k \\ & - \mathcal{G}_k \mathcal{E}_k (\mathbf{1}_N \otimes I_{n_\mu}) \mu_k - \mathcal{H}_k \Psi_k \tilde{x}_k - \mathcal{H}_k \Sigma \Delta_2 \tilde{x}_k \\ & - \mathcal{H}_k \mathcal{E}_k (\mathbf{1}_N \otimes I_{n_\mu}) \mu_k - \alpha_k (\mathcal{H}_k \circ \mathcal{W}) \phi_k \\ & + \bar{\alpha} \sum_{i=1}^N \Theta_{n_x,i} (\mathcal{H}_k \circ \hat{\mathcal{W}}_k)(\mathbf{1}_N \otimes \mathcal{R}_{n_y,i}) \phi_k \\ =& \hat{f}_k - \mathcal{F}_k \hat{x}_k + \big(\Phi_k + \mathcal{L} \Delta_1 - \mathcal{G}_k \Psi_k \\ & - \mathcal{G}_k \Sigma \Delta_2 - \mathcal{H}_k \Psi_k - \mathcal{H}_k \Sigma \Delta_2 \big) \tilde{x}_k \\ & + \mathcal{B}_k \nu_k - (\mathcal{G}_k + \mathcal{H}_k) \mathcal{E}_k (\mathbf{1}_N \otimes I_{n_\mu}) \mu_k \\ & - \alpha_k \mathcal{H}_k \delta_k - \alpha_k (\mathcal{H}_k \circ \mathcal{W}) \phi_k \\ & + \bar{\alpha} \sum_{i=1}^N \Theta_{n_x,i} (\mathcal{H}_k \circ \hat{\mathcal{W}}_k)(\mathbf{1}_N \otimes \mathcal{R}_{n_y,i}) \phi_k \quad (16) \end{aligned}$$

where $\mathcal{H}_k \triangleq \left[\theta_{ij} H_{ij,k} \right]_{N \times N}$. Note that since $\theta_{ij} = 0$ when $j \notin \mathcal{N}_i$, $\mathcal{H}_k$ is a sparse matrix described by

$$\mathcal{H}_k \in \mathscr{T}_{n_x \times n_y}$$

where $\mathscr{T}_{n_x \times n_y} \triangleq \big\{ \mathcal{T} = [T_{ij}] \in \mathbb{R}^{Nn_x \times Nn_y} | T_{ij} \in \mathbb{R}^{n_x \times n_y}, T_{ij} = 0 \text{ if } j \notin \mathcal{N}_i \big\}$. Moreover, in equation (16), for two block matrices $A = [A_{ij}]_{N \times N}$ and $B = [B_{ij}]_{N \times N}$ where A_{ij} and B_{ij} are matrices with compatible dimensions, the product of $A \circ B$ is defined by $A \circ B \triangleq [A_{ij} B_{ij}]_{N \times N}$.

A. Tuning of $\hat{W}_{ij,k}$

The adaptive tuning law of the estimated weight matrix $\hat{W}_{ij,k}$ is designed as follows:

$$\hat{W}_{ij,k+1} = \lambda_{ij,k}^{(1)} \hat{W}_{ij,k} + \lambda_{ij,k}^{(2)} \varpi_k \phi^{\mathrm{T}}(z_{i,k}) \quad (17)$$

where

$$\varpi_k \triangleq \bar{\alpha} \sum_{j \in \mathcal{N}_i} \theta_{ij} H_{ij,k}^{\mathrm{T}} \Psi_{i,k+1}^{\mathrm{T}} z_{i,k+1}$$

and $\lambda_{ij,k}^{(1)}$ and $\lambda_{ij,k}^{(2)}$ are positive tuning scalars that will be determined later.

Remark 4: The adaptive tuning law (17) is proposed according to the gradient descent algorithm aiming to minimize the cost function defined by $J \triangleq \|z_{i,k+1}\| = \|y_{i,k+1} - \hat{y}_{i,k+1}\|$. The tuning scalars $\lambda_{ij,k}^{(1)}$ and $\lambda_{ij,k}^{(2)}$ will be co-designed with the parameters of state estimators subsequently.

The following theorem gives a sufficient condition under which the matrix $\tilde{W}_{ij,k}$ will be bounded in the sense of weighted Frobenius norm.

Theorem 1: For nonlinear system (1), the tuning law of $\hat{W}_{ij,k}$ is given by (17). Let the family of matrices $\{Q_{ij,k}\}_{k\in[0,T]}$ be given. If there exist families of positive scalars $\{\varrho_{ij,k}^{(1)}, \varrho_{ij,k}^{(2)}\}_{k\in[0,T-1]}$ and families of tuning scalars $\{\lambda_{ij,k}^{(1)}, \lambda_{ij,k}^{(2)}\}_{k\in[0,T-1]}$ satisfying:

$$\begin{bmatrix} \Omega_{ij,k} & * \\ \Upsilon_{ij,k} & -\mathcal{Q}_{ij,k+1} \end{bmatrix} \leq 0 \tag{18}$$

where

$$\mathcal{Q}_{ij,k} \triangleq Q_{ij,k} \otimes I_{n_y},$$
$$\Omega_{ij,k} \triangleq \mathrm{diag}\{-1 + \varrho_{ij,k}^{(1)} + \epsilon_{1j}^2 \varrho_{ij,k}^{(1)}, -\varrho_{ij,k}^{(1)} I, -\varrho_{ij,k}^{(2)} I\},$$
$$\Upsilon_{ij,k} \triangleq \begin{bmatrix} -\lambda_{ij,k}^{(2)} \varpi_k \otimes \phi(z_{i,k}) & \lambda_{ij,k}^{(1)} M_{ij,k} & (1 - \lambda_{ij,k}^{(1)}) I \end{bmatrix}$$

with $M_{ij,k}$ being a factorization of the matrix $\mathcal{Q}_{ij,k}$ (i.e., $\mathcal{Q}_{ij,k} = M_{ij,k} M_{ij,k}^{\mathrm{T}}$), then the following inequality holds:

$$\mathrm{tr}[\tilde{W}_{ij,k+1}^{\mathrm{T}} \mathcal{Q}_{ij,k+1}^{-1} \tilde{W}_{ij,k+1}] \leq 1. \tag{19}$$

Proof: See Appendix VI-A. ∎

Remark 5: Theorem 1 gives a sufficient condition to constrain the estimation errors of NN weight matrix (i.e., $\tilde{W}_{ij,k}$) to be bounded by a pre-specified criterion (i.e., $Q_{ij,k}$) in the sense of weighted Frobenius norm. From (17) we know that the matrix $\hat{W}_{ij,k+1}$ will be tuned adaptively by making the trade-off between the current step weight matrix (i.e., $\hat{W}_{ij,k}$) and the innovation information (i.e., $z_{i,k+1}$). This theorem will play a paramount role in the following analysis and design of the desired state estimator gains. The connection between the tuning errors of NN weight and the estimation errors of system state will be discussed in detail in Remark 6 of Subsection III-C.

B. Ellipsoidal constraint in probability $\mathbf{p}$

In this subsection, we shall establish the sufficient condition under which the state estimation error will be driven to reside within an allowable ellipsoidal area in a pre-specified probability. First, we introduce the following lemmas that will be used in the derivation.

Lemma 3: [36] Given a random variable υ of suitable dimension which belongs to an ellipsoid $\mathfrak{X}(a, Y)$ as follows:

$$\upsilon \in \mathfrak{X}(a, Y) \triangleq \{\upsilon | (\upsilon - a)^{\mathrm{T}} Y^{-1} (\upsilon - a) \leq 1\}$$

where a and Y are given in Definition 1. If, for any given $0 < \mathbf{p} < 1$, the following inequality

$$\mathbb{E}\{(\upsilon - a)^{\mathrm{T}} Y^{-1} (\upsilon - a)\} \leq 1 - \mathbf{p}$$

is true, then we have

$$\mathbb{P}\{\upsilon \in \mathfrak{X}(a, Y)\} \geq \mathbf{p}.$$

Define a matrix P_k by

$$P_k \triangleq \frac{1}{1 - \mathbf{p}} \mathfrak{P}_k,$$

and then, on basis of Lemma 3, we immediately obtain the following lemma.

Lemma 4: If $\mathbb{E}\{(x_k - \hat{x}_{i,k})^{\mathrm{T}} P_k^{-1} (x_k - \hat{x}_{i,k})\} \leq 1$, then the following holds:

$$\mathbb{P}\{x_k \in \mathfrak{X}(\hat{x}_{i,k}, \mathfrak{P}_k)\} \geq \mathbf{p}. \tag{20}$$

Proof: Lemma 4 is easily proved according to Lemma 3, and the proof is therefore omitted here. ∎

Defining $\tilde{\alpha}_k \triangleq \alpha_k - \bar{\alpha}$, we know that

$$\mathbb{E}\{\tilde{\alpha}_k\} = 0, \quad \mathbb{E}\{\tilde{\alpha}_k^2\} = \bar{\alpha}(1 - \bar{\alpha}) \triangleq \sigma^2. \tag{21}$$

Theorem 2: Let the estimator gains $F_{i,k}$, $G_{i,k}$ and $H_{ij,k}$ be given. Under the condition of Theorem 1, for a pre-specified family of positive definite matrices $\{P_k\}_{k\in[0,T]}$ with a factorization of $P_k = S_k S_k^{\mathrm{T}}$, if there exist families of non-negative scalars $\{\beta_{i,k}^{(1)}, \beta_k^{(2)}, \beta_k^{(3)}, \beta_k^{(4)}, \beta_{i,k}^{(5)}, \beta_{i,k}^{(6)}, \beta_{i,k}^{(7)}\}_{k\in[0,T-1]}$ satisfying the following recursive matrix inequalities:

$$\begin{bmatrix} -\Gamma_k & * & * \\ \mathcal{R}_{n_x,i} \bar{\Xi}_k & -P_{k+1} & * \\ \sigma \mathcal{R}_{n_x,i} \tilde{\Xi}_k & 0 & -P_{k+1} \end{bmatrix} \leq 0 \tag{22}$$

where

$$\Gamma_k \triangleq \mathrm{diag}\Big\{ 1 - \sum_{i=1}^{N} (\beta_{i,k}^{(1)} + \beta_{i,k}^{(5)} \epsilon_{3i}) - \beta_k^{(2)} - \beta_k^{(3)} - N^2 \beta_k^{(4)},$$
$$\sum_{i=1}^{N} \beta_{i,k}^{(1)} \Theta_{n_s,i} - \sum_{i=1}^{N} (\beta_{i,k}^{(6)} + \beta_{i,k}^{(7)}) \Theta_{n_s,i} \mathcal{S}_k^{\mathrm{T}} \mathcal{S}_k,$$
$$\sum_{i=1}^{N} \beta_{i,k}^{(6)} \Theta_{n_l,i}, \sum_{i=1}^{N} \beta_{i,k}^{(7)} \Theta_{n_\Sigma,i}, \beta_k^{(2)} V_k^{-1}, \beta_k^{(3)} U_k^{-1},$$
$$\sum_{i=1}^{N} \beta_{i,k}^{(5)} \Theta_{n_\delta,i}, \beta_k^{(4)} I_{N^2} \Big\}, \tag{23}$$

$$\bar{\Xi}_k \triangleq \begin{bmatrix} \bar{\Xi}_k^{(11)} & \Xi_k^{(12)} & \mathcal{L} & -(\mathcal{G}_k + \mathcal{H}_k)\Sigma \\ & \mathcal{B}_k & \Xi_k^{(16)} & -\bar{\alpha}\mathcal{H}_k & -\bar{\alpha}\mathscr{H}_k \end{bmatrix}, \tag{24}$$

$$\bar{\Xi}_k^{(11)} \triangleq \hat{f}_k - \mathcal{F}_k \hat{x}_k - \bar{\alpha}(\mathcal{H}_k \circ \hat{\mathcal{W}}_k)\phi_k$$
$$+ \bar{\alpha} \sum_{i=1}^{N} \Theta_{n_x,i}(\mathcal{H}_k \circ \hat{\mathcal{W}}_k)(\mathbf{1}_N \otimes \mathcal{R}_{n_y,i})\phi_k, \tag{25}$$

$$\tilde{\Xi}_k \triangleq \begin{bmatrix} -(\mathcal{H}_k \circ \hat{\mathcal{W}}_k)\phi_k & 0 & 0 & 0 \\ & 0 & 0 & -\mathcal{H}_k & -\mathscr{H}_k \end{bmatrix}, \tag{26}$$

$$\Xi_k^{(12)} \triangleq (\Phi_k - \mathcal{G}_k \Psi_k - \mathcal{H}_k \Psi_k) \mathcal{S}_k, \tag{27}$$

$$\Xi_k^{(16)} \triangleq -(\mathcal{G}_k + \mathcal{H}_k) \mathcal{E}_k (\mathbf{1}_N \otimes I_{n_\mu}), \tag{28}$$

then the following inequality holds:

$$\mathbb{E}\{(x_{k+1} - \hat{x}_{i,k+1})^{\mathrm{T}} P_{k+1}^{-1} (x_{k+1} - \hat{x}_{i,k+1})\} \leq 1. \tag{29}$$

Proof: See Appendix VI-B. ∎

C. Distributed Estimator Design

In this subsection, a sufficient condition is provided for the solvability of the addressed probability-guaranteed distributed state estimator design problem. The desired estimator gains at each time step can be acquired by solving the corresponding matrix inequalities.

Theorem 3: Given a pre-specified $0 < \mathbf{p} < 1$ and a family of positive definite matrices $\{\mathfrak{P}_k\}_{k\in[0,T]}$. Under the condition given in Theorem 1, the design objective (11) is satisfied if there exist families of non-negative scalars $\{\beta_{i,k}^{(1)}, \beta_k^{(2)}, \beta_k^{(3)}, \beta_{i,k}^{(4)}, \beta_{i,k}^{(5)}, \beta_{i,k}^{(6)}, \beta_{i,k}^{(7)}\}_{k\in[0,T-1]}$ such that

$$
\begin{bmatrix}
-\Gamma_k & * & * \\
\mathcal{R}_{n_x,i}\bar{\bar{\Xi}}_k & -\frac{1}{1-\mathbf{p}}\mathfrak{P}_{k+1} & * \\
\sigma\mathcal{R}_{n_x,i}\tilde{\bar{\Xi}}_k & 0 & -\frac{1}{1-\mathbf{p}}\mathfrak{P}_{k+1}
\end{bmatrix} \le 0. \quad (30)
$$

Moreover, the estimator gains $F_{i,k}$, $G_{i,k}$ and $H_{ij,k}$ can be calculated via solving the corresponding matrix inequalities.

Proof: Theorem 3 can be proved easily from Lemma 4 and Theorem 2 by taking $P_{k+1} = \frac{1}{1-\mathbf{p}}\mathfrak{P}_{k+1}$ into account, and thus is omitted here. ∎

Remark 6: Theorem 3 provides a scheme to obtain the desired estimator guaranteeing the state estimation errors are confined within an allowable ellipsoidal area in probability $\mathbf{p}$. Note that Theorem 3 has an inherent connection with Theorem 1. On one hand, the boundedness of $\tilde{W}_{ij,k}$, ensured by Theorem 1, is of help to restrict the state estimation error $\tilde{x}_{i,k}$; on the other hand, the satisfaction of the requirement imposed on the state estimation errors, guaranteed by Theorem 3, contributes to constrain $\tilde{W}_{ij,k}$. In specific, Theorem 1 influences the state estimation via (37) and (53), while Theorem 3 helps to regulate the tuning of $\hat{W}_{ij,k}$ via state estimate $\hat{x}_{i,k}$ and the resulting innovation characterized by formula (17). In such a coupling yet collaborative way, the desired state estimation performance can be reached.

Remark 7: Notice that the RLMI algorithm proposed in this paper is based on LMI approach. As discussed in [2], the computational complexity of an LMI system is bounded by $O(\mathscr{P}\mathscr{Q}3\log(\mathscr{U}/\varepsilon))$ where $\mathscr{P}$ represents the row size, $\mathscr{Q}$ stands for the number of scalar decision variables, $\mathcal{U}$ is a data-dependent scaling factor and ε is relative accuracy set for algorithm. For instance, let us now look at the condition proposed in Theorem 1, where the number of sensing nodes is N, the iteration time is $T+1$ (since the time interval is $[0,T]$) and the dimensions of variables are known from $x_k \in \mathbb{R}^{n_x}$, $y_{i,k} \in \mathbb{R}^{n_y}$, $\nu_k \in \mathbb{R}^{n_\nu}$ and $\mu_k \in \mathbb{R}^{n_\mu}$. Moreover, we assume that $r_{i,k} \in \mathbb{R}^{n_r}$ and $\chi(z_{j,k}) \in \mathbb{R}^{n_\chi}$. The RLMI-based algorithm is implemented recursively for $T+1$ steps and, at each step, we need to solve the LMI (18) with $\mathcal{P} = N^2(1 + n_r + n_\chi \dot{n}_y)$ rows and $\mathcal{Q} = N^2(\frac{n_\chi^2 + n_\chi}{2} + 4)$ scalar variables. Accordingly, the computational complexity of the proposed RLMI algorithm can be represented by $O((T+1)\mathcal{P}\mathcal{Q})$. We can now conclude that the computational complexity of our proposed RLMI algorithm depends linearly on the length of time interval $T+1$. It is worth mentioning that the study on LMI optimization is very active in recent years within the communities of applied mathematics, control science and signal processing. We can expect substantial speedups in the near future.

The following algorithm provides a computing method to iteratively obtain the required estimator gains $\{F_{i,k}, G_{i,k}, H_{ij,k}\}$.

Algorithm 1 Distributed estimator design algorithm

Input: Topology information $\mathscr{G} = (\mathscr{V}, \mathscr{E}, \mathscr{L})$; Initial values: $x_{i,0}, \hat{x}_{i,0}, W_{ij,0}, \hat{W}_{ij,0}, Q_{ij,0}$, and P_0; Performance constraints $\{Q_{ij,k}\}_{k>0}$, $\{\mathfrak{P}_k\}_{k>0}$ and $\mathbf{p}$; Activation function $\phi(\cdot)$; Other coefficients: $\bar{\alpha}, \epsilon_{1i}, \epsilon_2, \epsilon_{3i}, U_k$ and V_k; Maximum time step T.

1: Set $k = 0$
2: **if** $k < T$ **then**
3:　Solve (30) for estimator gains $\{F_{i,k}, G_{i,k}, H_{ij,k}\}$.
4:　Obtain one-step estimate $\hat{x}_{i,k+1}$ according to (8).
5:　Obtain innovation $z_{i,k+1}$.
6:　Solve (18) for tuning coefficients $\lambda_{ij,k}^{(1)}$ and $\lambda_{ij,k}^{(2)}$.
7:　Obtain $\hat{W}_{ij,k+1}$ according to (17).
8: **end if**

D. Optimization Problems

It should be noticed that within the design framework we proposed in Subsection III-C, the acquired estimator parameters could be a set if existing. Therefore, in this subsection, we aim to present the optimization problems to seek the locally optimal estimator parameters by making use of trade-off between performance specifications.

OP1: Minimization of $Q_{ij,k}$ (in the sense of matrix trace) to ensure the minimal estimation errors of the NN weight matrices (in the sense of weighted Frobenius norm)

Corollary 1: Under the conditions given in Theorem 1, a family of minimized matrices $\{Q_{ij,k}\}_{k\in[0,T]}$ can be obtained in the sense of matrix trace if the following minimization problem is feasible:

$$
\min_{\{\varrho_{ij,k}^{(1)}, \varrho_{ij,k}^{(2)}, \lambda_{ij,k}^{(1)}, \lambda_{ij,k}^{(2)}, Q_{ij,k+1}\}} \text{trace}[Q_{ij,k+1}] \quad (31)
$$

$$\text{subject to } (18)$$

In the following, on basis of Corollary 1, we shall further present two optimization problems to guarantee locally optimal state estimation performance from different perspectives. To this end, for brevity of development, denote a set of variables by

$$
\mathfrak{S}_k \triangleq \{\mathcal{F}_k, \mathcal{G}_k, \mathcal{H}_k, \beta_{i,k}^{(1)}, \beta_k^{(2)}, \beta_k^{(3)}, \beta_k^{(4)}, \beta_{i,k}^{(5)}, \beta_{i,k}^{(6)}, \beta_{i,k}^{(7)}\}.
$$

OP2: Minimization of $\mathfrak{P}_k$ (in the sense of matrix trace) to guarantee the locally optimal estimation subject to the fixed pre-specified probability specification.

Corollary 2: Let $\mathbf{p}$ be given. Under the conditions given in Theorem 2 and Corollary 1, a family of minimized matrices $\{\mathfrak{P}_k\}_{k\in[0,T]}$ can be ensured in the sense of matrix trace if the following minimization problem is feasible:

$$
\min_{\{\mathfrak{S}_k, \mathfrak{P}_{k+1}\}} \text{trace}[\mathfrak{P}_{k+1}] \quad (32)
$$

$$\text{subject to } (22)$$

Next, suppose $\mathbf{p}$ is time-varying and denote by $\mathbf{p}_k$ the probability criterion at instant k. Define

$$
\mathbf{q}_k \triangleq \frac{1}{1 - \mathbf{p}_k}.
$$

OP3: Minimization of $\mathbf{q}_k$ subject to fixed $\mathfrak{P}_k$ to determine the lower bound on the probability criterion at each time step.

Corollary 3: Let $\{\mathfrak{P}_k\}_{k\in[0,T]}$ be given. Under the conditions in Theorem 3 and Corollary 1, the lower bound on $\mathbf{p}_k$ is guaranteed if the following minimization problem is feasible:

$$\min_{\{\mathfrak{S}_k,\mathbf{q}_k\}} \quad \mathbf{q}_k \tag{33}$$

subject to $\begin{cases} 1 < \mathbf{q}_k < +\infty \\ \begin{bmatrix} -\Gamma_k & * & * \\ \mathcal{R}_{n_x,i}\bar{\Xi}_k & -\mathbf{q}_k\mathfrak{P}_{k+1} & * \\ \sigma\mathcal{R}_{n_x,i}\bar{\Xi}_k & 0 & -\mathbf{q}_k\mathfrak{P}_{k+1} \end{bmatrix} \leq 0 \end{cases}$

The proofs of Corollaries 1–3 are quite simple on basis of the obtained theorems, and are thus omitted here.

Remark 8: So far, we have discussed the probability-guaranteed distributed state estimation problem for nonlinear system over sensor networks under random falsified innovations. The solvability of the addressed problem is cast into the feasibility of a set of recursive linear matrix inequalities. A methodology has been proposed which utilizes neural networks to approximate the unknown falsified innovations. Within the established framework, the tuning parameters of NN weight matrices and the probabilistic ellipsoidal constraint can be investigated systematically. The desired estimator gains ensuring required performance criterion can be obtained by solving the recursive linear matrix inequalities at each time step. Three optimization problems are proposed to ensure the sub-optimal estimation performances from different perspectives. It should be emphasized that, it is not difficult to extend our obtained results to the case where more performance indices are considered simultaneously within the proposed theoretical framework.

Remark 9: Our proposed algorithm possesses the following merits: i) the recursive nature of the algorithm is a true reflection of the time-varying feature of the addressed estimation error dynamics; ii) the idea of using NN to approximate the unknown falsified innovations is of more practical significance in comparison to those existing methods where the attack functions are usually assumed to be known to the defenders; and iii) within the established generic framework, it is not difficult to apply our developed methodology to deal with systems with more complex dynamics such as quantization effects and Markovian jump parameters, or handle more performance indices such as H_∞ specification and robustness.

IV. A Numerical Example

In this section, an illustrative example is presented to demonstrate the usefulness of the proposed algorithm in this paper. First, for a vector $c \in \mathbb{R}^2$, denote by $c^{(1)}$ and $c^{(2)}$ the first and second entries of c, respectively.

Consider a nonlinear system whose parameters are given as follows:

$$B_k = \begin{bmatrix} 0.1 + 0.01\cos(k+1) \\ 0.2 + 0.01\sin(k) \end{bmatrix},$$

$$E_{1,k} = \begin{bmatrix} 0.25 \\ 0.2 + 0.1\cos(k+1) \end{bmatrix},$$

$$E_{2,k} = \begin{bmatrix} 0.2 \\ 0.15 + 0.1\sin(2k) \end{bmatrix},$$

$$E_{3,k} = \begin{bmatrix} 0.25 + 0.05\sin(k) \\ 0.15 \end{bmatrix},$$

$$f(x_k) = \begin{bmatrix} 0.5x_k^{(1)} + 0.45\sin(x_k^{(1)}) + 0.25x_k^{(2)} \\ 0.4\cos(x_k^{(2)}) + 0.3x_k^{(2)} + 0.6x_k^{(1)} \end{bmatrix},$$

$$g_1(x_k) = \begin{bmatrix} 0.15x_k^{(1)} + 0.35\sin(2x_k^{(1)}) + 0.2x_k^{(2)} \\ 0.5\sin(x_k^{(2)}) + 0.1x_k^{(2)} + 0.3x_k^{(1)} \end{bmatrix},$$

$$g_2(x_k) = \begin{bmatrix} 0.3x_k^{(1)} + 0.3\sin(x_k^{(1)}+1) + 0.4x_k^{(2)} \\ 0.45\cos(x_k^{(2)}) + 0.25x_k^{(2)} + 0.5x_k^{(1)} \end{bmatrix},$$

$$g_3(x_k) = \begin{bmatrix} 0.5\sin(x_k^{(1)}) + 0.4x_k^{(1)} + 0.2x_k^{(2)} \\ 0.4\sin(x_k^{(2)}) + 0.1x_k^{(2)} + 0.3x_k^{(1)} \end{bmatrix}.$$

We select $\nu_k = 0.3\cos(2k)$ and $\mu(k) = 0.4\sin(k)$. Set $V = 1$ and $U = 1$. It can be easily verified that Assumption 1 is satisfied.

Set $\epsilon_{1i} = \epsilon_2 = \epsilon_{3i} = 3$. Choose the activation function of the neural networks as follows:

$$\phi(z_{i,k}) = \begin{bmatrix} \tanh(z_{i,k}^{(1)}) \\ \tanh(z_{i,k}^{(2)}) \end{bmatrix}.$$

Assume that the number of sensor nodes is three, and the adjacency matrix $\mathscr{L}$ is set as follows:

$$\mathscr{L} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

Set $\mathbf{p} = 0.8$. The initial values in this simulation are selected as follows:

$$\hat{x}_{1,0} = \begin{bmatrix} 2.0 \\ 1.4 \end{bmatrix}, \quad \hat{x}_{2,0} = \begin{bmatrix} 1.5 \\ 1.8 \end{bmatrix},$$

$$\hat{x}_{3,0} = \begin{bmatrix} 1.7 \\ 1.9 \end{bmatrix}, \quad x_0 = \begin{bmatrix} 1.8 \\ 1.5 \end{bmatrix},$$

$$P_0 = \begin{bmatrix} 0.6887 & -0.0588 \\ -0.0588 & 0.7390 \end{bmatrix},$$

$$\hat{W}_{ij,0} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad Q_{ij,0} = \begin{bmatrix} 20 & 0 \\ 0 & 20 \end{bmatrix}.$$

The random falsified innovations are governed by the following parameters:

$$\bar{\alpha} = 0.6, \quad \chi(z_{i,k}) = \begin{bmatrix} 0.15\sin(z_{i,k}^{(1)}) \\ 0.30\cos(z_{i,k}^{(2)}) \end{bmatrix}.$$

By solving the optimization problems in Corollaries 1–3, the simulation results are obtained in Figs. 1–4. Specifically, Figs. 1–2 depict the trajectories of the system state x_k (i.e., $x_k^{(1)}$ and $x_k^{(2)}$) and the corresponding estimates. Figs. 3–4

present the trajectories of the estimation errors $\tilde{x}_{i,k}^{(1)}$ and $\tilde{x}_{i,k}^{(2)}$, respectively. It can be seen from Fig. 1 and Fig. 2 that the developed technique can estimate the entries of state (i.e., x_k and x_k) with a satisfactory precision, which is also demonstrated via Fig. 3 and Fig. 4 where both the estimation errors are confined within the required range. From the simulation figures we can clearly observe that the proposed algorithm can effectively provide state estimates on each sensor node.

Next, we proceed to show the superiority of the proposed algorithm on mitigating the influence from falsified innovations over the conventional distributed estimator characterized by (3). A comparative simulation is carried out where we apply the conventional algorithm (3) to generate local state estimates under the identical manipulations of innovations. The simulation results by using the conventional algorithm are recorded in Figs. 5–8. In specific, Figs. 5–6 plot the trajectories of the system state x_k (i.e., $x_k^{(1)}$ and $x_k^{(2)}$) and their estimates, respectively. Figs. 7–8 depict the evolutions of the estimation errors $\tilde{x}_{i,k}^{(1)}$ and $\tilde{x}_{i,k}^{(2)}$, respectively. From the figures we can see that the conventional algorithm is also capable of providing state estimation with certain precision. However, in comparison to our developed approach, the conventional algorithm shows worse estimation performance in the existence of falsified innovations.
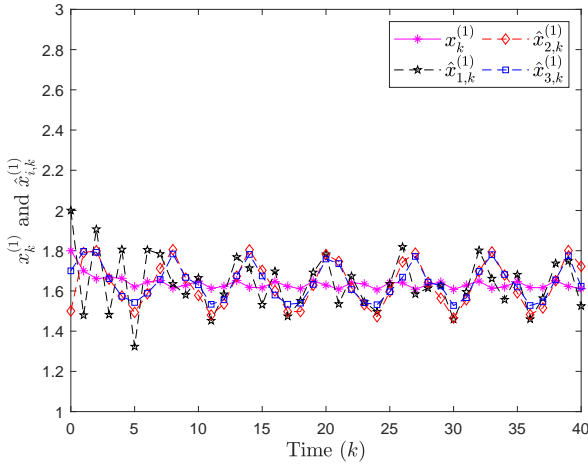

Fig. 1.   Trajectories of $x_k^{(1)}$ and $\hat{x}_{i,k}^{(1)}$ (by estimator with NN).

V. CONCLUSION

In this paper, we have investigated the distributed state estimation problem for a class of general nonlinear systems subject to falsified innovations. A neural-network-based mechanism has been proposed to approximate the unknown manipulated data, thereby mitigating the effects on the state estimation. With the help of certain convex optimization methods, sufficient conditions have been established for the solvability of the addressed estimator design issue, ensuring that the state estimation errors at each sensing node reside within required ellipsoidal regions in a prescribed probability. The desired estimator gains have been given in terms of the solution to a set of recursive matrix inequalities. On basis of
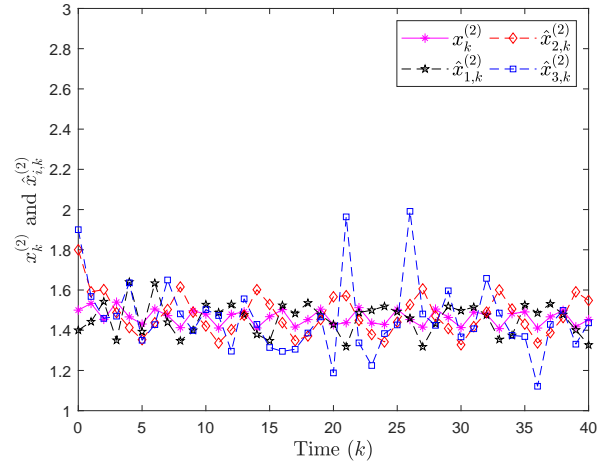

Fig. 2.   Trajectories of $x_k^{(2)}$ and $\hat{x}_{i,k}^{(2)}$ (by estimator with NN).
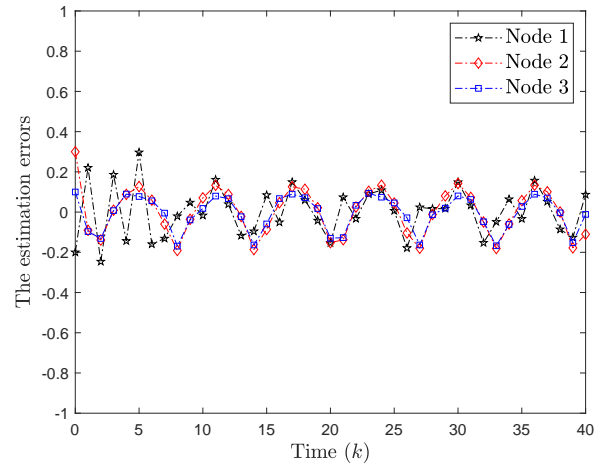

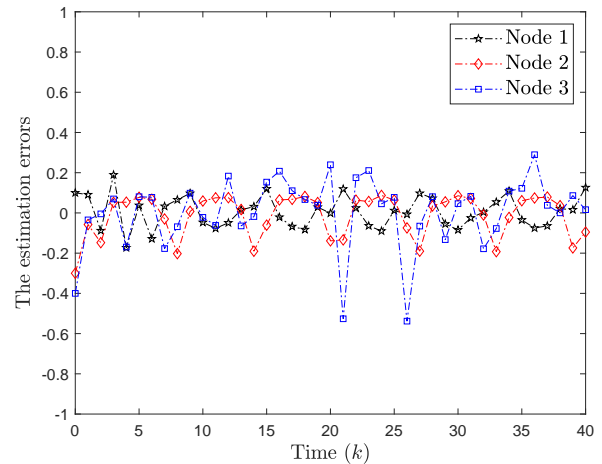Fig. 3. $\tilde{x}_{i,k}^{(1)}$ (by estimator with NN).

Fig. 4. $\tilde{x}_{i,k}^{(2)}$ (by estimator with NN).

the established framework, optimization problems have been considered with the aim to ensure locally optimal estimation
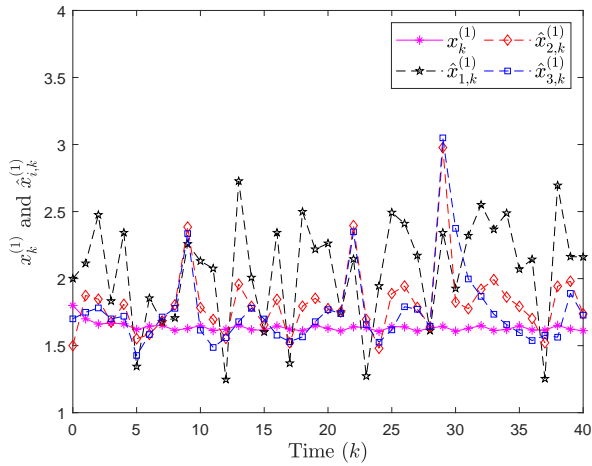
Fig. 5. Trajectories of $x_k^{(1)}$ and $\hat{x}_{i,k}^{(1)}$ (by conventional estimator).
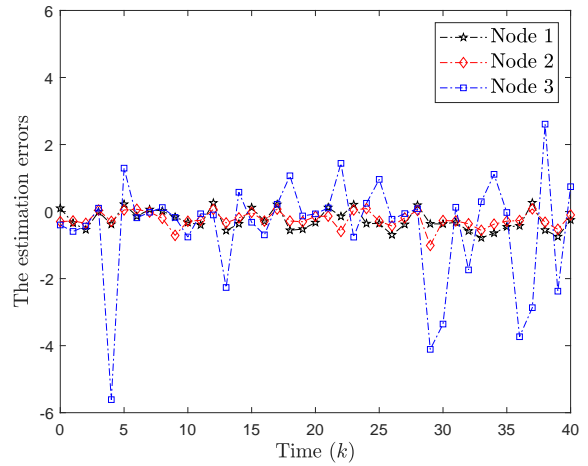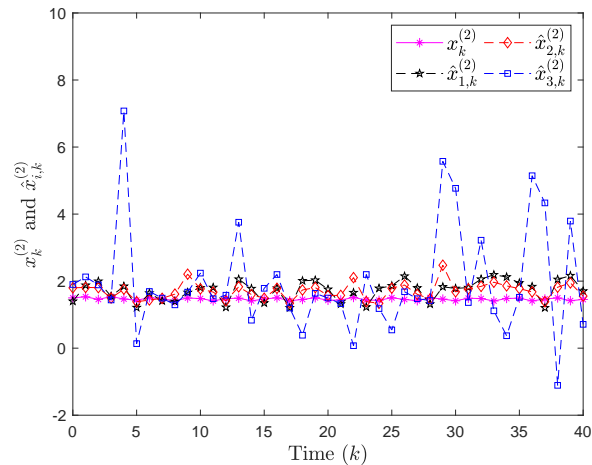
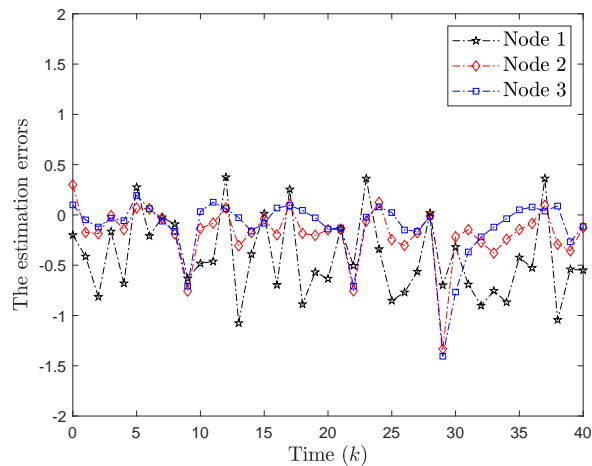Fig. 6. Trajectories of $x_k^{(2)}$ and $\hat{x}_{i,k}^{(2)}$ (by conventional estimator).

Fig. 7. $\tilde{x}_{i,k}^{(1)}$ (by conventional estimator).

Fig. 8. $\tilde{x}_{i,k}^{(2)}$ (by conventional estimator).

theoretical results.

REFERENCES

[1] D. P. Bertsekas and I. B. Rhodes,Recursive state estimation for a set-membership description of uncertainty, *IEEE Transactions on Automatic Control*, Vol. 16, No. 2, pp. 117–128, 1971.

[2] S. Boyd, L. Ghaoui, E. Feron and V. Balakrishnan, *Linear matrix inequalities in system and control theory*, Philadelphia: SIAM Studies in Applied Mathematics, 1994.

[3] G. Calariore, Reliable localization using set-value nonlinear filters, *IEEE Transactions on Systems, Man and Cybernetics–Part A: Systems and Humans*, Vol. 35, No. 2, pp. 189–197, 2005.

[4] R. Carli, A. Chiuso, L. Schenato and S. Zampieri, Distributed Kalman filtering based on consensus strategies, *IEEE Transactions on Selected Areas in Communications*, Vol. 26, No. 4, pp. 622–633, 2010.

[5] F. S. Cattivelli and A. H. Sayed, Diffusion strategies for distributed Kalman filtering and smoothing, *IEEE Transactions on Automatic Control*, Vol. 55, No. 9, pp. 2069–2084, 2010.

[6] Y. Chen, Z. Chen, Z. Chen and A. Xue, Observer-based passive control of non-homogeneous Markov jump systems with random communication delays, *International Journal of Systems Science*, vol. 51, no. 6, pp. 1133–1147, 2020.

[7] Y. Chen, Z. Wang, Y. Yuan and P. Date, Distributed H_∞ filtering for switched stochastic delayed systems over sensor networks with fading measurements, *IEEE Transactions on Cybernetics*, vol. 50, no. 1, pp. 2–14, 2020.

[8] L. El Ghaoui and G. Calafiore, Robust filtering for discrete-time systems with bounded noise and parametric uncertainty, *IEEE Transactions on Automatic Control*, Vol. 46, No. 7, pp. 1084–1089, 2001.

[9] G. Franze, F. Tedesco and W. Lucia, Resilient control for cyber-physical systems subject to replay attacks, *IEEE Control Systems Letters*, Vol. 3, No. 4, pp. 984–985, 2019.

[10] X. Ge, Q.-L. Han, X. M. Zhang, L. Ding and F. Yang, Distributed event-triggered estimation over sensor networks: a survey, *IEEE Transactions on Cybernetics*, Vol. 50, No. 3, pp. 1306–1320, 2020.

[11] X. Ge, Q.-L. Han, L. Ding, Y. Wang, X.-M. Zhang, Dynamic event-triggered distributed coordination control and its applications: A survey of trends and techniques, IEEE Transactions on Systems, Man, and Cybernetics: Systems, 50 (9), 3112-3125, Sep. 2020.

[12] X. Ge, Q.-L. Han, M. Zhong, X.-M. Zhang, Distributed Krein space-based attack detection over sensor networks under deception attacks, *Automatica*, 109, art. no. 108557, Nov. 2019.

[13] F. Han, Z. Wang, H. Dong and H. Liu, Partial-nodes-based scalable H_∞-consensus filtering with censored measurements over sensor networks, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 3, pp. 1892–1903, 2021.

[14] J. Hu, Z. Wang, G.-P. Liu, H. Zhang and R. Navaratne, A prediction-based approach to distributed filtering with missing measurements and communication delays through sensor networks, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, in press, DOI: 10.1109/TSM-C.2020.2966977.

performance. Finally, an illustrative numerical example has been presented to demonstrate the effectiveness of the obtained

[15] J. Hu, H. Zhang, H. Liu and X. Yu, A survey on sliding mode control for networked control systems, *International Journal of Systems Science*, vol. 52, no. 6, pp. 1129–1147, 2021.

[16] X. Li, F. Han, N. Hou, H. Dong and H. Liu, Set-membership filtering for piecewise linear systems with censored measurements under Round-Robin protocol, *International Journal of Systems Science*, vol. 51, no. 9, pp. 1578–1588, 2020.

[17] M. Jorjani, H. Seifi and A. Varjani, A graph theory-based approach to detect false data injection attacks in power system AC state Eestimation, *IEEE Transactions on Industrial Informatics*, Vol. 17, No. 4, pp. 2465–2475, 2021.

[18] Q. Li, B. Shen, Z. Wang and W. Sheng, Recursive distributed filtering over sensor networks on Gilbert-Elliott channels: A dynamic event-triggered approach, *Automatica*, vol. 113, art. no. 108681, 2020.

[19] X. Li, Y. Zhao, S. Zhang and X. Fan, Adaptive particle filter for nonparametric estimation with measurement uncertainty in wireless sensor networks, *Sensors*, Vol. 16, No. 6, Art.No. 786, 2016.

[20] C. Li, H. Dong, J. Li and F. Wang, Distributed Kalman filtering for sensor network with balanced topology, *Systems & Control Letters*, Vol. 131, Art.No. 104500, 2019.

[21] H. Liu, Z. Wang, W. Fei and J. Li, H_∞ and $l_2 - l_\infty$ state estimation for discrete-time delayed memristive neural networks on finite horizon: the Round-Robin protocol, *Neural Networks*, vol. 132, pp. 121–130, 2020.

[22] L. Liu, L. Ma, J. Zhang and Y. Bo, Distributed non-fragile set-membership filtering for nonlinear systems under fading channels and bias injection attacks, *International Journal of Systems Science*, vol. 52, no. 6, pp. 1192–1205, 2021.

[23] S. Liu, Z. Wang, G. Wei and M. Li, Distributed set-membership filtering for multi-rate systems under the Round-Robin scheduling over sensor networks, *IEEE Transactions on Cybernetics*, vol. 50, no. 5, pp. 1910–1920, 2020.

[24] L. Ma, Z. Wang, Y. Chen and X. Yi, Probability-guaranteed distributed filtering for nonlinear systems with innovation constraints over sensor networks, *IEEE Transactions on Control of Network Systems*, in press, DOI: 10.1109/TCNS.2021.3049361.

[25] L. Ma, Z. Wang, H.-K. Lam and N. Kyriakoulis, Distributed event-based set-membership filtering for a class of nonlinear systems with sensor saturations over sensor networks, *IEEE Transactions on Cybernetics*, Vol. 47, No. 11, pp. 3772–3783, 2017.

[26] L. Ma, Z. Wang, Q.-L. Han and Y. Liu, Distributed filtering for nonlinear time-delay systems over sensor networks subject to multiplicative link noises and switching topology , *International Journal of Robust and Nonlinear Control*, Vol. 29, pp. 2941–2959, 2019.

[27] L. Ma, Z. Wang, J. Hu and Q.-L. Han, Probability-guaranteed envelope-constrained filtering for nonlinear systems subject to measurement outliers, *IEEE Transactions on Automatic Control*, 2021, DOI: 10.1109/TAC.2020.3016767.

[28] J. Mao, Y. Sun, X. Yi, H. Liu and D. Ding, Recursive filtering of networked nonlinear systems: A survey, *International Journal of Systems Science*, vol. 52, no. 6, pp. 1110–1128, 2021.

[29] A. De Paola, S. Gaglio, G. Lo Re, F. Milazzo and M. Ortolani, Adaptive distributed outlier detection for WSNs, *IEEE Transactions on Cybernetics*, Vol. 45, No. 5, pp. 888-899, 2015.

[30] A. Ribeiro, I. Schizas, S. Roumeliotis and G. Giannakis, Kalman filtering in wireless sensor networks, *IEEE Control Systems Magazine*, vol. 30, pp. 66–86, 2010.

[31] S. Seifzadeh, B. Khaleghi and F. Karray, Distributed soft-data-constrained multi-model particle filter, *IEEE Transactions on Cybernetics*, Vol. 45, No. 3, pp. 384-394, 2015.

[32] B. Shen, Z. Wang, D. Wang and H. Liu, Distributed state-saturated recursive filtering over sensor networks under Round-Robin protocol, *IEEE Transactions on Cybernetics*, vol. 50, no. 8, pp. 3605–3615, Aug. 2020.

[33] B. Shen, Z. Wang and Y. S. Hung, Distributed H_∞-consensus filtering in sensor networks with multiple missing measurements: The finite-horizon case, *Automatica*, Vol. 46, No. 10, pp. 1682–1688, 2011.

[34] Y. Sun, J. Mao, H. Liu and D. Ding, Distributed recursive filtering for discrete time-delayed stochastic nonlinear systems based on fuzzy rules, *Neurocomputing*, vol. 400, pp. 412–419, 2020.

[35] H. Tan, B. Shen, K. Peng and H. Liu, Robust recursive filtering for uncertain stochastic systems with amplify-and-forward relays, *International Journal of Systems Science*, vol. 51, no. 7, pp. 1188–1199, 2020.

[36] E. Tian, Z. Wang, L. Zou and D. Yue, Probabilistic-constrained filtering for a class of nonlinear systems with improved static event-triggered communication, *International Journal of Robust and Nonlinear Control*, Vol. 29, No. 5, pp. 1484–1498, 2019.

[37] V. Ugrinovskii, Distributed robust filtering with H_∞ consensus of estimates, *Automatica*, Vol. 47, No. 1, pp. 1–13, 2011.

[38] M. Wang, L. Huang and C. Yang, NN-based adaptive tracking control of discrete-time nonlinear systems with actuator saturation and event-triggering protocol, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, in press, DOI: 10.1109/TSMC.2020.2981954.

[39] G. Wei, W. Li, D. Ding and Y. Liu, Stability analysis of covariance intersection-based Kalman consensus filtering for time-varying systems, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 50, no. 11, pp. 4611–4622, 2020.

[40] I. Yaesha, S. Boyarskia and U. Shaked, Probability-guaranteed robust H_∞ performance analysis and state-feedback design, *Systems and Control Letters*, Vol. 48, pp. 351–364, 2003.

[41] N. D. Zade, S. N. Deshpande and D. Sita, Investigative analysis of suboptimal filter for state estimation in object tracking wireless sensor network, *IEEE Sensors Letters*, Vol. 4, No. 10 pp. 7003604, 2020.

[42] C. Zhao, J. Lam and H. Lin, State estimation of CPSs with deception attacks: Stability analysis and approximate computation, *Neurocomputing*, vol. 423, pp. 318–326, 2021.

[43] D. Zhao, Z. Wang, D. W. C. Ho and G. Wei, Observer-based PID security control for discrete time-delay systems under cyber-attacks, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 6, pp. 3926–3938, 2021.

[44] D. Zhao, Z. Wang, G. Wei and Q.-L. Han, A dynamic event-triggered approach to observer-based PID security control subject to deception attacks, *Automatica*, vol. 120, art. no. 109128, 2020.

[45] Z. Zhao, Z. Wang, L. Zou and J. Guo, Set-membership filtering for time-varying complex networks with uniform quantisations over randomly delayed redundant channels, *International Journal of Systems Science*, vol. 51, no. 16, pp. 3364–3377, 2020.

[46] K. Zhu, J. Hu, Y. Liu, N. D. Alotaibi and F. E. Alsaadi, On ℓ_2-ℓ_∞ output-feedback control scheduled by stochastic communication protocol for two-dimensional switched systems, *International Journal of Systems Science*, in press. DOI: 10.1080/00207721.2021.1914768.

[47] L. Zou, Z. Wang, H. Geng and X. Liu, Set-membership filtering subject to impulsive measurement outliers: A recursive algorithm, *IEEE/CAA Journal of Automatica Sinica*, vol. 8, no. 2, pp. 377–388, 2021.

[48] L. Zou, Z. Wang, J. Hu, Y. Liu and X. Liu, Communication-protocol-based analysis and synthesis of networked systems: Progress, prospects and challenges, *International Journal of Systems Science*, in press, DOI: 10.1080/00207721.2021.1917721.

VI. APPENDIX

A. Proof of Theorem 1

Proof: The proof is carried out by mathematical induction.

First, according to Assumption 3, at time step $k = 0$, $\mathrm{tr}[\tilde{W}_{ij,0}^{\mathrm{T}} Q_{ij,0}^{-1} \tilde{W}_{ij,0}] \leq 1$ holds.

Second, suppose at time step k, $\mathrm{tr}[\tilde{W}_{ij,k}^{\mathrm{T}} Q_{ij,k}^{-1} \tilde{W}_{ij,k}] \leq 1$ is true. Then, we only need to verify that at time step $k+1$, under the condition given in this theorem, the following inequality also holds:

$$\mathrm{tr}[\tilde{W}_{ij,k+1}^{\mathrm{T}} Q_{ij,k+1}^{-1} \tilde{W}_{ij,k+1}] \leq 1. \tag{34}$$

To this end, it is readily from the tuning law (17) that

$$
\begin{aligned}
\tilde{W}_{ij,k+1} &= W_{ij} - \hat{W}_{ij,k+1} \\
&= W_{ij} - \left(\lambda_{ij,k}^{(1)}\hat{W}_{ij,k} + \lambda_{ij,k}^{(2)}\varpi_k\phi^{\mathrm{T}}(z_{i,k})\right) \\
&= (1 - \lambda_{ij,k}^{(1)})W_{ij} + \lambda_{ij,k}^{(1)}\tilde{W}_{ij,k} - \lambda_{ij,k}^{(2)}\varpi_k\phi^{\mathrm{T}}(z_{i,k}).
\end{aligned} \tag{35}
$$

Define a function $\varphi(\cdot) : \mathbb{R}^{m \times n} \mapsto \mathbb{R}^{mn}$ as follows:

$$\varphi(A) \triangleq \begin{bmatrix} A^{(1)} & A^{(2)} & \cdots & A^{(m)} \end{bmatrix}^{\mathrm{T}}$$

where $A^{(\iota)}$ stands for the ι-th row of matrix A. Then, by denoting

$$\varphi(\tilde{W}_{ij,k}) \triangleq \begin{bmatrix} \tilde{W}_{ij,k}^{(1)} & \tilde{W}_{ij,k}^{(2)} & \cdots & \tilde{W}_{ij,k}^{(n_y)} \end{bmatrix}^{\mathrm{T}},$$

we know that the following inequality

$$\text{tr}[\tilde{W}_{ij,k}^{\mathrm{T}} Q_{ij,k}^{-1} \tilde{W}_{ij,k}] \le 1$$

can be equivalently expressed by

$$\varphi^{\mathrm{T}}(\tilde{W}_{ij,k}) \mathcal{Q}_{ij,k}^{-1} \varphi(\tilde{W}_{ij,k}) \le 1. \qquad (36)$$

Consequently, according to [3], taking into account $\mathcal{Q}_{ij,k} = M_{ij,k} M_{ij,k}^{\mathrm{T}}$, we have

$$\varphi(\tilde{W}_{ij,k}) = M_{ij,k} r_{ij,k} \qquad (37)$$

where $r_{ij,k}$ satisfies

$$r_{ij,k}^{\mathrm{T}} r_{ij,k} \le 1. \qquad (38)$$

By defining a vector $\pi_{ij,k}$ as follows:

$$\pi_{ij,k} \triangleq \begin{bmatrix} 1 & r_{ij,k}^{\mathrm{T}} & \varphi^{\mathrm{T}}(W_{ij}) \end{bmatrix}^{\mathrm{T}}, \qquad (39)$$

we rewrite the error dynamics (35) by

$$\varphi(\tilde{W}_{ij,k+1}) = \Upsilon_{ij,k} \pi_{ij,k}. \qquad (40)$$

Next, in terms of $\pi_{ij,k}$, the inequality $r_{ij,k}^{\mathrm{T}} r_{ij,k} \le 1$ can be reformulated by

$$\pi_{ij,k}^{\mathrm{T}} \bar{M}_{ij,k} \pi_{ij,k} \le 0 \qquad (41)$$

where $\bar{M}_{ij,k} \triangleq \text{diag}\{-1, I, 0\}$.

Similarly, based on Assumption 2, one has

$$\text{tr}[W_{ij}^{\mathrm{T}} W_{ij}] \le \epsilon_{1j}^2, \qquad (42)$$

which, in terms of $\pi_{ij,k}$, is equivalently described by

$$\pi_{ij,k}^{\mathrm{T}} \bar{W}_{ij,k} \pi_{ij,k} \le 0 \qquad (43)$$

where $\bar{W}_{ij,k} \triangleq \text{diag}\{-\epsilon_{1j}^2, 0, I\}$.

On the other hand, it is easily inferred that

$$\text{tr}[\tilde{W}_{ij,k+1}^{\mathrm{T}} Q_{ij,k+1}^{-1} \tilde{W}_{ij,k+1}] \le 1 \qquad (44)$$

is equivalent to

$$\varphi^{\mathrm{T}}(\tilde{W}_{ij,k+1}) \mathcal{Q}_{ij,k+1}^{-1} \varphi(\tilde{W}_{ij,k+1}) \le 1, \qquad (45)$$

which can be further expressed as follows:

$$\pi_{ij,k}^{\mathrm{T}} (\Upsilon_{ij,k}^{\mathrm{T}} \mathcal{Q}_{ij,k+1}^{-1} \Upsilon_{ij,k} - \text{diag}\{1,0,0\}) \pi_{ij,k} \le 0. \qquad (46)$$

Therefore, we know from Lemma 1 (S-procedure) that, if there exist positive scalars $\varrho_{ij,k}^{(1)}$ and $\varrho_{ij,k}^{(2)}$ such that the following inequality holds:

$$\Upsilon_{ij,k}^{\mathrm{T}} \mathcal{Q}_{ij,k+1}^{-1} \Upsilon_{i,k} - \text{diag}\{1,0,0\}$$
$$- \varrho_{ij,k}^{(1)} \bar{M}_{ij,k} - \varrho_{ij,k}^{(2)} \bar{W}_{ij,k} \le 0, \qquad (47)$$

then inequality (45) also holds.

Finally, according to Lemma 2 (Schur Complement Equivalence), we derive that (47) holds if and only if (18) holds. The proof is complete now. ∎

B. Proof of Theorem 2

Proof: The proof is performed by resorting to mathematical induction approach. First, it is readily inferred from Assumption 3 that

$$\mathbb{E}\{\tilde{x}_{i,0}^{\mathrm{T}} P_0^{-1} \tilde{x}_{i,0}\} \le 1. \qquad (48)$$

Second, suppose that at step $k > 0$, the following inequality holds:

$$\mathbb{E}\{\tilde{x}_{i,k}^{\mathrm{T}} P_k^{-1} \tilde{x}_{i,k}\} \le 1. \qquad (49)$$

Then, on basis of the principle of mathematical induction, the rest of the proof is to demonstrate that, under the given condition, inequality (29) also holds at time step $k+1$. In fact, since (49) is true, we can always find a vector $s_{i,k} \in \mathbb{R}^{n_s}$ ($i = 1, 2, \ldots, N$) with $\mathbb{E}\{s_{i,k}^{\mathrm{T}} s_{i,k}\} \le 1$ such that

$$x_k = \hat{x}_{i,k} + S_k s_{i,k}. \qquad (50)$$

By denoting $s_k \triangleq \text{col}_N\{s_{i,k}\}$ and $\mathcal{S}_k \triangleq \text{diag}_N\{S_k\}$, (50) is described by

$$\xi_k = \hat{x}_k + \mathcal{S}_k s_k. \qquad (51)$$

Hence, (16) is rewritten by

$$\begin{aligned}
\tilde{x}_{k+1} =& \hat{f}_k - \mathcal{F}_k \hat{x}_k + (\Phi_k - \mathcal{G}_k \Psi_k - \mathcal{H}_k \Psi_k) \mathcal{S}_k s_k \\
& + \mathcal{L}\rho_{1k} - \mathcal{G}_k \Sigma \rho_{2k} - \mathcal{H}_k \Sigma \rho_{2k} + \mathcal{B}_k \nu_k \\
& - (\mathcal{G}_k + \mathcal{H}_k) \mathcal{E}_k (\mathbf{1}_N \otimes I_{n_\mu}) \mu_k - \alpha_k \mathcal{H}_k \delta_k \\
& - \alpha_k \mathcal{H}_k \circ \tilde{\mathcal{W}}_k \phi_k - \alpha_k \mathcal{H}_k \circ \hat{\mathcal{W}}_k \phi_k \\
& + \bar{\alpha} \sum_{i=1}^N \Theta_{n_x,i} \mathcal{H}_k \circ \hat{\mathcal{W}}_k (\mathbf{1}_N \otimes \mathcal{R}_{n_y,i}) \phi_k \\
=& \hat{f}_k - \mathcal{F}_k \hat{x}_k + (\Phi_k - \mathcal{G}_k \Psi_k - \mathcal{H}_k \Psi_k) \mathcal{S}_k s_k \\
& + \mathcal{L}\rho_{1k} - \mathcal{G}_k \Sigma \rho_{2k} - \mathcal{H}_k \Sigma \rho_{2k} + \mathcal{B}_k \nu_k \\
& - (\mathcal{G}_k + \mathcal{H}_k) \mathcal{E}_k (\mathbf{1}_N \otimes I_{n_\mu}) \mu_k - (\bar{\alpha} + \tilde{\alpha}_k) \mathcal{H}_k \delta_k \\
& - (\bar{\alpha} + \tilde{\alpha}_k)(\mathcal{H}_k \circ \tilde{\mathcal{W}}_k) \phi_k - (\bar{\alpha} + \tilde{\alpha}_k)(\mathcal{H}_k \circ \hat{\mathcal{W}}_k) \phi_k \\
& + \bar{\alpha} \sum_{i=1}^N \Theta_{n_x,i} (\mathcal{H}_k \circ \hat{\mathcal{W}}_k)(\mathbf{1}_N \otimes \mathcal{R}_{n_y,i}) \phi_k \\
=& \hat{f}_k - \mathcal{F}_k \hat{x}_k + (\Phi_k - \mathcal{G}_k \Psi_k - \mathcal{H}_k \Psi_k) \mathcal{S}_k s_k \\
& + \mathcal{L}\rho_{1k} - \mathcal{G}_k \Sigma \rho_{2k} - \mathcal{H}_k \Sigma \rho_{2k} + \mathcal{B}_k \nu_k \\
& - (\mathcal{G}_k + \mathcal{H}_k) \mathcal{E}_k (\mathbf{1}_N \otimes I_{n_\mu}) \mu_k - (\bar{\alpha} + \tilde{\alpha}_k) \mathcal{H}_k \delta_k \\
& - (\bar{\alpha} + \tilde{\alpha}_k) \sum_{i=1}^N \Theta_{n_x,i} \mathcal{H}_k \tilde{\mathcal{W}}_{i,k} \phi_k \\
& - (\bar{\alpha} + \tilde{\alpha}_k)(\mathcal{H}_k \circ \hat{\mathcal{W}}_k) \phi_k \\
& + \bar{\alpha} \sum_{i=1}^N \Theta_{n_x,i} \mathcal{H}_k \circ \hat{\mathcal{W}}_k (\mathbf{1}_N \otimes \mathcal{R}_{n_y,i}) \phi_k \qquad (52)
\end{aligned}$$

where

$$\tilde{\mathcal{W}}_{i,k} \triangleq \text{diag}\{\tilde{W}_{i1,k}, \tilde{W}_{i2,k}, \ldots, \tilde{W}_{iN,k}\},$$
$$\rho_{1k} \triangleq \text{col}_N\{\rho_{1i,k}\} = \Delta_1 \mathcal{S}_k s_k,$$
$$\rho_{2k} \triangleq \text{col}_N\{\rho_{2i,k}\} = \Delta_2 \mathcal{S}_k s_k.$$

It is known that, under the condition of Theorem 1, we acquire

$$\varphi(\tilde{W}_{ij,k}) = M_{ij,k} r_{ij,k} \qquad (53)$$

with vector $r_{ij,k}$ satisfying $r_{ij,k}^{\mathrm{T}}r_{ij,k} \leq 1$.

By defining

$$\tilde{\phi}_k \triangleq I_{n_y} \otimes \phi^{\mathrm{T}}(z_{i,k}),$$

we have

$$\tilde{W}_{ij,k}\phi_i = \tilde{\phi}_k\varphi(\tilde{W}_{ij,k}) = \tilde{\phi}_k M_{ij,k}r_{ij,k}.$$

Therefore, one obtains

$$\tilde{\mathcal{W}}_{i,k}\phi_k = \mathrm{diag}_N\{\tilde{\phi}_k\} \cdot \mathrm{diag}\{M_{i1,k}, M_{i2,k}, \ldots, M_{iN,k}\} \cdot r_{i,k}$$
$$\triangleq \mathcal{M}_{i,k}r_{i,k}$$

where $r_{i,k} \triangleq \left[\begin{array}{cccc} r_{i1,k}^{\mathrm{T}} & r_{i2,k}^{\mathrm{T}} & \cdots & r_{iN,k}^{\mathrm{T}} \end{array}\right]^{\mathrm{T}}$.

By defining a new vector ζ_k as

$$\zeta_k \triangleq \left[\begin{array}{cccccccc} 1 & s_k^{\mathrm{T}} & \rho_{1k}^{\mathrm{T}} & \rho_{2k}^{\mathrm{T}} & \nu_k^{\mathrm{T}} & \mu_k^{\mathrm{T}} & \delta_k^{\mathrm{T}} & r_k^{\mathrm{T}} \end{array}\right]^{\mathrm{T}}$$

where $r_k \triangleq \left[\begin{array}{cccc} r_{1,k}^{\mathrm{T}} & r_{2,k}^{\mathrm{T}} & \cdots & r_{N,k}^{\mathrm{T}} \end{array}\right]^{\mathrm{T}}$, the dynamics of state estimation error are formulated as follows:

$$\tilde{x}_{k+1} = \Xi_k\zeta_k \triangleq (\bar{\Xi}_k + \tilde{\alpha}_k\tilde{\Xi}_k)\zeta_k \qquad (54)$$

where

$$\Xi_k \triangleq \left[\begin{array}{ccccc} \Xi_k^{(11)} & \Xi_k^{(12)} & \mathcal{L} & -(\mathcal{G}_k + \mathcal{H}_k)\Sigma & \mathcal{B}_k \end{array}\right.$$
$$\left.\begin{array}{ccc} \Xi_k^{(16)} & -(\bar{\alpha} + \tilde{\alpha}_k)\mathcal{H}_k & -(\bar{\alpha} + \tilde{\alpha}_k)\mathscr{H}_k \end{array}\right],$$

$$\Xi_k^{(11)} \triangleq \hat{f}_k - \mathcal{F}_k\hat{x}_k - (\bar{\alpha} + \tilde{\alpha}_k)(\mathcal{H}_k \circ \hat{\mathcal{W}}_k)\phi_k$$
$$+ \bar{\alpha}\sum_{i=1}^N \Theta_{n_x,i}(\mathcal{H}_k \circ \hat{\mathcal{W}}_k)(\mathbf{1}_N \otimes \mathcal{R}_{n_y,i})\phi_k,$$

$$\Xi_k^{(12)} \triangleq (\Phi_k - \mathcal{G}_k\Psi_k - \mathcal{H}_k\Psi_k)\mathcal{S}_k,$$
$$\Xi_k^{(16)} \triangleq -(\mathcal{G}_k + \mathcal{H}_k)\mathcal{E}_k(\mathbf{1}_N \otimes I_{n_\mu}),$$
$$\bar{\Xi}_k \triangleq \left[\begin{array}{ccccc} \bar{\Xi}_k^{(11)} & \Xi_k^{(12)} & \mathcal{L} & -(\mathcal{G}_k + \mathcal{H}_k)\Sigma & \mathcal{B}_k \end{array}\right.$$
$$\left.\begin{array}{ccc} \Xi_k^{(16)} & -\bar{\alpha}\mathcal{H}_k & -\bar{\alpha}\mathscr{H}_k \end{array}\right],$$

$$\bar{\Xi}_k^{(11)} \triangleq \hat{f}_k - \mathcal{F}_k\hat{x}_k - \bar{\alpha}(\mathcal{H}_k \circ \hat{\mathcal{W}}_k)\phi_k$$
$$+ \bar{\alpha}\sum_{i=1}^N \Theta_{n_x,i}(\mathcal{H}_k \circ \hat{\mathcal{W}}_k)(\mathbf{1}_N \otimes \mathcal{R}_{n_y,i})\phi_k,$$

$$\tilde{\Xi}_k \triangleq -\left[\begin{array}{cccccccc} (\mathcal{H}_k \circ \hat{\mathcal{W}}_k)\phi_k & 0 & 0 & 0 & 0 & 0 & \mathcal{H}_k & \mathscr{H}_k \end{array}\right],$$
$$\mathscr{H}_{i,k} \triangleq \Theta_{n_x,i}\mathcal{H}_k\mathcal{M}_{i,k},$$
$$\mathscr{H}_k \triangleq \left[\begin{array}{cccc} \mathscr{H}_{1,k} & \mathscr{H}_{2,k} & \cdots & \mathscr{H}_{N,k} \end{array}\right].$$

It is already known that the following conditions hold:

$$\mathbb{E}\{s_{i,k}^{\mathrm{T}}s_{i,k}\} \leq 1,$$
$$\nu_k^{\mathrm{T}}V_k^{-1}\nu_k \leq 1,$$
$$\mu_k^{\mathrm{T}}U_k^{-1}\mu_k \leq 1,$$
$$r_{ij,k}^{\mathrm{T}}r_{ij,k} \leq 1,$$
$$\|\delta_i\| \leq \epsilon_{3_i},$$

which can be equivalently expressed by utilizing ζ_k as follows:

$$\mathbb{E}\left\{\zeta_k^{\mathrm{T}}\mathrm{diag}\left\{-1, \Theta_{n_s,i}, 0, 0, 0, 0, 0, 0\right\}\zeta_k\right\} \leq 0,$$
$$\zeta_k^{\mathrm{T}}\mathrm{diag}\{-1, 0, 0, 0, V_k^{-1}, 0, 0, 0\}\zeta_k \leq 0,$$

$$\zeta_k^{\mathrm{T}}\mathrm{diag}\{-1, 0, 0, 0, 0, U_k^{-1}, 0, 0\}\zeta_k \leq 0,$$
$$\zeta_k^{\mathrm{T}}\mathrm{diag}\{-N^2, 0, 0, 0, 0, 0, 0, I_{N^2}\}\zeta_k \leq 0,$$
$$\zeta_k^{\mathrm{T}}\mathrm{diag}\{-\epsilon_{3i}, 0, 0, 0, 0, 0, \Theta_{n_\delta,i}, 0\}\zeta_k \leq 0.$$

Moreover, from

$$\|\Delta_{1i}\| \leq 1 \quad \text{and} \quad \|\Delta_{2i}\| \leq 1$$

we have

$$\rho_{1i,k}^{\mathrm{T}}\rho_{1i,k} - s_{i,k}^{\mathrm{T}}S_k^{\mathrm{T}}S_k s_{i,k} \leq 0,$$
$$\rho_{2i,k}^{\mathrm{T}}\rho_{2i,k} - s_{i,k}^{\mathrm{T}}S_k^{\mathrm{T}}S_k s_{i,k} \leq 0,$$

which can be equivalently described by

$$\zeta_k^{\mathrm{T}}\mathrm{diag}\left\{0, -\Theta_{n_s,i}\mathcal{S}_k^{\mathrm{T}}\mathcal{S}_k, \Theta_{n_l,i}, 0, 0, 0, 0, 0\right\}\zeta_k \leq 0,$$
$$\zeta_k^{\mathrm{T}}\mathrm{diag}\left\{0, -\Theta_{n_s,i}\mathcal{S}_k^{\mathrm{T}}\mathcal{S}_k, 0, \Theta_{n_\Sigma,i}, 0, 0, 0, 0\right\}\zeta_k \leq 0.$$

Next, by virtue of Lemma 2 (Schur Complement Equivalence), we know that inequalities (22) hold if and only if

$$\bar{\Xi}_k^{\mathrm{T}}\mathcal{R}_{n_x,i}^{\mathrm{T}}P_{k+1}^{-1}\mathcal{R}_{n_x,i}\bar{\Xi}_k$$
$$+ \sigma^2\tilde{\Xi}_k^{\mathrm{T}}\mathcal{R}_{n_x,i}^{\mathrm{T}}P_{k+1}^{-1}\mathcal{R}_{n_x,i}\tilde{\Xi}_k - \Gamma_k \leq 0, \qquad (55)$$

which, by considering the statistical property of $\tilde{\alpha}_k$ (i.e., formula (21)), implies that

$$\mathbb{E}\{(\bar{\Xi}_k + \tilde{\alpha}_k\tilde{\Xi}_k)^{\mathrm{T}}\mathcal{R}_{n_x,i}^{\mathrm{T}}P_{k+1}^{-1}\mathcal{R}_{n_x,i}(\bar{\Xi}_k + \tilde{\alpha}_k\tilde{\Xi}_k)\} - \Gamma_k \leq 0. \qquad (56)$$

By taking (23) and (54) into account, inequality (56) is equivalent to

$$\mathbb{E}\{\tilde{x}_{i,k+1}^{\mathrm{T}}P_{k+1}^{-1}\tilde{x}_{i,k+1}\} - \zeta_k^{\mathrm{T}}\mathrm{diag}\{1, 0, 0, 0, 0, 0, 0, 0\}\zeta_k$$
$$- \sum_{i=1}^N \beta_{i,k}^{(1)}\zeta_k^{\mathrm{T}}\mathrm{diag}\{-1, \Theta_{n_s,i}, 0, 0, 0, 0, 0, 0\}\zeta_k$$
$$- \beta_k^{(2)}\zeta_k^{\mathrm{T}}\mathrm{diag}\{-1, 0, 0, 0, V_k^{-1}, 0, 0, 0\}\zeta_k$$
$$- \beta_k^{(3)}\zeta_k^{\mathrm{T}}\mathrm{diag}\{-1, 0, 0, 0, 0, U_k^{-1}, 0, 0\}\zeta_k$$
$$- \beta_k^{(4)}\zeta_k^{\mathrm{T}}\mathrm{diag}\{-N^2, 0, 0, 0, 0, 0, 0, I_{N^2}\}\zeta_k$$
$$- \sum_{i=1}^N \beta_{i,k}^{(5)}\zeta_k^{\mathrm{T}}\mathrm{diag}\{-\epsilon_{3i}, 0, 0, 0, 0, 0, \Theta_{n_\delta,i}, 0\}\zeta_k$$
$$- \sum_{i=1}^N \beta_{i,k}^{(6)}\zeta_k^{\mathrm{T}}\mathrm{diag}\{0, -\Theta_{n_s,i}\mathcal{S}_k^{\mathrm{T}}\mathcal{S}_k, \Theta_{n_l,i}, 0, 0, 0, 0, 0\}\zeta_k$$
$$- \sum_{i=1}^N \beta_{i,k}^{(7)}\zeta_k^{\mathrm{T}}\mathrm{diag}\{0, -\Theta_{n_s,i}\mathcal{S}_k^{\mathrm{T}}\mathcal{S}_k, 0, \Theta_{n_\Sigma,i}, 0, 0, 0, 0\}\zeta_k$$
$$\leq 0. \qquad (57)$$
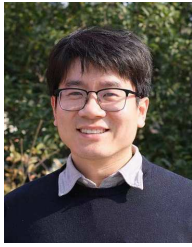
We now acquire from Lemma 1 (S-procedure) that

$$\mathbb{E}\{\tilde{x}_{i,k+1}^{\mathrm{T}}P_{k+1}^{-1}\tilde{x}_{i,k+1}\}$$
$$- \zeta_k^{\mathrm{T}}\mathrm{diag}\{1, 0, 0, 0, 0, 0, 0, 0\}\zeta_k \leq 0, \qquad (58)$$

or equivalently,

$$\mathbb{E}\{\tilde{x}_{i,k+1}^{\mathrm{T}}P_{k+1}^{-1}\tilde{x}_{i,k+1}\} \leq 1, \qquad (59)$$

which indicates the proof is complete. ∎

Lifeng Ma received the B.Sc. degree in Automation from Jiangsu University, Zhenjiang, China, in 2004 and the Ph.D. degree in Control Science and Engineering from Nanjing University of Science and Technology, Nanjing, China, in 2010. From August 2008 to February 2009, he was a Visiting Ph.D. Student in the Department of Information Systems and Computing, Brunel University London, U.K. From January 2010 to April 2010 and May 2011 to September 2011, he was a Research Associate in the Department of Mechanical Engineering, the University of Hong Kong. From March 2015 to Februry 2017, he was a Visiting Research Fellow at the King's College London, U.K.

He is currently a Professor in the School of Automation, Nanjing University of Science and Technology, Nanjing, China. His current research interests include control and signal processing, machine learning and deep learning. He has published more than 50 papers in refereed international journals. He serves as an editor for *Neurocomputing* and *International Journal of Systems Science*.

Zidong Wang (SM'03-F'14) was born in Jiangsu, China, in 1966. He received the B.Sc. degree in mathematics in 1986 from Suzhou University, Suzhou, China, and the M.Sc. degree in applied mathematics in 1990 and the Ph.D. degree in electrical engineering in 1994, both from Nanjing University of Science and Technology, Nanjing, China.

He is currently Professor of Dynamical Systems and Computing in the Department of Computer Science, Brunel University London, U.K. From 1990 to 2002, he held teaching and research appointments in universities in China, Germany and the UK. Prof. Wang's research interests include dynamical systems, signal processing, bioinformatics, control theory and applications. He has published more than 600 papers in international journals. He is a holder of the Alexander von Humboldt Research Fellowship of Germany, the JSPS Research Fellowship of Japan, William Mong Visiting Research Fellowship of Hong Kong.

Prof. Wang serves (or has served) as the Editor-in-Chief for *International Journal of Systems Science*, the Editor-in-Chief for *Neurocomputing*, the Editor-in-Chief for *Systems Science & Control Engineering*, and an Associate Editor for 12 international journals including IEEE Transactions on Automatic Control, IEEE Transactions on Control Systems Technology, IEEE Transactions on Neural Networks, IEEE Transactions on Signal Processing, and IEEE Transactions on Systems, Man, and Cybernetics-Part C. He is a Member of the Academia Europaea, a Member of the European Academy of Sciences and Arts, an Academician of the International Academy for Systems and Cybernetic Sciences, a Fellow of the IEEE, a Fellow of the Royal Statistical Society and a member of program committee for many international conferences.

Yun Chen was born in Zhejiang Province, China. He received the B.E. degree in thermal engineering in 1999 from Central South University of Technology (Central South University), Changsha, China, and the M.E. degree in engineering thermal physics in 2002 and Ph.D. degree in control science and engineering in 2008, both from Zhejiang University, Hangzhou, China.

From August 2009 to August 2010, he was a visiting fellow with the School of Computing, Engineering and Mathematics, University of Western Sydney, Australia. From December 2016 to December 2017, he was an academic visitor with the Department of Mathematics, Brunel University London, UK. In 2002, he joined Hangzhou Dianzi University, China, where he is currently a Professor. His research interests include stochastic and hybrid systems, robust control and filtering.

Xiaojian Yi was born in 1987. He received the B.S. degree in control technology in 2010 from the North University of China, Taiyuan, China, and the M.S. degree in 2012 and Ph.D. degree in 2016 both in reliability engineering from Beijing Institute of Technology, Beijing, China. During 2015-2016, he was a jointly trained PhD student in the University of Ottawa, Canada, to study robot reliability and maintenance. From 2016 to 2020, he was an Associate Professor with the China North Vehicle Research Institute. He is currently an Associate Professor with the Beijing Institute of Technology, Beijing, China. He is the author of two books and more than 100 articles, and is also the holder of 8 patents. His research interests include system reliability analysis, intelligent control, fault diagnosis and health management.