Security Analysis of Distributed Consensus Filtering Under Replay Attacks

Huang, Jiahao; Yang, Wen; Ho, Daniel W. C.; Li, Fangfei; Tang, Yang

Published in:
IEEE Transactions on Cybernetics

Accepted/In press/Filed: 13/09/2023

Document Version:
Post-print, also known as Accepted Author Manuscript, Peer-reviewed or Author Final version

Publication record in CityU Scholars:
Go to record

Published version (DOI):
10.1109/TCYB.2022.3209820

Publications details:

Citing this paper
Please note that where the full-text provided on CityU Scholars is the Post-print version (also known as Accepted Author Manuscript, Peer-reviewed or Author Final version), it may differ from the Final Published version. When citing, ensure that you check and use the publisher's definitive version for pagination and other details.

General rights
Copyright for the publications made accessible via the CityU Scholars portal is retained by the author(s) and/or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights. Users may not further distribute the material or use it for any profit-making activity or commercial gain.

Publisher permission
Permission for previously published items are in accordance with publisher's copyright policies sourced from the SHERPA RoMEO database. Links to full text versions (either Published or Post-print) are only available if corresponding publishers allow open access.

Take down policy
Contact lbscholars@cityu.edu.hk if you believe that this document breaches copyright and provide us with details. We will remove access to the work immediately and investigate your claim.

Download date: 11/05/2024
Security Analysis of Distributed Consensus Filtering under Replay Attacks

Jiahao Huang, Wen Yang, Member, IEEE, Daniel W. C. Ho, Fellow, IEEE, Fangfei Li, Member, IEEE, and Yang Tang, Senior Member, IEEE

Abstract—This work studies the security of consensus-based distributed filtering under the replay attack, which can freely select a part of sensors and modify their measurements into previously recorded ones. We analyze the performance degradation of distributed estimation caused by the replay attack, and utilize the Kullback-Leibler (K-L) divergence to quantify the attack stealthiness. Specifically, for a stable system, we prove that under any replay attack, the estimation error is not only bounded, but also can re-enter the steady state. In that case, we prove that the replay attack is ε-stealthy, where ε can be calculated based on two Lyapunov equations. On the other hand, for an unstable system, we prove that the trace of estimation error covariance is lower bounded by an exponential function, which indicates that the estimation error may diverge due to the attack. In view of this, we provide a sufficient condition to ensure that any replay attack is detectable. Furthermore, we analyze the case that the adversary starts to attack only if the current measurement is close to a previously recorded one. Finally, we verify the theoretical results via several numerical simulations.

Index Terms—Cyber-physical systems, cyber security, replay attack, distributed consensus filtering.

I. INTRODUCTION

OVER the past few decades, due to huge advantages including real-time perception, dynamic control, and high working efficiency, Cyber-Physical Systems (CPSs) have been greatly developed in many practical applications such as smart grids and environmental monitoring [1], [2]. However, due to the interconnection between the communication layers and the physical systems, CPSs are also vulnerable to malicious third parties, which may cause serious consequences to industrial production and life safety. A variety of recently reported security incidents, e.g., the StuxNet worm [3], evidently reflect that the security is unprecedentedly important to CPSs.

This work is supported in part by the Natural Science Foundation of China (62233005, 62173142), National Natural Science Fund for Distinguished Young Scholars (61725301), the Program of Shanghai Academic Research Leader under Grant 20XD1401300, the Programme of Introducing Talents of Discipline to Universities (the 111 Project) under Grant B17017, and the Special Project of Discipline to Universities (the 111 Project) under Grant B17017 and Leader under Grant 20XD1401300, the Programme of Introducing Talents Young Scholars (61725301), the Program of Shanghai Academic Research (62233005, 62173142), National Natural Science Fund for Distinguished Scholars under Grant 61725301, the Key Laboratory of Smart Manufacturing in Energy Chemical Process, Ministry of Education, East China University of Science and Technology, Shanghai 200237, China (e-mail: madaniel@cityu.edu.hk). Fangfei Li is with the Department of Mathematics, East China University of Science and Technology, Shanghai 200237, China (e-mail: li_fangfei@163.com).

There exist two main types of attacks: Denial-of-Service (DoS) attacks and integrity attacks [4]. DoS attacks prevent information exchange by blocking wireless channels, and their impact on estimation or control systems has been studied in existing works such as [5]–[9]. Integrity attacks modify the content of transmitted data, and can be further categorized into false data injection attacks and replay attacks. Due to the widespread usage of false data detectors, integrity attacks need to be designed to remain stealthy. In the remote estimation scenario, [10] proposed an innovation-based integrity attack, which degrades estimation quality and deceives residue-based detectors. [11] analyzed the tradeoff between detectability and control performance in stochastic control systems under false data injection attacks. To mitigate the impact of attacks, many efforts have been devoted to the design of detection mechanisms. For instance, to ensure the security of remote estimation, [12], [13] proposed their detection schemes based on causality and watermarking, respectively. Based on trusted sensors, [14] proposed some data verification algorithms to protect the multi-sensor state estimation against attacks. Due to huge advantages such as scalability, distributed filtering has become a hot research topic [15]. However, the distributed structure also brings more risks to the state estimation, since any undetectable attack may spread its impact to the entire network [16]. Hence, many works have studied how to protect the distributed filtering against integrity attacks [17]–[24]. For instance, [18] proposed a saturated innovation update algorithm to improve the resilience of distributed filtering under false data injection attacks. To ensure the stability of distributed filtering in linear Gaussian systems, [22] proposed a defense method to resist the attack that injects Gaussian noises into sensor networks.

As a kind of integrity attacks, the replay attack can replace actual signals with previous ones. Compared with other attacks, the replay attack does not require sufficient system knowledge and consumes fewer computing resources. Besides, without the need to identify system parameters, the replay attack can be easily implemented. In closed-loop control systems, [25]–[28] proved that the replay attack can bypass residue-based detectors, and proposed a watermarking approach to improve the detection rate at the cost of sacrificing some control performance. By employing the zero-sum game, [29] proposed a switching control policy to balance the control performance and detection rate of the replay attack. Moreover, [30] utilized the model predictive control to improve the resilience of closed-loop control systems against the replay attack. In multi-agent systems, [31], [32] developed some
resilient algorithms based on model predictive control, so that agents can still achieve consensus under the replay attack. [33] proposed an event-based control protocol to ensure the mean-square exponential consensus of multi-agent systems subject to both replay attacks and DoS attacks.

We can summarize that in the existing works, the replay attack is mainly studied in closed-loop control systems [25]–[30] and multi-agent systems [31]–[33], rather than distributed estimation. However, such an attack is likely to be adopted to damage distributed estimation, especially when the adversary lacks sufficient system knowledge or computing resources. Motivated by this, we are interested in studying the security of distributed filtering under replay attacks from the aspects of estimation performance and attack detectability. Specifically, to quantify the impact of replay attacks on distributed filtering, we derive the analytical expression of the estimation error covariance. Then, we adopt the K-L divergence to measure the stealthiness of the replay attack. Note that compared with [25]–[30], we consider a more general and complex scenario, where multiple sensors form a distributed sensor network, and a part of them may be tampered by the replay attack. Hence, due to inherent coupling of sensor networks, we need to further utilize the graph theory for security analysis. Moreover, due to the coupling relationship between the state estimate and the data replayed by the attacker, our results show that the estimation error covariance under the replay attack has a more complicated structure than those under other types of attacks in [17]–[24]. Accordingly, to derive some intuitive theoretical results, it is necessary to develop some mathematical methods to address the above challenge.

The main contributions of this work are listed as follows:

1) From the perspectives of estimation performance and attack detectability, we analyze the security of distributed consensus filter under the replay attack. Compared with [25], [26], we consider a more complex and general system that is composed of distributed sensor networks, and propose a more general replay attack model that allows the attacker to freely select the attacked sensors. Compared with [34], [35], we consider a more general case that the distributed consensus filter is subject to attacks, and introduce a new index to measure the attack stealthiness, i.e., the K-L divergence.

2) For a stable system, we prove that the estimation error is bounded under any replay attack (Theorem 1) and can even return to the steady state (Theorem 2). In addition, we prove that the corresponding K-L divergence is bounded as well (Theorem 3), and the replay attack is ε-stealthy, where ε can be obtained by two Lyapunov equations (Theorem 4).

3) For an unstable system, although the estimation error may diverge due to the replay attack (Theorem 5), we can ensure that any attack in the proposed attack model is detectable when system parameters satisfy the proposed sufficient condition (Theorem 6). Then, we extend the security analysis to another different attack scenario, where the replay attack can elaborately select the time to begin. In that case, no matter how to design system parameters, it is proved that the replay attack has at least one strategy to keep strictly stealthy (Theorem 7). Hence, a corresponding countermeasure is further provided to solve this system vulnerability.

The remainder of the paper is organized as follows. Section II introduces the system model. For the stable and unstable systems, Sections III and IV analyze the impact of the replay attack on estimation performance and evaluate its stealthiness, respectively. Numerical examples are given in Section V. Some concluding remarks are provided in Section VI.

Notations: Let $\mathbb{N}$ and $\mathbb{R}$ be the sets of nonnegative integers and real numbers, respectively. $\mathbb{R}^n$ is the $n$-dimensional Euclidean space. For a matrix $X$, its 2-norm, trace, determinant, transpose, and inverse matrix are $\|X\|_2$, $\text{tr}(X)$, $\det(X)$, $X^T$ and $X^{-1}$, respectively. If $X$ is positive semi-definite (positive definite), we write $X \succeq 0$ (or $X > 0$). $\text{diag}(X_i)$ denotes the block diagonal matrix with its main diagonal elements varying with $X_i$. $\lambda_i(X)$ stands for the $i$th eigenvalue of $X$ such that $\lambda_1(X) \leq \ldots \leq \lambda_n(X)$, where the maximal and minimal ones are rewritten as $\lambda_{\max}(X)$ and $\lambda_{\min}(X)$, respectively. $\text{vec}(X)$ is a linear transformation that converts $X$ into a column vector. For matrices $X$ and $Y$, $X \otimes Y$ is their Kronecker product. $I_N$ denotes the identity matrix. $1_N$ is a vector where all elements equal to one. $\theta_i$ is a vector where its $i$th element equals to one and all the other elements equal to zero. $\mathbb{E}[\cdot]$ denotes the expectation of a random variable. For a set $S$, its cardinality is $\text{card}(S)$. For a scalar $x \in \mathbb{R}$, $|x|$ denotes its absolute value. $e$ means the natural logarithm, and $\log(\cdot)$ denotes the log function. To simplify the symmetrical notation, $(X)Y(\cdot)^T$ means $(X)Y(X)^T$, and $(X)\cdot^T$ means $(X)(X)^T$.

II. PROBLEM FORMULATION

The system architecture is shown in Fig. 1. Each node estimates the process state based on its measurements and neighbor information. Moreover, each node is configured with a detector to detect the replay attack that can modify the measurements. The detailed models are introduced as follows.

A. System Model

Consider the discrete linear time-invariant (LTI) process:

$$x(k+1) = Ax(k) + w(k),$$

where $k \in \mathbb{N}$ is the time index, $A \in \mathbb{R}^{n \times n}$ represents the state matrix of the LTI process (1), $x(k) \in \mathbb{R}^n$ denotes the process state, and $w(k)$ is the zero-mean i.i.d. Gaussian noise of the process with covariance $Q > 0$. Assume that $x(0)$ also satisfies the zero-mean Gaussian distribution with covariance $P_0 > 0$. A wireless sensor network consisting of $N$ sensors is applied to measure $x(k)$, and the measurement vector $y_i(k) \in \mathbb{R}^{m_i}$ of the $i$th sensor is governed by the following equation:

$$y_i(k) = C_i x(k) + v_i(k),$$

where $v_i(k)$ is the zero-mean i.i.d. Gaussian measurement noise with covariance $R_i > 0$. Note that $x(0)$, $w(k)$, $v_i(k)$, and $v_i(k)$, $\forall i \neq j$ are independent of each other for all $k$. The topology of the sensor network is a directed graph $G = (V, E)$, where the nodes $V = \{1, 2, \ldots, N\}$ and the edges $E \subset V \times V$ represent the set of sensors and the set of
their communication links, respectively. If there exists an edge \((i, j) \in E\), it means that the \(i\)th sensor can receive data from the \(j\)th sensor. We denote the in-neighbors of the \(i\)th sensor as \(N_i = \{j : (i, j) \in E\}\), whose dimension is \(d_i = \text{card}(N_i)\). The Laplacian matrix utilized to describe \(G\) is \(L = [l_{ij}]\), where

\[
l_{ij} = \begin{cases} 
-1, & \text{if } (i, j) \in E, i \neq j, \\
- \sum_{j \in N_i} l_{ij}, & \text{if } i = j, \\
0, & \text{otherwise.}
\end{cases}
\]

We utilize the distributed consensus filter [34], [35] to jointly estimate \(x(k)\) of the LTI process (1), and define the local state estimate as \(\hat{x}_i(k)\) for the \(i\)th sensor. Note that \(\hat{x}_i(k)\) is also the data transmitted over the sensor network. In other words, the \(i\)th sensor can utilize \(y_i(k)\) and \(\hat{x}_j(k), \forall j \in N_i\) to update \(\hat{x}_i(k+1)\) on the basis of the following equation:

\[
\hat{x}_i(k+1) = A\hat{x}_i(k) + K_i(k)[y_i(k) - C_i\hat{x}_i(k)] \\
- \varepsilon A \sum_{j \in N_i} [\hat{x}_i(k) - \hat{x}_j(k)],
\]

where \(K_i(k)\) and \(\varepsilon \in (0, 1/\max_i(d_i))\) are the Kalman and consensus gains, respectively. The estimation error of the \(i\)th sensor is \(e_i(k) = x(k) - \hat{x}_i(k)\), whose covariance is \(P_i(k) = \mathbb{E}[e_i(k)e_i(k)^T]\). Similarly, \(P_{ij}(k)\) is the cross estimation error covariance between the \(i\)th and \(j\)th sensors. Based on [34], [35], the optimal Kalman gain in (3) is given as follows.

**Lemma 1:** [34], [35] To minimize the estimation error covariance \(P_i(k)\), the optimal Kalman gain of the \(i\)th sensor for the distributed consensus filter (3) is \(K_i(k) = K_i^*(k) = A[P_i(k) + \varepsilon \sum_{j \in N_i} (P_{ij}(k) - P_i(k))]C_i^T C_i P_i(k) C_i^T + R_i)^{-1}\).

By stacking \(e_i(k)\) for \(\forall i \in V\) into a column, the estimation error of the whole sensor network is written as \(e(k) = [e_1(k)^T, ..., e_N(k)^T]^T\) with covariance \(P(k) = \mathbb{E}[e(k)e(k)^T]\), where its ith diagonal block matrix is \(P_i(k)\). Under the following assumptions, the distributed filter (3) is mean-square stable.

**Assumption 1:** The graph \(G\) is strongly connected.

**Assumption 2:** \(\{A, Q^{1/2}\}\) is stabilizable.

**Lemma 2:** [34], [35] Under Assumptions 1 and 2, for an arbitrary but fixed initial nonnegative symmetric \(P(0)\), if the pair \((I_N - \varepsilon L) \otimes A, \text{diag}(C_i)\) is detectable, then \(P(k)\) is bounded for all \(k\) and converges to a unique limit \(\bar{P} > 0\).

**Remark 1:** The distributed consensus filter (3) is derived from [34], [35], and can be regarded as a special case of [34], [35]. Hence, it is straightforward to transform the existing results of [34], [35] into Lemmas 1 and 2. Specifically, it should be noted that for the distributed consensus filter (3), Assumption 3 in [35] is equivalent to the condition that \((I_N - \varepsilon L) \otimes A, \text{diag}(C_i)\) is detectable, and thus we do not need to make the similar assumption in Lemma 2. Moreover, without the consensus term, the distributed consensus filter (3) can be rewritten in the form of general Kalman filter. Accordingly, Lemmas 1 and 2 can also be simplified to the same ones for the general Kalman filter.

**B. Attack Model**

Based on Lemmas 1 and 2, the distributed filter (3) can guarantee its estimation accuracy and mean-square stability in the absence of the attack. However, it may fail to estimate \(x(k)\) due to a malicious third party. In this paper, we consider that the distributed filter (3) is subject to the replay attack. Following the definition in [25], [26], we suppose that the adversary eavesdrops the measurements of several preset sensors during \([k_0, k_0 + \tau - 1]\), and then replays those recorded data to the corresponding sensors during \([k_0 + \tau, k_0 + 2\tau - 1]\). Moreover, due to the limited energy budget or resources especially in the large-scale sensor network, the adversary may only select part of sensors to launch the replay attack. Therefore, a binary variable \(\gamma_i\) is introduced to indicate whether the \(i\)th sensor is attacked, i.e., \(\gamma_i = 0\) represents that the attacker tampers the \(i\)th sensor, while \(\gamma_i = 1\) is on the contrary. In a word, for the \(i\)th sensor, we define its received measurement that may be modified by the attacker during \([k_0 + \tau, k_0 + 2\tau - 1]\) as

\[
y_i^\alpha(k) = \begin{cases} 
y_i(k - \tau), & \text{if } \gamma_i = 0, \\
y_i(k), & \text{otherwise},
\end{cases}
\]

where \(k \in [k_0 + \tau, k_0 + 2\tau - 1]\). Recall that the sensor network is composed of \(N\) sensors, and thus the adversary has \(2^N\) strategies to determine how to corrupt the distributed filter (3). To comprehensively evaluate the security of distributed filter under such a type of attack, we consider that the adversary can freely choose its strategy from the whole attack space. Similar to [25], [26], we assume that the attacker records a sufficient number of measurements during \([k_0, k_0 + \tau - 1]\), i.e., \(\tau\) is sufficiently large. Moreover, based on Lemma 2, we further assume that the distributed filter (3) is already in the steady state before the attack occurs. That is, \(P(k) = \bar{P}\) at the time \(k_0\), which results in a fixed Kalman gain, i.e., \(K_i(k) = \bar{K}_i\).

**Remark 2:** Besides the attacking pattern (4), the adversary may adopt other replay attack strategies. For instance, the start time \(k_0\) may be chosen to be different for each sensor. Moreover, the adversary may record a sequence of measurements of the \(i\)th sensor, but replay those data to the \(j\)th sensor, \(\forall j \neq i\). Thus, it is one of our future directions to study the security of distributed filter (3) in a more complex attack scenario.

**Remark 3:** Following the real case such as the Stuxnet worm [3], the replay attack needs to record a sufficient number of sensor readings [25], [26]. In the following sections, \(\tau\) is defined to be ‘sufficiently large’ such that for a given stable matrix \(M, M^\tau\) converges to the zero matrix. Note that it is a mild assumption since \(M^\tau\) decays exponentially fast. Hence, the attack duration \(\tau\) is not required to be a extremely large.
value. For instance, it is chosen as 80 in our simulations. On the other hand, the limited duration $\tau$ also helps the adversary to save its energy. Thus, the selection of $\tau$ and number of attacked sensors is a trade-off for the replay attack (4).

C. False Data Detector

Each sensor is configured with a residue-based detector to detect potential attacks. Note that many existing detectors such as the $\chi^2$ detector basically depend on the statistical properties of residues. While the K-L divergence is a non-negative measure that quantifies the distance between two probability distributions [36]. Thus, the K-L divergence can be taken to fundamentally measure the performance of the residue-based detectors, and its definition is given as follows.

Definition 1: Let $h$ and $g$ be two random sequences, whose joint probability density functions are $f_h(x)$ and $f_g(x)$, respectively. The K-L divergence between $h$ and $g$ is

$$D(h\|g) = \int_{x|f_h(x) > 0} \log \frac{f_h(x)}{f_g(x)} f_h(x) dx.$$  

For the $i$th sensor, its residue is $z_i(k) = y_i(k) - C_i \hat{x}_i(k)$ when there is no attack, and the attacker induced residue is $z^*_i(k) = y_i^*(k) - C_i \hat{x}^*_i(k)$, where $\hat{x}_i(k)$ is the local state estimate under attack. Thus, we can take the K-L divergence $D(z_i^*(k)\|z_i(k))$ to measure the attack stealthiness. A larger $D(z_i^*(k)\|z_i(k))$ means that the attack is easier to be detected.

D. Problem of interest

We are interested in analyzing the estimation accuracy and mean-square stability of the distributed filter (3) under the replay attack (4). Besides, we utilize the K-L divergence to analyze the attack stealthiness. According to whether the state matrix $A$ satisfies $|\lambda_{\text{max}}(A)| < 1$ or $|\lambda_{\text{max}}(A)| > 1$, the LTI process (1) is called as a stable or unstable system, respectively. The main results for the two different systems will be introduced in the following two sections.

III. SECURITY ANALYSIS OF STABLE SYSTEMS

Under the attack (4), this section first gives the closed-form expression of estimation error covariance for the distributed filter (3). Given that $A$ is stable, we can derive a constant upper bound of its trace, which demonstrates the boundness of estimation performance degradation. Then, we prove the mean-square stability of the estimator. Finally, we prove that the corresponding K-L divergence is bounded as well, and derive its steady state value to quantify the attack stealthiness.

A. Attack effect

First, we define some quantities that will be used in the remainder of the paper. By stacking all the respective elements into a column, we can obtain the vector forms of $\hat{x}_i(k)$, $y_i(k)$, $v_i(k)$, and $z_i(k)$, which are defined as $\hat{x}(k)$, $y(k)$, $v(k)$, and $z(k)$, respectively. The covariance of $z_i(k)$ is defined as $\Sigma_i(k) = E[z_i(k)z_i^T(k)]$. According to the properties of $v_i(k)$, the covariance of $v(k)$ is $R = E[v(k)v(k)^T] = \text{diag}(R_i)$.

Furthermore, we add the superscript ‘a’ to the above quantities to emphasize that the system is under attack, e.g., $\hat{x}^a(k)$. During $[k_0 + \tau, k_0 + 2\tau - 1]$, the iterative formula of the distributed filter (3) under the replay attack (4) is rewritten as

$$\hat{x}^a_i(k+1) = A\hat{x}^a_i(k) + \hat{K}_i[y^a_i(k) - C_i\hat{x}^a_i(k)] - \varepsilon A \sum_{j \in N_i}[\hat{x}^a_j(k) - \hat{x}^a_j(k)],$$  

where its initial state is $\hat{x}^a_i(k_0 + \tau) = \hat{x}_i(k_0 + \tau)$, $y^a_i(k) = (1 - \gamma_i)y_i(k - \tau) + \gamma_i y_i(k)$ according to (4), and $\hat{K}_i(k) = \hat{K}_1$ due to the steady-state assumption. Based on (6), the state estimate of the whole sensor network under attack is as follows:

$$\hat{x}^a_i(k+1) = \Gamma \hat{x}^a_i(k) + \text{diag}(\hat{K}_i)y^a_i(k),$$  

where $\Gamma = (I_N - \varepsilon I) \otimes A - \text{diag}(\hat{K}_iC_i)$, and $y^a_i(k) = \text{diag}((1 - \gamma_i)I_m)y(k - \tau) + \text{diag}(\gamma_iI_m)y(k)$. Then, from (1) and (7), the error dynamic that describes the global estimation accuracy, i.e., $e^a(k) = 1_N \otimes (x(k) - \hat{x}^a(k))$, is given by

$$e^a(k+1) = \Gamma e^a(k) + 1_N \otimes w(k) - \text{diag}(\gamma_i\hat{K}_i)v(k) - \text{diag}((1 - \gamma_i)\hat{K}_i)v(k - \tau) + \text{diag}((1 - \gamma_i)\hat{K}_iC_i)1_N \otimes [x(k) - x(k - \tau)].$$  

Note that in the absence of the attack, i.e., $\gamma_i = 1, \forall i \in V$, (7) and (8) can be simplified into the form of

$$\hat{x}(k+1) = \Gamma \hat{x}(k) + \text{diag}(\hat{K}_i)y(k),$$  

$$e(k+1) = \Gamma e(k) + 1_N \otimes w(k) - \text{diag}(\hat{K}_i)v(k).$$  

The following lemma illustrates the relation between $\Gamma$ and the mean-square stability of the normal estimation system (10).

Lemma 3: For the distributed estimator in the absence of the attack, the estimation error (10) can be mean-square stable, if and only if $\Gamma$ is stable, i.e., $|\lambda_{\text{max}}(\Gamma)| < 1$.

Proof: See Appendix.

Lemma 3 implies that even if it is only to ensure the stability of the normal estimation system (10), the Kalman gain $\hat{K}_i$ should be designed to stabilize $\Gamma$. Then, according to [37], it is worth mentioning that due to the stability of $\Gamma$ and $A$, there exist constants $\eta_1$ and $\eta_2$ in the scope of $[0, 1]$ such that $\|\Gamma\|_2 \leq \rho_V(\eta_1)$ and $\|A\|_2 \leq \rho_A(\eta_2)$, where $\rho_V$ and $\rho_A$ are positive constants. Following this, we can directly obtain the boundness of estimation error covariance with respect to (8).

Theorem 1: When the LTI process (1) is stable and the distributed consensus filter (3) is subject to any replay attack (4), the trace of the estimation error covariance $\text{tr}(P^a(k))$ is upper bounded by

$$\omega \triangleq \omega_1 \text{tr}(Q) + \omega_2 \text{tr}(\Pi_0) + \omega_3 \text{tr}(\text{diag}(\hat{K}_i)R_i(k)^T),$$  

where

$$\omega_1 \triangleq \frac{3N(\rho_A)^2\sigma_3}{(1 - (\eta_1)^2)(\log(N))^2} + \frac{4N(1 + \hat{\sigma}_2^2)\sigma_3}{(1 - \rho_V^2)^3} + \frac{2N(\rho_A)^2}{(1 - (\eta_2)^2)^2},$$  

$$\omega_2 \triangleq N(\rho_A)^2(\eta_2)^2\sigma_3^2, \quad \omega_3 \triangleq (\rho_V)^2(\sigma_3^2),$$  

and $\hat{\sigma}_3$ is the maximum of $\sigma_1$ and $\sigma_2$. $\omega_1$ and $\omega_2$ are positive constants. Following this, we can always find a constant upper bound of $\text{tr}(P^a(k))$ from (11). It implies
that when $A$ is stable, the boundness of $\text{tr}(P^a(k))$ can be
guaranteed under an arbitrary replay attack (4).

Remark 5: Based on (28), the closed-form expression of
the estimation error covariance is $P^a(k + 1) = E[(\bar{\Xi}(\cdot))^T] + \sum_{s=0}^{\infty} [\Gamma^s \text{diag}(K_i)] R_i^{(s)}$, where $E[(\bar{\Xi}(\cdot))^T]$ is given by (30).

Note that $P^a(k + 1)$ has a very complex structure due to the
coupling relation between $y(k)$ during $[k_0, k_0 + \tau - 1]$ and
$y^s(k)$ during $[k_0 + \tau, k_0 + 2\tau - 1]$. Hence, we utilize the tools
such as norm inequality to derive the upper bound in (11) at the
cost of some conservatism.

Theorem 1 means that $P^a(k)$ is not beyond a certain limit
during $[k_0 + \tau, k_0 + 2\tau - 1]$. With the increase of time $k$, we
further prove that $P^a(k)$ can converge to a fixed matrix $P^a$, and
thus the filtered output can re-enter the steady state.

Theorem 2: When the LTI process (1) is stable and the
distributed consensus filter (3) is subject to any replay attack
(4), the estimation error covariance $P^a(k)$ converges to $P^a = 1_N 1_N^T \otimes Q + \Delta_1$, where $\Delta_1$ can be solved by the following
Lyapunov equation:

$$\Delta_1 - \Gamma \Delta_1 \Gamma^T = \text{diag}(K_i) R_i^{(s)} + [\Gamma + \bar{\Xi}] 1_N 1_N^T \otimes Q + \bar{\Xi} + 2 \bar{\Xi} 1_N 1_N^T \otimes [\varphi_{AQ} - Q] \bar{\Xi}^T,$$

where $\bar{\Xi} \triangleq \text{diag}((1 - \gamma_i) K_i C_i)$, $\varphi_{AQ}$ satisfies $\varphi_{AQ} - A \varphi_{AQ} A^T = Q$, and $\Delta_2$ can be solved by the following
Lyapunov equation:

$$\Delta_2 - (I_N \otimes A) \Delta_2 \Gamma^T = (I_N \otimes A) 1_N 1_N^T \otimes Q [\Gamma + 2T] \Gamma^T + 21_N 1_N^T \otimes [\varphi_{AQ} - Q] \bar{\Xi}^T.$$

Proof: See Appendix.

Remark 6: Based on (34), (35), the steady-state estimation
covariance error covariance in the absence of the attack is $P \triangleq \sum_{s=0}^{\infty} [\Gamma^s \text{diag}(K_i)] R_i^{(s)} + [\Gamma + \bar{\Xi}] 1_N 1_N^T \otimes Q + \bar{\Xi} + 2 \bar{\Xi} 1_N 1_N^T \otimes [\varphi_{AQ} - Q] \bar{\Xi}^T$. By comparing $P$ and $P^a$, it is intuitive to see the impact of replay attack (4) on the distributed filter (3) in a stable LTI process (1). Note that when $\gamma_i = 1, \forall i \in V$, $\bar{\Xi}$ is a zero matrix such that $P$ can be simplified into $P$. Hence, $P$ is a special case of $P^a$.

Remark 7: Theorem 2 implies that once the attack occurs, the
estimation process can successively enter the transient
and steady states. Thus, we utilize Theorems 1 and 2 to
comprehensively analyze the estimation performance in the
two states, respectively. Note that in the transient state, its
duration is short in general, and its estimation performance
degradation is also limited based on Theorem 1. Thus, it is
more urgent to propose some methods to enhance the
distributed estimation in the steady state. Then, $P^a$ in Theorem
2 can be taken as a benchmark to evaluate their effectiveness.

B. Attack Stealthiness

Recall that the attack stealthiness is evaluated based on the
K-L divergence between $z^o(k)$ and $z_i(k)$. One can verify that
$z^o(k)$ is still a linear combination of $x(0)$, $w(k)$ and $v(k)$ such
that $z^o(k)$ still has the Gaussian property. Hence, according to (38), the right hand side of (5) is equivalent to

$$\frac{1}{2} \text{tr}(\Sigma^o_i(k) \Sigma_i^{-1}(k)) - \frac{m_i}{2} + \frac{1}{2} \log \frac{\det(\Sigma_i(k))}{\det(\Sigma^o_i(k))},$$

which describes $D(z^o_i(k) || z_i(k))$ on the basis of the difference between $\Sigma^o_i(k)$ and $\Sigma_i(k)$. Under an arbitrary attack (4), we can prove that the K-L divergence (14) is also bounded for each sensor, since it does not exceed the derived constant.

Theorem 3: When the LTI process (1) is stable and the
distributed consensus filter (3) is subject to any replay attack
(4), the K-L divergence $D(z^o_i(k) || z_i(k))$ of the $i$th sensor is upper bounded by

$$\begin{align*}
\frac{\omega_4}{2} + \omega_5 \omega_6 - \frac{m_i}{2} + \frac{m_i}{2} & \log(\lambda_{\min}(\Sigma_i(k) - \frac{1}{2} R_i(k))),
\end{align*}$$

where $\omega_4 \triangleq \text{tr}(\Sigma_i(k) - R_i(k))$, $\omega_5 \triangleq \lambda_{\max}(C_i^T \Sigma_i(k)^{-1} C_i)$,
$\omega_6 \triangleq 2 \text{tr}((\varphi_{AQ}) + N(\rho_A)^2 \text{tr}(\Pi_i) + \omega_i \Sigma_i(k) = C_i P_i (C_i^T + R_i)$, $P_i$ is the $i$th block diagonal matrix of $P$.

Proof: See Appendix.

Similar to Theorem 2, we can prove that as time $k$ increases,
the K-L divergence (14) can enter the steady state as well.

Theorem 4: When the LTI process (1) is stable and the
distributed consensus filter (3) is subject to any replay attack
(4), the K-L divergence $D(z^o_i(k) || z_i(k))$ of the $i$th sensor can enter the steady state, since $\Sigma^o_i(k)$ in (14) converges to a
unique limit $\Sigma^o_i(k) = C_i (\Delta_1 + Q) (C_i^T + R_i)$, where $\Delta_1$ is the $i$th block diagonal matrix of $\Delta_1$, and $\Delta_1$ can be solved by the following
Lyapunov equation:

$$\Delta_1 - \Gamma \Delta_1 \Gamma^T = \text{diag}(K_i) R_i^{(s)} + [\Gamma + \bar{\Xi}] 1_N 1_N^T \otimes Q + \bar{\Xi} + 2 \bar{\Xi} 1_N 1_N^T \otimes [\varphi_{AQ} - Q] \bar{\Xi}^T,$$

where $\bar{\Xi} \triangleq [(\gamma_i I_n) - \bar{\Xi}] 1_N 1_N^T \otimes Q + \bar{\Xi} 1_N 1_N^T \otimes [\varphi_{AQ} - Q] \bar{\Xi}^T$.

Proof: The proof is similar to that of Theorem 2, and
thus is omitted for simplicity.

Remark 8: Based on the definition of $z^o_i(k)$, the attacked
measurement $y^a_i(k)$ on the $i$th node can directly affect the
local residue $z^o_i(k)$ at the time step $k$. Moreover, due to the
information exchange over the sensor network and the data
fusion process in (6), $y^a_i(k)$ can indirectly affect $z^o_j(k + 2)$ of
the neighbor node $j \in N_i$ at the time step $k + 2$. It indicates that the K-L divergence of each sensor is related to the attack on the whole sensor network. Besides, $\Sigma^o_i(k)$ depends on $x(k - \tau) - x(k)$ according to (37). Thus, if $x(k_0 + \tau)$ deviates from
functions with maximal right eigenvector of $A$ to detect the replay attack or mitigate its impact on distributed future work is to propose some methods that can effectively error and maintains a low K-L divergence. On the other hand, when the state matrix $A$ can see that when the state matrix $A$ increases, while all the steady-state $A$ are detectable. Since the replay attack may enhance its estimation against the attack, by utilizing the game theory to estimate the LTI process (1) is unstable and the distributed consensus filter (3) is subject to any replay attack (4), the K-L divergence $D(\pi_0(k))$ is lower bounded by

$$
\varpi_5(k) / (\lambda_{\min}(\Pi_0) f_A^1(k) + \lambda_{\min}(Q) f^2_A(k)) - m_s^i / 2,
$$

where $\varpi_5(k) \triangleq \left(1 - \frac{1}{2e} \right) \| \Phi_5(k-\kappa_0-\tau) \|_2^2$ and $\Phi_5(k - k_0 - \tau) \triangleq C_i[\theta_i \otimes I_n] \Phi_4(k - k_0 - \tau) - (1 - \gamma_i) \lambda_{\max}(A) C_i X_A.$

Proof: See Appendix.

Remark 10: In the absence of the attack, i.e., $\gamma_i = 1, \forall i \in V$, both $\varpi_4(k)$ and $\varpi_5(k)$ are equal to zero. Hence, the results in (18) and (19) are consistent with the fact that neither the estimation error nor the K-L divergence diverges in this case. While in the presence of the attack, $f_A^1(k)$ and $f_A^2(k)$ have diverged to large values even if at the beginning time $k_0 + \tau$. Hence, even if the attack may damage the estimator in an unstable LTI process (1), the $i$th sensor is able to quickly detect the attack and identify $k_0 + \tau$ when $\varpi_5(k) \neq 0$.

Remark 11: There is no doubt that both $P^\omega(k)$ and $D(\pi_0(k))$ are related to the Laplacian matrix $L$. Our results also depend on $L$ due to their relationship to the matrix $\Gamma$ that contains $L$. For instance, the lower bound in Theorem 5 consists of $\varpi_4(k)$, whose value is based on $\Gamma$.

To ensure that the system does not have any safety loophole, we should guarantee that $\varpi_5(k) \neq 0$ for an arbitrary attack (4). However, the attack has $2^N$ strategies, and the computation complexity to verify $\varpi_5(k)$ is $O(2^N)$. Thus, the traversing method is not suitable for a large-scale network. We provide a judging criteria as follows, whose complexity is $O(N)$.

Theorem 6: When the LTI process (1) is unstable, if

$$
[C_i K_i - \lambda_{\max}(A) I_m] C_i X_A \neq 0, \forall i \in V
$$

is satisfied, then the residue-based detector is guaranteed to detect all the attacks in (4).

Proof: As stated in Remark 10, it is sufficient to detect the attack, if $\varpi_5(k_0 + \tau) \neq 0$. Note that $\Phi_5(0) = (1 - \gamma_i) [C_i K_i - \lambda_{\max}(A) I_m] C_i X_A$. Hence, when the $i$th sensor is attacked, i.e., $\gamma_i = 0$, the attack can be detected if (20) is satisfied.

Remark 12: If $A$ has multiple maximum eigenvectors, we only need to ensure that one of them satisfies (20). Moreover, Theorem 6 provides a method to resist the replay attack (4) by designing system parameters. Specifically, we can first design the measurement matrix $C_i$ such that $C_i X_A \neq 0, \forall i \in V$. Then, we need to check whether the optimal steady-state Kalman gain $K_1^*$ in Lemma 1 satisfies (20). If not, we can add a perturbation matrix $\Delta K$ to $K_1^*$, which slightly sacrifices the estimation performance but guarantees the security.

Remark 13: Compared with Theorem 3, the selection of $C_i$ in Theorem 6 has more effect on the judgment of whether the replay attack can be detected, since it helps to determine whether the K-L divergence diverges in that case. Therefore, different form Theorem 3, $C_i$ is required to further satisfy the inequality constraint (20) in Theorem 6. On the other hand, (20) is a sufficient condition for the attack detection in unstable.
systems, and thus we have no requirement on the Laplacian matrix $L$ in Theorem 6.

B. Attack scenario that $x(k_0) = x(k_0 + \tau)$

As stated in [26], to remain undetected at $k_0 + \tau$, the replay attack may wait until the current measurement is close to the initial part of the recording such that $x(k_0 + \tau) = x(k_0)$. Hence, we further analyze the security of distributed filter in such an attack scenario. In the following theorem, we prove that if the system is designed to satisfy (21), the necessary and sufficient condition for the adversary to avoid being detected is to tamper the whole sensor network.

**Theorem 7:** When the LTI process (1) is unstable and the time $k_0 + \tau$ is chosen such that $x(k_0 + \tau) = x(k_0)$, the replay attack (4) is strictly stealthy, i.e., $D(z^A_t(k)\|z_t(k)) = 0$, if its strategy is $\gamma_i = 0, \forall i \in V$. While for any other attack strategy in (4), if the following condition holds

$$\text{diag}(C_t) \hat{\Upsilon} \neq 0,$$

where $\hat{\Upsilon} = (I_{Nn} - \frac{\tau}{\max(A)})^{-1}[\varepsilon L \text{diag}(1 - \gamma_i)1_N] \otimes X_A$, then the attack is detectable.

**Proof:** See Appendix. 

To effectively limit the attack space, we follow (21) to design system parameters including $C_t, K_t, \eta, L$. Similarly, the complexity to calculate (21) is $O(2^N)$. In two special cases of Corollary 1, we prove that (21) can be simplified and thus the complexity can be reduced to $O(N)$.

**Corollary 1:** When $A$ has a single maximum eigenvalue or the sensor nodes are homogeneous in the form of $C_t = \beta_i C, \forall i \in V$, where $\beta_i$ is a nonzero constant and $C$ is a constant matrix, if the following condition is satisfied

$$C_t X_A \neq 0, \forall i \in V,$$

for the attack (4) which selects $k_0 + \tau$ such that $x(k_0 + \tau) = x(k_0)$, all strategies except for $\gamma_i = 0, \forall i \in V$ are detectable.

**Proof:** See Appendix.

Then, we further summarize the estimation error covariance under the replay attack (4) with its strategy $\gamma_i = 0, \forall i \in V$.

**Corollary 2:** When the LTI process (1) is unstable, the replay attack (4) adopts its strategy as $\gamma_i = 0, \forall i \in V$ and chooses the time $k_0 + \tau$ such that $x(k_0 + \tau) = x(k_0)$, the complexity to calculate (21) is $O(2^N)$. Here, $\text{diag}(C_t) \hat{\Upsilon} \neq 0, \forall i \in V$.

**Proof:** See Appendix.

**Remark 14:** Corollary 2 implies that when $\gamma_i = 0, \forall i \in V$, the replay attack (4) has effect on delaying the estimation of distributed filter. By comparing $f_3^2(k)$ and $f_2^3(k)$, $\text{tr}(P^3(t))$ in (23) diverges slower than the one in (18), which is consistent with our intuitive understanding since the modified data is close to the true measurement at the time $k_0 + \tau$.

**Remark 15:** Intuitively, it is the inconsistency of sensor inputs that makes the replay attacks except for $\gamma_i = 0, \forall i \in V$ detectable. While in the case of $\gamma_i = 0, \forall i \in V$, we can see that no matter how to design system parameters, (21) does not hold since $L1_N = 0$ for a Laplacian matrix $L$. Nevertheless, due to the energy budget of the attack, it is a high requirement to tamper all the nodes in a large-scale sensor network. Moreover, when the LTI process (1) is unstable, $x(k_0 + \tau)$ has a low probability to approach $x(k_0)$. Hence, it is also a strong constraint to find $k_0 + \tau$ such that $x(k_0 + \tau) = x(k_0)$.

Finally, safety sensors with hardware protection can avoid the intrusion of attacks [41], and thus we can easily fill up such a loophole by allocating only one safety sensor.

V. SIMULATION EXAMPLES

In this section, some numerical examples are given to verify the analytic results. Specifically, a 3-order LTI process (1) is monitored by a sensor network with $N = 30$ nodes. The system parameters are as follows:

$$A_1 = \begin{pmatrix} 0.5 & 0 & -0.2 \\ 0.1 & -0.1 & 0 \\ -0.1 & 0 & -0.3 \end{pmatrix}, Q = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 4 \end{pmatrix},$$

$$C_t = \beta_i \begin{pmatrix} 1 \\ -1 \\ 0 \\ 2 \end{pmatrix}, R_i = \nu_i \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix},$$

where $\beta_i, \nu_i \in (0, 1), \forall i \in V$. Besides, we set $\Pi_0 = I_3$ and $\varepsilon = 0.05$. For comparison, we further consider two other LTI processes (1), whose state matrices are $A_2$ and $A_3$, respectively. Note that the only difference between $A_1, A_2$ and $A_3$ is that their 1st diagonal elements are 0.5, 0.9 and 1.2 respectively. Clearly, since $|\max(A_1)| = 0.5137, |\max(A_2)| = 0.9196$ and $|\max(A_3)| = 1.2152$, $A_1$ and $A_2$ are stable, while $A_3$ is not. Via 10000 Monte Carlo simulations, all of the numerical examples are obtained during the period [0, 200], while the adversary records the measurements during [21, 100] and utilizes the replay attack to tamper the distributed filter during [101, 180]. Besides, the replay attack has two alternative strategies $\Omega_1$ and $\Omega_2$ to select attacked sensors. Specifically, $\Omega_1$ means that the 3rd, 10th, 15th, and 28th sensors are under...
Furthermore, the simulated value of\( \text{tr}(P_1(\delta)) \) in Case 2. It implies that as the number of attacked
of estimation error in Case 3 is significantly larger than the
3, which have the same
\( \Omega \)
vergence of replay attack in Cases 1-3 of Fig. 2. By comparing
estimation performance of distributed filter and the K-L diver-
ance calculated as
\( \sum_{i=1}^{N} D(z_i^a(k)||z_i(k)) \) of each case diverges over time. Moreover,
by comparing Cases 4 and 5, we can see that the divergence
speed of the K-L divergence in Case 4 is faster than the one
in Case 5. It may be because that through the information
exchange on the sensor network, the 3rd sensor in Case 5
is indirectly affected by attacks on other sensors, while it is
directly attacked by the adversary in Case 4.

We further compare the replay attack (4) with the false
data injection (FDI) attack studied in [12]. Specifically, in
Fig. 4, we take the replay attack in Case 2 of Fig. 2 as a benchmark to compare the FDI attacks in Cases 6 and 7.
That is, in all the three cases, the same sensor nodes in \( \Omega_1 \)
are attacked during [101, 180] for the stable system \( A_2 \).
Note that the FDI attack in [12] injects Gaussian random noise to the
measurements, and can be modeled as
\( g_i^a(k) = y_i(k) + a_i(k) \),
where \( a_i(k) \) is the zero-mean i.i.d. Gaussian noise with its
covariance \( \delta_i^a > 0 \). Therefore, we define the covariances of
false data in Cases 6 and 7 as \( \delta_i^6 \) and \( \delta_i^7 \), respectively, where
\( \delta_i^6 = \iota_i \delta_i^6, \quad \iota_i > 1 \), both \( \iota_i \) and \( \delta_i^6 \) are randomly generated.
In Fig. 4, the simulated value of \( \text{tr}(P^a(k)) \) in Case 2 is
the largest, while the corresponding \( D(z_i^a(k)||z_i(k)) \) is the
smallest. Clearly, the choice of \( \iota_i \) is a tradeoff between the
attack effect and stealthiness for the FDI attack. Hence, with
the increase of \( \iota_i \), it is possible that the FDI attack in Case 7
can cause larger estimation error than the replay attack in Case 2.
However, the corresponding \( D(z_i^a(k)||z_i(k)) \) in Case 7
will increase, and the stealthiness of the FDI attack will become
worse meanwhile. By comprehensively evaluating the attack
effect and stealthiness, we can conclude that for the distributed
estimation, the replay attack (4) performs better than the FDI
attack [12] in some cases.

VI. CONCLUSION
In this paper, we have analyzed the impact of the replay
attack on the distributed filter and evaluated its stealthiness
with a measure of the K-L divergence. For stable and unstable
systems, their theoretical results are exactly the opposite. Specifically, despite having a limited effect on degrading the estimation performance, the replay attack remains $\epsilon$-stealthy when the system is stable. For the unstable system, even if the estimation error may diverge under attack, we can ensure that all the attack strategies are detectable if the system parameters satisfy the proposed criteria. When the starting time is subtly chosen, we have proven that the replay attack falsifying all sensors is strictly stealthy. However, by properly designing the system parameters, all the other strategies are detectable, and thus we can effectively limit the attack behaviors. One of our future works is to analyze the optimal replay attack, which has sufficient system knowledge to maximize the estimation error. Another future work is to analyze the security of distributed estimation under the replay attack, when $|\lambda_{\text{max}}(A)| = 1$.

APPENDIX

A. Proof of Lemma 3

Since $w(k)$ is independent of $v(k)$, $P(k+1) = \Gamma P(k)\Gamma^T + \text{diag}(\bar{K}_i) R_i \tau T + I_N T_N^T \otimes Q$. Based on Lemma 2, $P(k+1)$ will converge to $\bar{P}$, and thus we have

$$\bar{P} = \Gamma \bar{P} \Gamma^T + \text{diag}(\bar{K}_i) R_i \tau T + I_N T_N^T \otimes Q. \quad (24)$$

Note that (24) is solvable if $\Gamma$ is stable. Hence, it is a sufficient condition for the stability of (10). Define $X_\Gamma$ as the maximal left eigenvector of $\Gamma$ such that $X_\Gamma \Gamma = \lambda_{\text{max}}(\Gamma) X_\Gamma$. Since $(X_\Gamma \Gamma) X_\Gamma \geq 0$, we have $(X_\Gamma \Gamma) X_\Gamma \leq \mu_\Gamma I_N$, where $\mu_\Gamma = \|X_\Gamma\|_2$. By recursively iterating (24) through $l$ steps, we have

$$\text{tr}(P) = \text{tr}[(\Gamma^l \bar{P} \Gamma^T)^T I_{Nn}(\cdot)] + \sum_{s=0}^{l-1} \text{tr}[(\Gamma^s \text{diag}(\bar{K}_i) R_i \tau T)^T I_{Nn}(\cdot)]$$

$$\geq \underbrace{\mu_1 \|X_\Gamma\|^2}_w + \underbrace{\mu_2 \lambda_{\text{max}}(\Gamma)^2}_{\text{var}}, \quad (25)$$

where $\mu_1 \equiv \mu_\Gamma^2 (\|X_\Gamma I_N \otimes \bar{Q}_i \|^2 + \|X_\Gamma \text{diag}(\bar{K}_i) R_i \|^2)$, and $\mu_2 \equiv \mu_\Gamma^2 (\|X_\Gamma \bar{P} \|^2_2)$.

B. Proof of Theorem 1

Based on the iteration of (7) and (9), we have

$$\dot{x}(k+1) - \dot{x}_0(k+1) \equiv (a) \equiv \Xi + \sum_{s=0}^{k-\ell} \Phi_1(s) [v(k-s) - v(k-\tau - s)], \quad (26)$$

where (a) follows from $\dot{x}(0) = x_0(k+\tau) \equiv \Xi \equiv \sum_{s=0}^{k-\ell} \Phi_2(s) I_N \otimes [x(k-s) - x(k-\tau - s)]$, $\Phi_2(s) \equiv \Phi_1(s) \text{diag}(C_i)$, and $\Phi_1(s) \equiv \Gamma^s \text{diag}((1-\gamma_i) K_i)$. Then, based on (10) and (26), we have

$$e^s(k+1) = e(k+1) + \underbrace{\Gamma^s \sum_{s=0}^{\tau-1} v(k-s) - \sum_{s=0}^{\tau-1} \Gamma^s}_{(b)}$$

$$\text{diag}(\bar{K}_i) v(k-s) - \sum_{s=0}^{k-\ell} \Phi_1(s) v(k-s), \quad (27)$$

where $k \not\equiv \ell$, $\Xi \equiv \Xi + \sum_{s=0}^{\tau-1} \Gamma^s I_N \otimes w(k-s)$, and the first term in (b) converges to zero vector since $\Gamma$ is stable and $\tau$ is sufficient large. Based on (27), we have

$$P^a(k+1) \equiv \mathbb{E}[e^a(k+1) (k)^T]$$

$$\geq \mathbb{E}[(\Xi)(\cdot)^T] + \sum_{s=0}^{k-\ell} \text{diag}(\bar{K}_i) R_i \tau T \mathbb{E}[(\cdot)(\cdot)^T] + \sum_{s=0}^{k-\ell} \Phi_1(s) R_i \tau T + \sum_{s=0}^{\tau-1} \sum_{s=0}^{k-\ell} \text{diag}(\bar{K}_i) R_i \tau T, \quad (28)$$

where (c) is based on the independence among $x(k), v(k)$ and $v(l), \forall k \not\equiv l$. Note that $A^r$ converges to the zero matrix since $A$ is stable and $\tau$ is sufficient large. Hence, we can follow the evolution of $x(k)$ in (1) to rewrite $\Xi$ as

$$\Xi = I_N \mathbb{E} \otimes w(k) + \sum_{s=0}^{k-\ell} \text{diag}(\bar{K}_i) R_i \tau T \mathbb{E}[(\cdot)(\cdot)^T]$$

$$+ \sum_{s=0}^{\tau-1} \sum_{s=0}^{k-\ell} \text{diag}(\bar{K}_i) R_i \tau T \mathbb{E}[(\cdot)(\cdot)^T] + \sum_{s=0}^{\tau-1} \sum_{s=0}^{k-\ell} \text{diag}(\bar{K}_i) R_i \tau T$$

where $\Phi_3(s) \equiv \sum_{t=0}^{k-\ell} \Phi_2(t) I_N \otimes A^{s-t}$. Since each term in (29) is independent with each other, $\mathbb{E}[(\Xi)(\cdot)^T]$ is given by

$$\mathbb{E}[(\Xi)(\cdot)^T] = \Phi_2(k-\ell) I_N \otimes A^\ell \mathbb{E} I_N T_N^T \mathbb{E} + \sum_{s=0}^{k-\ell} \mathbb{E} I_N T_N^T \mathbb{E} + \sum_{s=0}^{\tau-1} \sum_{s=0}^{k-\ell} \Phi_1(s) \Phi_2(s) \Phi_3(s) \Phi_2(s) \mathbb{E} I_N T_N^T \mathbb{E} + \sum_{s=0}^{\tau-1} \sum_{s=0}^{k-\ell} \text{diag}(\bar{K}_i) R_i \tau T$$

$$+ \mathbb{E} I_N T_N^T \mathbb{E} + \sum_{s=0}^{\tau-1} \sum_{s=0}^{k-\ell} \text{diag}(\bar{K}_i) R_i \tau T.$$
Notice that for a matrix \( A \geq 0 \), we have \( \text{tr}(BAB^T) \leq \lambda_{\text{max}}(B^TB)\text{tr}(A) \), where \( B \) is any matrix with appropriate dimensions. Thus, an upper bound of \( \text{tr}([\bar{\Xi}]^{(\cdot)}T) \) is
\[
\text{tr}([\bar{\Xi}]^{(\cdot)}T) \leq \text{tr}(1_{N_T^0} \otimes \Pi_0) \left\| \Phi_3(k-\tilde{k})II_N \otimes A^{k_0} \right\|_2^2 + \text{tr}(1_{N_T^0} \otimes Q) \left\{ 1 + \sum_{s=0}^{k-k-1} \left\| \Gamma^{s+1} + \Phi_3(s) \right\|_2^2 + \sum_{s=0}^{k-\tilde{k}+s+1} \left\| \Gamma^{k-\tilde{k}+s+1} + \Phi_3(s) \right\|_2^2 \right\}
\]
(where \( \equiv \eta \)).

From (33), we have \( P^a(k+1) \geq P^a(k) \). Hence, the estimation system (8) is stable, if and only if \( I_1 \equiv \sum_{s=0}^{\infty} (\Theta_1(s) + \Theta_2(s)) \) converges to a fixed matrix such that the steady-state of \( P^a(k) \) is \( P^a \equiv 1_{N_T}^0 \otimes Q + \Delta_1 \). Note that
\[
\Phi_3(s+1) \equiv \sum_{t=0}^{s+1} \Gamma^t \text{diag}((1-\gamma_1)\bar{K}_CC_i)I_N \otimes A^{s+1-t} + \gamma I_N \otimes A^{s+1} - \gamma \bar{\Theta}_s(s) + \sum_{s=0}^{\infty} (\Theta_3(s) - \Theta_3(s)) T^{s+1-t} + \gamma \bar{\Theta}_s(s),
\]
where \( \gamma \equiv \text{diag}((1-\gamma_1)\bar{K}_CC_i) \). Thus, if \( \Delta_1 \) converges, it must satisfy the following Lyapunov equation:
\[
\Delta_1 \Gamma^T \equiv \sum_{s=0}^{\infty} \gamma \left( \Theta_1(s) + \Theta_2(s) \right) \Gamma^T
\]
\[
\Delta_1 \equiv \Theta_1(0) - \Theta_2(0) + 2\gamma I_N \otimes A \cdot [\varphi AQ - Q] Y^T
\]
\[
- \gamma \sum_{s=1}^{\infty} \Theta_3(s) \equiv \sum_{s=1}^{\infty} (\Theta_3(s) - \Theta_3(s)) T^{s+1-t} + \gamma \bar{\Theta}_s(s),
\]
where \( \Theta_1(s) \equiv (I_N \otimes A^s) !_{N_T} \otimes Q [\Gamma^{s+1} + 2\Phi_3(s)] T^s \), and \( \varphi AQ \equiv \sum_{s=0}^{\infty} A^s Q \). The solution of \( \varphi AQ - A\varphi AQ A T = Q \) since \( A \) is stable. Suppose that the last two terms of (36) converges to \( \gamma \Delta_2 \) and \( \Delta_2 ^T \gamma ^T \) respectively. Similarly, we have
\[
\Delta_2 \equiv \sum_{s=1}^{\infty} (I_N \otimes A^s) !_{N_T} \otimes Q [\Gamma^{s+1} + 2\Phi_3(s)] T^s
\]
\[
= (I_N \otimes A) \Delta_2 \Gamma^T + (I_N \otimes A) \Theta_2(0) \Gamma^T + 2 \gamma I_N \otimes [\varphi AQ - Q] Y^T.
\]
Note that for a equation in the form of \( X - AXB = C \), where \( A, B \in \mathbb{R} \times n \times n \) are stable, we can convert it into the vector form, i.e., \( vec(X - AXB) = (I_n \otimes -B^T \otimes A) vec(X) = vec(C) \). One can derive that \( I_{n_2} - B^T \otimes A \) is invertible, since it does not have zero eigenvalue. It indicates that the above equation is solvable with the solution \( vec(X) = (I_{n_2} - B^T \otimes A)^{-1} vec(C) \). Similarly, based on the stability of \( A \) and \( \Gamma \), we can prove that both (35) and (36) are solvable and have unique solutions, from which we verify the stability of \( P^a(k) \).
D. Proof of Theorem 3

Based on the definition of $z_i(k)$ and steady state assumption of $P(k)$, we have $\Sigma_i(k) = (C_i) P_i(\cdot)^T + R_i$, where $P_i$ is the $i$th block diagonal matrix of $P$. From (27), we have $E[e^{a}(k+1) + (x(k) - x(k))\Sigma_i(k) + (1 - \gamma_i) I_m, \nu(k+1)] = 0,$ which converges to zero matrix such that $\nu(k+1)$ is independent of $e^a(k+1)$. Hence, $\Sigma_i(k) \triangleq E[\xi^T(k)\xi(k)]$ can be calculated as

$$\Sigma_i(k+1) = E[\xi^T(k)\xi(k+1)] - \text{diag}(\dot{C}_i)\dot{x}_i(k+1)\Sigma_i(k+1)$$

$= E[\{\text{diag}(C_i)\{\Psi(k+1) + e^{a}(k+1)\} + \text{diag}(1 - \gamma_i) I_m\} \cdot \{\nu(k+1)\}^T]$

$= (\text{diag}(C_i)) E[\{\Psi(k+1) + e^{a}(k+1)\} \cdot \{\nu(k+1)\}^T + R_i]$  

$= (\text{diag}(C_i)) P^a(k+1) + \text{diag}(C_i)$

(37)

where $\Psi(k+1) \triangleq \text{diag}((1 - \gamma_i) I_m) \Sigma_i(k) \otimes [x(k) - x(k+1)]$, and $P^a(k+1) \triangleq E[\{\Psi(k+1) + e^{a}(k+1)\} \cdot \{\nu(k+1)\}^T]$. Define $\tilde{P}^a(k+1)$ as the $i$th block diagonal matrix of $P^a(k+1)$. From (37), we have $\Sigma_i(k+1) = C_i \tilde{P}^a(k+1) C_i^T + R_i$. As shown in (14), $D(z_i^a(k)) \|z_i(k)\|$ is composed of three terms, where the second term is a constant and, thus we need to derive the upper bounds of the rest of terms in (14). Based on

$$\text{tr} \{E[\Psi(k)\cdot\cdot\cdot] \}$$

$$= \text{tr} \{\text{diag}(I) \text{E}[x(k) - x(k)]^T \}$$

$$\leq N \text{tr}(E[\{x(k)^T - x(k)\}^T])$$

$$\leq 2N \text{tr}(\sum_{s=0}^\infty \ldots + N A^{k-\gamma} \text{tr}(\Pi_0))$$

$$\leq 2N \text{tr}(\nu A + N(\rho A)^2 \text{tr}(\Pi_0)), \quad (38)$$

we can derive that

$$\text{tr} \{\Sigma_i(k) \Sigma_i(k) \} = \text{tr} \{\{C_i \tilde{P}^a(k+1) C_i^T + R_i\} \Sigma_i(k) \}$$

$$\leq \text{tr}(C_i \tilde{P}^a(k+1) C_i^T) + \text{tr}(\Sigma_i(k) \otimes I_m \Sigma_i(k) \otimes I_m)\text{tr}(\Pi_0)$$

$$\leq \omega_4 + \lambda_{\min}(\Sigma_i(k)^2) \Sigma_i(k) \otimes I_m \otimes I_m) \text{tr}(\Pi_0)$$

$$\leq \omega_4 + 2\omega_5 \text{tr}(\nu A) + 2N \text{tr}(\nu A) + N(\rho A)^2 \text{tr}(\Pi_0) + \omega_6,$$

where $\omega_4 \triangleq \text{tr}(\Sigma_i(k)^{-1} R_i)$, $\omega_5 \triangleq \lambda_{\max}(\Sigma_i(k)^T \Sigma_i(k)^2) C_i$, and $\omega_6 \triangleq 2N \text{tr}(\nu A) + N(\rho A)^2 \text{tr}(\Pi_0) + \omega_6$. Based on the positive semi-definite property of the first term of (37), we have $\Sigma_i(k) \geq R_i$. Then, we can obtain that

$$\log(\text{det}(\Sigma_i(k)^{-1})) = -\log(\prod_{s=1}^m A_s(\Sigma_i(k)^{-1} \Sigma_i(k)(\cdot)))$$

$$\leq -m_i \log(\lambda_{\min}((\Sigma_i(k)^{-2} \Sigma_i(k)(\cdot))))$$

$$= -m_i \log(\frac{1}{X_z(k) X_z(k)^T} [X_z(k) \Sigma_i(k)^{-2} R_i(k) X_z(k)^T])$$

$$+ X_z(k) \Sigma_i(k)^{-2} (\Sigma_i(k)^{-1} - R_i(k) X_z(k)^T)$$

$$\leq -m_i \log(\lambda_{\min}(\Sigma_i(k)^{-2} R_i(k))), \quad (40)$$

where $X_z$ represents the minimal left eigenvector of the matrix $\Sigma_i(k)^{-2} \Sigma_i(k)(\cdot)$, the last inequality is based on the Rayleigh Quotient and the fact that $\Sigma_i(k) - R_i \geq 0$. Hence, by substituting (39) and (40) into (14), the upper bound of $D(z_i^a(k)) \|z_i(k)\|$ can be derived as (15).

E. Proof of Theorem 3

For $P^a(k+1)$ in (28), its second term is a constant matrix since $\Gamma$ is stable, while its first term is $E[\xi(k)^T \cdot \cdot\cdot]$, such that $\text{tr}(P^a(k+1)) \geq \text{tr}(E[\xi(k)^T \cdot \cdot\cdot])$. Hence, we just derive a lower bound of $E[\xi(k)^T \cdot \cdot\cdot]$ based on (29). Note that when $A$ is unstable, $A^*$ in the second last equality of (29) no longer converges to the zero matrix. Thus, the first and last terms in (30) should be rewritten as $[\Phi_3(k-\tilde{k}) I_N \otimes (A^{\kappa_0} - A^{\kappa})] 1_N 1_N^T \otimes \Pi_0 \|z_i(k)\|^T$ and $E[\Phi_3(k-\tilde{k}) I_N \otimes (\gamma^* - A^{\kappa})] 1_N 1_N^T \otimes Q_1^T$, respectively. We first focus on the first term in (30), and derive a low bound of its trace as follows:

$$\text{tr} \{[\Phi_3(k-\tilde{k}) I_N \otimes (A^{\kappa_0} - A^{\kappa})] 1_N 1_N^T \otimes \Pi_0 \|z_i(k)\|^T$$

$$\geq \text{tr} \{\sum_{s=0}^\infty \text{tr}(\frac{\Gamma^s Y_1^N \otimes (A^{\kappa-s} - A^{\kappa=s})] 1_N 1_N^T \otimes (\frac{\lambda_{\min}(\Pi_0)}{\|A\|^2}) X_A(1_X^A)^T) \})$$

$$= \lambda_{\min}(\Pi_0) \|z_i(k)\|^2 (\frac{\lambda_{\max}(A)^{k-\gamma} - \lambda_{\max}(A)^{k-\gamma}}{\|A\|^2})^T$$

$$\leq \lambda_{\min}(\Pi_0) \|z_i(k)\|^2 f_A^T,$$

(41)

where $f_A$ is the maximal right eigenvector of $A$, i.e., $A^T X_A = \lambda_{\max}(A) X_A$, the first inequality is based on $\lambda_{\min}(\Pi_0) I_{1_N} \otimes \Pi_0 \|z_i(k)\|^T \leq X_A(1_X^A)^T \Pi_0 \|z_i(k)\|^T$. Define $\Delta_3$ as the rest terms of (30). Similarly, its trace is lower bounded by

$$\text{tr} \{\Delta_3 \} \geq \frac{\lambda_{\min}(Q)}{\|X_A\|^2} \|I_N \otimes \sum_{s=0}^\infty \gamma^s Y_1^N \otimes (A^{\kappa=s} - A^{\kappa=s})] 1_N 1_N^T \otimes \Pi_0 \|z_i(k)\|^T$$

$$\geq \frac{\lambda_{\min}(Q)}{\|X_A\|^2} \|I_N \otimes \sum_{s=0}^\infty \gamma^s Y_1^N \otimes (A^{\kappa=s} - A^{\kappa=s})] 1_N 1_N^T \otimes \Pi_0 \|z_i(k)\|^T$$

$$\geq \lambda_{\min}(Q) \|z_i(k)\|^T \|I_N \otimes \sum_{s=0}^\infty \gamma^s Y_1^N \otimes (A^{\kappa=s} - A^{\kappa=s})] 1_N 1_N^T \otimes \Pi_0 \|z_i(k)\|^T$$

$$= \lambda_{\min}(Q) \|z_i(k)\|^T \|I_N \otimes (A^{\kappa-s} - A^{\kappa=s})] 1_N 1_N^T \otimes \Pi_0 \|z_i(k)\|^T$$

$= \lambda_{\min}(Q) \|z_i(k)\|^T \|I_N \otimes (A^{\kappa-s} - A^{\kappa=s})] 1_N 1_N^T \otimes \Pi_0 \|z_i(k)\|^T$
large, and \( f_2^T(k) \triangleq \frac{f_2^T(k)}{\kappa(k+1)} \). Based on (41) and (42), we can thus complete the proof.

\[ \frac{1}{k+1}(1 - \lambda_{\max}(A))^2 \]

**F. Proof of Lemma 4**

Similar to the proof of (33), one can first derive the closed-form expression of \( P_{\alpha}(k+1) \) in (37). Then, the lower bound of the first term in (14) is given by

\[
\tr(S^\alpha_i(k+1)S_i(k)+1) \geq \tr((\Sigma_i(k+1)^{-\frac{1}{2}} \Sigma_i^\alpha(k+1)[\cdot])^T) \geq \min \left\{ \Sigma_i(k+1)^{-\frac{1}{2}}(\cdot) \right\} \geq \frac{\| \Phi_\alpha(k) - \tilde{k} \|_2^2}{\| X_{A} \|_2^2 \lambda_{\max}(\Sigma_i(k))} \left( \min (\Pi) f_2^T(k) + \min (Q) f_3^T(k) \right),
\]

where the proof of the last inequality is similar to (41) and (42), and thus is omitted for simplicity, and \( \Phi_\alpha(k) \triangleq \Sigma_i(k+1)^{-\frac{1}{2}} \Sigma_i^\alpha(k+1)[\cdot] \). Note that the continues function \( f_2(x) = x - e \log(x) \geq 0 \) when \( x > 0 \). Hence, for the K-L divergence in (14), we can obtain that

\[
\tr(S^\alpha_i(k)S_i(k)-1) + \frac{\log(\det(\Sigma_i(k)))}{\det(\Sigma_i(k))} = \sum_{s=1}^{m} \lambda_s(\Sigma_i(k)^{-\frac{1}{2}} \Sigma_i^\alpha(k)[\cdot]) - \log(\lambda_s(\Sigma_i(k)^{-\frac{1}{2}} \Sigma_i^\alpha(k)[\cdot]))) \geq (1 - \frac{1}{e}) \tr(S^\alpha_i(k)S_i(k)-1),
\]

which completes the proof.

\[ \text{G. Proof of Theorem 7} \]

If the adversary chooses the time slot \( k \triangleq k_0 + \tau \) to attack the estimation system (3) when \( x(\tilde{k}) = x(k_0) \), it implies that \( \Phi_\alpha(k) \triangleq \Sigma_i(k+1)^{-\frac{1}{2}} \Sigma_i^\alpha(k+1)[\cdot] \) of the first equality in (29) equals to zero. Then, due to the fact that

\[
\Gamma^{\ast+1}(1 - \gamma_i)[I_{N_n}] = \Gamma^{\ast+1}[I_{N_n} - \epsilon L] \Lambda_A \lambda_{\lambda_{\max}(A)}(1 - \gamma_i)[I_{N_n}] = \Lambda_A \lambda_{\lambda_{\max}(A)}(1 - \gamma_i)[I_{N_n}, \Lambda_A],
\]

where \( \Phi_\alpha(s) \triangleq \sum_{s=0}^{\tau-1} [\cdot]^T \), \( \tau - 1 \) in (37) can be rewritten as:

\[
\tilde{P}_{\alpha}(k+1) = \frac{1}{k+1} \left( \Psi \tilde{P}_{\alpha}(k) + \epsilon \tilde{P}_{\alpha}(k) \right).
\]

Note that \( L \) satisfies \( L_{1N} = 0 \), which implies that \( \Phi_\alpha(s) \Lambda_{\lambda_{\max}(A)}(1 - \gamma_i)[I_{N_n}] = \Lambda_{\lambda_{\max}(A)}(1 - \gamma_i)[I_{N_n}, \Lambda_{\lambda_{\max}(A)}], \) and its particular solution is \( \Lambda \Lambda_{\lambda_{\max}(A)}(1 - \gamma_i)[I_{N_n}] \). Note that if the \( \gamma_i \) is a zero eigenvalue of \( \Lambda \), then \( \Lambda_{\lambda_{\max}(A)}(1 - \gamma_i)[I_{N_n}] = 0 \), and thus \( -1 < 1 - \epsilon \lambda_{\lambda_{\max}(A)}(1 - \gamma_i)[I_{N_n}] \) in (64), which is based on \( \epsilon < \frac{1}{\lambda_{\lambda_{\max}(A)}} \). Besides, we have \( \lambda_{\lambda_{\max}(A)}(1 - \gamma_i)[I_{N_n}] = \Lambda_{\lambda_{\max}(A)}(1 - \gamma_i)[I_{N_n}, \Lambda_{\lambda_{\max}(A)}] \). Hence, \( \Lambda \) has zero eigenvalues if and only if \( \lambda_{\lambda_{\max}(A)}(1 - \gamma_i)[I_{N_n}] = 0 \) and \( \lambda_{\lambda_{\max}(A)}(1 - \gamma_i)[I_{N_n}, \Lambda_{\lambda_{\max}(A)}] \). Based on Assumption 1, \( \lambda \) has only one zero eigenvalue with its eigenvector \( 1_N \). However, \( \Lambda \) may have multiple maximum eigenvalues, whose multiplicity is defined as \( \gamma_A \). Thus, \( \Lambda \gamma_{\gamma_{\gamma_{A}}} \) eigenvalues, and the number of general solutions of (64) is no more than \( \gamma_{\gamma_{A}} \). Rewriting the \( \gamma_i \) maximum eigenvector of \( \Lambda \) as \( X_{\gamma}^{(i)}, \gamma_i \in \{1, 2, ..., \gamma_{\gamma_{A}}\} \). By substituting \( 1_N \times X_{\gamma}^{(i)} \) into the left hand side of (46), we have \( \Lambda_{\lambda_{\max}(A)}(1 - \gamma_i)[I_{N_n}] = 0 \). Thus, (46) has \( \gamma_{\gamma_{A}} \) general solutions in the form of \( 1_N \times X_{\gamma}^{(i)} \), and \( \tilde{\Lambda} \) can be rewritten as

\[
\tilde{\Lambda} = \sum_{\gamma_i=1}^{\gamma_{\gamma_{A}}} \left[ \alpha_i \Lambda_{\lambda_{\max}(A)}(1 - \gamma_i)[I_{N_n}] + \Lambda_{\lambda_{\max}(A)}(1 - \gamma_i)[I_{N_n}, \Lambda_{\lambda_{\max}(A)}] \right],
\]

where \( \alpha_i \) is a constant coefficient of \( 1_N \times X_{\gamma}^{(i)} \). Note that the left
hand side of the above equation is a constant vector. Hence, if (22) is satisfied, the attack may keep stealthy only if \( \gamma_i \) is the same for \( \forall i \in V \). When \( A \) has only one maximum eigenvector, (47) can be simplified into \( \alpha_1 N \otimes X_A + [\text{diag}(1 - \gamma_i)I_N] \otimes X_A \), where \( \alpha \) is a constant coefficient of \( I_N \otimes X_A \). By substituting it into \( \text{diag}(C_i) \tilde{Y} = 0 \), we have \( \alpha + 1 - \gamma_i C_i X_A = 0, \forall i \in V \). Thus, if the condition (22) holds, the attacker must choose an identical \( \gamma_i \) to remain stealthy, i.e., \( \gamma_i = 1 + \alpha, \forall i \in V \). It implies that except for \( \gamma_i = 0, \forall i \in V \), any other attack (4) is detectable, since \( \text{diag}(C_i) \tilde{Y} \) does not equal to the zero vector.

I. Proof of Corollary 2

In the case that \( \gamma_i = 0, \forall i \in V \), \( \Phi_i(s) \) equals to the zero vector, since \( L_1 N = 0 \). Therefore, from (44), one can derive that \( \Gamma^k+1 N \otimes X_A = \lambda_{\text{max}}(A) \Gamma^k N \otimes X_A - \Phi_i(s) N \otimes X_A \). Then, according to (28) and (30),

\[
\text{tr} \left( P^k \left( k + 1 \right) \right) \geq \text{tr} \left( E \left[ \left( \tilde{Z}(k) \right)^T \right] \right)
\]

\[
\geq \lambda_{\text{min}}(Q) \left\{ \lambda_{\text{max}}(A) \right\}^2 \left( k + 1 \right)
\]

When \( \gamma_i = 0, \forall i \in V \), from (7) and (9),

\[
\text{tr} \left( E \left[ \left( \tilde{x}(k) \right)^T \right] \right)
\]

where the second last inequality is due to \( x(k_0 + \tau) = x(k_0) \), and the last equality is based on the steady state assumption of \( e(k_0) \) and \( e(k_0 + \tau) \). Since \( \Gamma \) is stable, \( \text{tr} \left( E \left[ \left( \tilde{x}(k) - x(k - \tau) \right)^T \right] \right) \) converges to zero, which thus completes the proof.

References


Jiahao Huang received the B.S. degree in school of control science and engineering from East China University of Science and Technology, in 2017. He is currently pursuing the Ph.D. degree from East China University of Science and Technology.

His research interests include cyber-physical system security, networked state estimation and their applications in practical engineering.

Wen Yang (Member, IEEE) is a Professor at East China University of Science and Technology. She received her BSc degree in Mineral Engineering in 2002 and MSc degree in Control Theory and Control Engineering from Central South University in 2005, Hunan, China, and PhD degree in Control Theory and Control Engineering from Shanghai Jiao Tong University, Shanghai, China, in 2009. She was a Visiting Student with the University of California, Los Angeles, from 2007 to 2008. Her research interests include Information fusion, state estimation, network security, coordinated control, complex networks and reinforcement learning.

Daniel W. C. Ho (Fellow, IEEE) received the B.S., M.S., and Ph.D. degrees in mathematics from the University of Salford, Greater Manchester, U.K., in 1980, 1982, and 1986, respectively.

From 1985 to 1988, he was a Research Fellow with the Industrial Control Unit, University of Strathclyde, Glasgow, U.K. In 1989, he joined the City University of Hong Kong, Hong Kong, where he is currently a Chair Professor of applied mathematics, and an Associate Dean with the College of Science. He has over 230 publications in scientific journals. His current research interests include control and estimation theory, complex dynamical distributed networks, multiagent systems, and stochastic systems.

Prof. Ho is a Fellow of the IEEE. He was a recipient of the Chang Jiang Chair Professor Awarded by the Ministry of Education, China, in 2012 and the ISI Highly Cited Researchers Award in Engineering by Clarivate Analytics, from 2014 to 2019. He has been on the Editorial Board of a number of journals including the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, IET Control Theory and Its Applications, the Journal of the Franklin Institute, and the Asian Journal of Control.

Fangfei Li (Member, IEEE) received the Ph.D. degree in applied mathematics from Tongji University, Shanghai, China, in 2012.

From Aug. 2016 to Aug. 2017, she was a Research Fellow with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. From Jan. 2018 to Feb. 2018, she was a Senior Research Assistant with the Department of Mathematics, City University of Hong Kong. She is currently a Professor with the School of Mathematics, East China University of Science and Technology, Shanghai. She has published more than 30 papers in international journals. Her research interests include logical systems, semi-tensor product of matrices, secure control of cyber physical systems, artificial intelligence, and their applications.

She is an Associate Editor of IEEE ACCESS.

Yang Tang (Senior Member, IEEE) received the B.S. and Ph.D. degrees in electrical engineering from Donghua University, Shanghai, China, in 2006 and 2010, respectively.

From 2008 to 2010, he was a Research Associate with The Hong Kong Polytechnic University, Hong Kong. From 2011 to 2015, he was a Post-Doctoral Researcher with the Humboldt University of Berlin, Berlin, Germany, and with the Potsdam Institute for Climate Impact Research, Potsdam, Germany. Since 2015, he has been a Professor with the East China University of Science and Technology, Shanghai. His current research interests include distributed estimation/control/optimization, cyber-physical systems, hybrid dynamical systems, computer vision, reinforcement learning and their applications.

Prof. Tang was a recipient of the Alexander von Humboldt Fellowship. He has been the ISI Highly Cited Researchers Award by Clarivate Analytics since 2017. He is also a Senior Board Member of SCIENTIFIC REPORTS, and an Associate Editor of IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, IEEE TRANSACTIONS ON CYBERNETICS, IEEE TRANSACTIONS ON CIRCUITS AND IEEE SYSTEMS-1:REGULAR PAPERS, IEEE TRANSACTIONS ON EMERGING TOPICS IN COMPUTATIONAL INTELLIGENCE, IEEE SYSTEMS JOURNAL and ENGINEERING APPLICATIONS OF ARTIFICIAL INTELLIGENCE (IFAC Journal). He is also a Leading Guest Editor for Special Issues in IEEE TRANSACTIONS ON EMERGING TOPICS IN COMPUTATIONAL INTELLIGENCE and IEEE TRANSACTIONS ON COGNITIVE AND DEVELOPMENTAL SYSTEMS.