2019 American Control Conference (ACC)
Philadelphia, PA, USA, July 10-12, 2019

Resilient Distributed Filter for State Estimation of Cyber-Physical Systems Under Attack

Raj Gautam Dutta[1], Teng Zhang[2], and Yier Jin[1]

Abstract— **Proliferation of distributed Cyber-Physical Systems has raised the need for developing computationally efficient security solutions. Toward this objective, distributed state estimators that can withstand attacks on agents (or nodes) of the system have been developed, but many of these works consider the estimation error to asymptotically converge to zero by restricting the number of agents that can be compromised. We propose Resilient Distributed Kalman Filter (RDKF), a novel distributed algorithm that estimates states within an error bound and does not depend on the number of agents that can be compromised by an attack. Our method is based on convex optimization and performs well in practice, which we demonstrate with the help of a simulation example. We theoretically show that, in a connected network, the estimation error generated by the Distributed Kalman Filter and our RDKF at each agent converges to zero in an attack free and noise free scenario. Furthermore, our resiliency analysis result shows that the RDKF algorithm bounds the disturbance on the state estimate caused by an attack.**

I. Introduction

Distributed (or networked) Cyber-Physical Systems (CPS) are the result of seamless integration of computational, physical, and network components. They are becoming increasingly ubiquitous in many industrial sectors such as energy, space, health-care, agriculture, transportation, building automation, and manufacturing. Application of distributed CPS in such diverse sectors is forcing system developers to think beyond their conventional design process. In addition to the standard design specifications, they are required to consider the uncertainties arising from such systems operating in complex, unpredictable, and contested environments. Security is another major concern for distributed system developers. Recent real-world attacks such as the massive power outage at the Ukrainian capital by the Crash Override malware [1], leveraging phishing emails to cause multiple component failure at a German steel mill [2], and compromising sensors and communication network of semi-automated ground vehicles [3], has raised the need for secure distributed systems.

Unlike IT systems where cybersecurity entails protection of data, cyber attacks on a networked CPS can impact the physical dynamics of the system by corrupting the state estimates of some of its components. Thus, such systems pose new security issues, which cannot be addressed with the existing cybersecurity or benign fault detection solutions. In the quest for building a secure distributed CPS, early researchers have developed centralized and decentralized attack resilient filters for CPS, which requires aggregation of sensor measurements at a particular location or at all the components of the system for state estimation [4], [5].

Due to the computational inefficiency of both centralized and decentralized approaches, efforts have been made towards the design of distributed state estimators, where a component (or agent) of the networked CPS asymptotically estimate the system state based on partial information of the state from its neighbors [6]–[8]. Most of these methods are variations of the Distributed Kalman Filter (DKF) of [6]. The earliest DKF algorithm solved the estimation problem in two steps: in the first step, a dynamic average-consensus filter was used for fusion of sensor and covariance data and in the next step Kalman filter update rules were used for recursively estimating the states. Convergence of the DKF depended on the topology of the communication network. Subsequently, single time scale strategies were developed for the DKF.

Until recently, very few attempts were made on designing attack resilient distributed state estimators [9]–[11]. Khan and Stankovic [9] proposed attack detection and single message exchange state estimation methods for a compromised communication scenario. Their estimator relied on statistical consistency of nodal and local data sets and physical-layer feedback. Matei *et al.* [10] designed a multi-agent filtering scheme in conjunction with a trust-based mechanism to secure the state estimates of power grids under a false data injection attack. In their approach, an agent of the grid computes local state estimates based on their own measurement and of their trusted neighbors. However, both [9], [10] did not provide any theoretical guarantees of their methods. Mitra and Sundaram [11] developed a secure distributed observer for the Byzantine adversary model, where some nodes of the network were compromised by an adversary. Prior to state estimation, they decomposed the linear system model using Kalman's observability decomposition method. Then, Luenberger observer at each node estimated the states corresponding to detectable eigenvalues. The undetectable portions of the states at each node were estimated using a secure consensus algorithm, which used measurements of well-behaving neighboring nodes. However, their method requires the network to be highly connected and they assume that only a small number of nodes are corrupted; and this number is known for their algorithm. In addition, they assume that the system matrix A only has simple and real eigenvalues, which might not hold in practice.

In this paper, we model the distributed CPS as a linear time-invariant system. A malicious attack corrupts the sensor measurements of some agents of the system. Consequently,

[1]Raj Gautam Dutta and Yier Jin are with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, 32611, USA (r.dutta, yier.jin)@ufl.edu

[2]Teng Zhang is with the Department of Mathematics, University of Central Florida, Orlando, FL, 32816, USA teng.zhang@ucf.edu

*This work is partially supported by National Science Foundation (CNS-1818500).

based on DKF, we develop Resilient Distributed Kalman Filter (RDKF), which is resilient to attacks and we provide theoretical guarantees. We show the asymptotic convergence of estimation error to zero for both DKF and RDKF when there is no attack and no noise and our resiliency analysis shows that the disturbance on the state estimate of RDKF caused by an attack is bounded. Compared to [11], our RDKF and its analysis does not make assumptions on the structure of the graph except being connected, the number of corrupted nodes, and the eigenvalues of system matrix A. The unique features of our method are:

- Our method can recursively estimate states of a distributed CPS without considering the number of neighbors of an agent that can be compromised. This particular characteristic separates our method from the rest and is effective in scenarios where adversaries might exist in greater numbers.
- Our method can *approximately* (within an error bound) reconstruct the system states and its performance does not degrade (beyond an upper bound) with the magnitude of the attack.

The rest of the paper is organized as follows: In Section II, we formulate the problem and describe the distributed CPS model and the measurement attack model. DKF, RDKF, and their performance analysis are explained in Section III. The effectiveness of RDKF is demonstrated on numerical examples in Section IV. Final conclusions are drawn in Section V and proofs are given in the Appendix.

II. Preliminaries and Problem Description

A. Notations

We assume that there are n agents, $\mathrm{X} \triangleq \{1, 2, \ldots, n\}$, in the distributed CPS, whose communication with each other can be described by an undirected graph $G = (\mathcal{V}, \mathcal{E})$. In the graph, nodes are the number of agents, $\mathcal{V} = \mathrm{X}$, and edges, $\mathcal{E} = \mathcal{V} \times \mathcal{V}$, represent communication between them. Here, $(i, j) \in \mathcal{E}$ is a bi-directional edge between i and j, that enable them to send and receive messages among themselves, but not simultaneously. We assume that every agent in X has a self loop i.e. $(i, i) \in \mathcal{E}$ for all $i = 1, 2, \ldots n$. Neighborhood of i is defined as the set of nodes that are adjacent to it i.e. $N^{(i)} = \{i\} \cup \{j \in \mathcal{V} : (i, j) \in \mathcal{E}\}$ and with whom it can communicate. Furthermore, we assume that each agent has an observer for estimating the state of the system. In the paper, we use the words agent and node interchangeably.

B. System and Measurement Models without Attack

We model the dynamics of the distributed CPS as a linear time-invariant (LTI) system, which is described below:

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k \tag{1}$$

where, $\mathbf{x}_k \triangleq [\mathbf{x}_k^{(1)}, \mathbf{x}_k^{(2)}, \ldots, \mathbf{x}_k^{(n)}] \in \mathbb{R}^n$ is the state vector at time $k \in \mathbb{N}$ of the distributed system and $\mathbf{A} \triangleq [A_{i,j}] \in \mathbb{R}^{n \times n}$ is the system matrix, with $A_{i,j}$ representing block matrix of i and its neighbors j.

In the distributed system, each agent measures the system state at time k, which is given by $\mathbf{y}_k^{(i)} = \mathbf{C}^{(i)}\mathbf{x}_k$ where, $\mathbf{y}_k^{(i)} \in \mathbb{R}^q$ is the measurement from q sensors of the agent and $\mathbf{C}^{(i)} \in \mathbb{R}^{q \times n}$ is the observation matrix. For analytical convenience, we represent the aggregated measurement vectors and observation matrices as

$$\mathbf{y}_k = \mathbf{C}\mathbf{x}_k, \tag{2}$$

where $\mathbf{y}_k \triangleq [\mathbf{y}_k^{(1)}, \mathbf{y}_k^{(2)}, \ldots, \mathbf{y}_k^{(n)}]$ and $\mathbf{C} \triangleq [\mathbf{C}^{(1)}, \mathbf{C}^{(2)}, \ldots, \mathbf{C}^{(n)}]$. In our model, we assume that each agent estimates the system state, $\mathbf{x}_k$, at each time-step k based on the measurements gathered from its neighbors and its own. Also, an agent (good or malicious) is assumed to transmit the same information to all its neighbors. This assumption appears in many practical scenarios such as in vehicular ad-hoc networks.

C. Measurement Model with Attack

We consider an insider attack, where an adversary has complete control over a set of nodes $\mathcal{V}_a \subset \mathcal{V}$ of the communication network. Such an attacker has knowledge of the observation matrices, $C^{(j)}$, of its neighbors, system matrix $\mathbf{A}$, and the communication topology. With this information, he/she can influence the state of the system without affecting the message scheduler of the network. The reason for considering such a strong adversary model is to show that our resilient estimator can withstand the worst-case scenario.

The attack is carried out by manipulating the sensor data of the compromised agents and can be represented by the following equation:

$$\mathbf{y}_k^{(i),a} = \mathbf{C}^{(i)}\mathbf{x}_k + \mathbf{a}_k^{(i)} \tag{3}$$

where, $\mathbf{a}_k^{(i)}$ is the attack vector and $\mathbf{y}_k^{(i),a}$ is the corrupted output of agent i. Such malicious measurements effect the state estimate of the agent which, when used by its neighbors affect their estimate as well. Consequently, the attack influences the state estimate of the distributed system. We provide the following definition of a compromised agent:

Definition II.1. *Compromised Agent*: An agent i is compromised at time $k \in \mathbb{N}$ if its attack vector $\mathbf{a}_k^{(i)} \neq 0$.

As the agents are completely controlled by an adversary, we do not make any assumption on the number of sensors that were manipulated. Also, unlike the f-adversarial attack model of [11], where they consider an upper bound on the number of adversarial neighbors of an agent, we do not put any such restriction. However, the performance of our estimator degrades within an error bound as the number of compromised neighbors of an agent increases.

D. Secure Distributed Estimation Problem

Given a linear time-invariant distributed system of n agents with a linear measurement model and an undirected communication graph G, our goal is to design a filter that can estimate system states such that $\lim_{k\to\infty} \|\hat{\mathbf{x}}_k^{(i)} - \mathbf{x}_k\| \to 0$, $\forall i \in \mathbb{R}^n$ when there is no attack and the estimation errors are bounded when sensor measurement of a subset of nodes $\mathcal{V}_a \subset \mathcal{V}$ are compromised by an insider attack.

For such an estimator, we make the following assumptions:

- Matrix $(\mathbf{A}, \mathbf{C})$ is detectable of the system. This assumption is in line with the assumption made in [7], [11], where it is considered a necessary condition for solving the distributed estimation problem with asymptotic guarantees.
- Each agent shares their estimated state information with neighbors via a secure communication channel. Thus, we do not consider any attack on the network.
- We assume that an agent cannot detect the sensor attack on its neighbors and thus accepts the corrupted state estimates from them.

III. Secure Distributed Estimation Methods

A. Distributed Kalman Filter

In this section, we first analyze the performance of the DKF for no attack and no noise scenarios. The motivation behind this analysis are two fold: (i) to the best of our knowledge, this is the first convergence result on DKF for no noise and no attack scenarios and thus, is one of the contributions of this paper and (ii) it forms the basis for the analysis of our attack resilient estimator in Section III-B.

We investigate the attack free case for the following model,

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k,\ \mathbf{y}_k^{(i)} = \mathbf{C}^{(i)}\mathbf{x}_k \tag{4}$$

We assume that the estimation error covariance matrix, $\mathbf{P}^{(i)}$, is chosen according to the following equation,

$$\mathbf{P}^{(i)} = \left(\frac{1}{d_i}\sum_{j\in N^{(i)}}(\mathbf{A}\mathbf{P}^{(j)}\mathbf{A}^T + \Sigma_w^{(j)\,-1}) + \mathbf{C}^{(i)\,T}\Sigma_v^{(i)\,-1}\mathbf{C}^{(i)}\right)^{-1} \tag{5}$$

where, $N^{(i)} = \{i\} \cup \{\text{neighbors of } i \text{ in } G\}$ and $d_i = |N^{(i)}|$, is the total number of neighbors of node i. By assuming $(\mathbf{A}, \mathbf{C})$ is observable in our model, we get a steady-state DKF with error covariance matrix, $\mathbf{P}^{(i)} = \lim_{k\to\infty}\mathbf{P}_k^{(i)}$, of equation (5). **Theorem III.1** of this paper states the convergence result of $\mathbf{P}^{(i)}$. While in Kalman filter, $\Sigma_v^{(i)}$ and $\Sigma_w^{(i)}$ are commonly used to denote the covariance of the noise in the system, they can also be treated as parameters for developing the algorithm in the noise-free setting (Kalman filter application in the noise-free setting is discussed in [12]). In principle, they can be chosen to be any positive definite matrices. The impact of the values of $\Sigma_v^{(i)}$ and $\Sigma_w^{(i)}$ on the estimate are discussed in Section III-B.

Now, the distributed estimator has following prediction rules,

$$\mathbf{P}_{|}^{(i)} = \mathbf{A}\mathbf{P}^{(i)}\mathbf{A}^T + \Sigma_w^{(i)} \tag{6}$$

$$\hat{\mathbf{x}}_k^{(i)} = \mathbf{P}_k^{(i)}\left(\frac{1}{d_i}\sum_{j\in N^{(i)}}\mathbf{P}_{|}^{(j)\,-1}\mathbf{A}\hat{\mathbf{x}}_{k-1}^{(j)} \quad + \mathbf{C}^{(i)\,T}\Sigma_v^{(i)\,-1}\mathbf{y}_k^{(i)}\right) \tag{7}$$

where, $\hat{\mathbf{x}}_k^{(i)}$ is the state estimate and $\mathbf{P}_{|}^{(i)}$ is *a priori* estimation error covariace of agent i. This estimator is motivated from the DKF studied in [6]–[8].

Before using this estimator, we need to ensure that a solution to equation (5) exists. Thus, we give the following theoretical guarantee on the existence of the solution.

Theorem III.1. If the graph G is connected, $\mathbf{A}$ is full-rank, $(\mathbf{A}, \mathbf{C})$ is observable, and $\Sigma_v^{(i)}$ is full rank for all $1 \le i \le n$, then there exist $\{\mathbf{P}^{(i)}\}_{i=1}^n$ that satisfy equation (5).

The proof of Theorem III.1 (in the Appendix) shows that the covariance matrices of the estimator converges when they are initialized as zero matrices. In comparison, there exist works on convergence of the covariance matrices of DKF: [13] proves the convergence of the covariance using probability theory and [8] performs convergence analysis on a modified DKF which has one prediction/update step at each time point. We remark that the convergence analysis in the standard Kalman filter uses observability assumption and as such it is the optimal assumption we could make as well.

The following theorem states the main result of distributed estimation without attacks and noises and its proof is in the Appendix. This theorem states that the estimation error converges to zero in the attack free and noise free scenarios. There exist work on convergence of estimation errors of DKF. For the noise free case, Li et al. [14] proves that the estimation error converges to a unique value. However, we are not aware of any work that has convergence result for the attack free and noise free scenarios, which we have considered. Our proof is mainly based on the observation that the estimation error do not increase over time and as a result, the estimation error converges.

Theorem III.2. (*Convergence of DKF*) Under the assumptions made in Theorem III.1, the estimate of equation (7) converges to the correct solution in the sense that for all $1 \le i \le n$, $\lim_{k\to\infty}\|\hat{\mathbf{x}}_k^{(i)} - \mathbf{x}_k\| \to 0$ and the convergence rate is linear.

The result described here is called the "omniscience property" in [7], [11], which is proved under the same system setting as Theorem III.2, but for a different estimation algorithm. We remark that while the condition "$(\mathbf{A}, \mathbf{C})$ is observable" is more restrictive than the condition in [7] that "$(\mathbf{A}, \mathbf{C})$ is detectable", in practice the difference could be addressed using the idea of decomposing the system ($\mathbf{A}$, $\mathbf{C}$, $\mathbf{x}$) into two parts corresponding to stable and unstable eigenvalues of $\mathbf{A}$. Note that for $\mathbf{x}$, the stable part converges to zero, thus it is sufficient to investigate the subsystem of ($\mathbf{A}$, $\mathbf{C}$, $\mathbf{x}$) that is associated with unstable eigenvalues of $\mathbf{A}$. More specifically, let $\mathbf{A} = \mathbf{U}\mathrm{diag}(\mathbf{S}_1, \mathbf{S}_2)\mathbf{U}^{-1}$ be the Jordan transformation of $\mathbf{A}$, where $\mathbf{U}$ is the similarity transformation matrix, $\mathbf{S}_1$ is a square matrix that contain all Jordan blocks with stable eigenvalues and $\mathbf{S}_2$ consist of all Jordan blocks with unstable eigenvalues. Then, with $\tilde{\mathbf{x}}_k = \mathbf{U}^{-1}\mathbf{x}_k$ and $\tilde{\mathbf{x}}_k = [\tilde{\mathbf{x}}_{1,k}, \tilde{\mathbf{x}}_{2,k}]$, the state evolution of equation (4) is equivalent to the following equations: $\tilde{\mathbf{x}}_{k+1,1} = \mathbf{S}_1\tilde{\mathbf{x}}_{k,1}, \tilde{\mathbf{x}}_{k+1,2} = \mathbf{S}_2\tilde{\mathbf{x}}_{k,2}$. Now, we have $\|\tilde{\mathbf{x}}_{k,1}\| \to 0$ as $k \to \infty$. Thus, it is sufficient to estimate $\tilde{\mathbf{x}}_{k,2}$. To have the "omniscience property" of the estimation of $\tilde{\mathbf{x}}_{k,2}$ from

$\mathbf{y}_k^{(i)} = \mathbf{C}^{(i)}\mathbf{U}\tilde{\mathbf{x}}_k \approx \mathbf{C}^{(i)}\mathbf{U}_2\tilde{\mathbf{x}}_{k,2}$ ($\mathbf{U}_2$ is a submatrix of $\mathbf{U}$ corresponding to the component $\mathbf{S}_2$), Theorem III.2 implies that it is sufficient to have the observability of $(\mathbf{S}_2, \mathbf{CU}_2)$. Applying the "Eigenvalue assignment" from [15, Table 15.1], it can be shown that the observability of $(\mathbf{S}_2, \mathbf{CU}_2)$ is equivalent to the detectability of $(\mathbf{A}, \mathbf{C})$.

B. Resilient Distributed Kalman Filter

We discuss the design of our optimization based estimator, RDKF, which is resilient to attack on sensors of the agents. We also analyze its performance and prove that when there is no attack, the estimation errors converge to zero and in the presence of attack, the estimation errors are bounded. Our results hold even when the magnitude of the attack is unbounded.

We investigate the case with attack, which is given by the following model,

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k,\ \mathbf{y}_k^{(i),a} = \mathbf{C}^{(i)}\mathbf{x}_k + \mathbf{a}_k^{(i)}$$

and we propose our RDKF based on optimization as follows:

$$\hat{\mathbf{x}}_k^{(i)} = \arg\min_{\mathbf{x}_k} \lambda \left\| \Sigma_v^{(i)\,-\frac{1}{2}} (\mathbf{y}_k^{(i),a} - \mathbf{C}^{(i)}\mathbf{x}_k) \right\| + \frac{1}{d_i} \sum_{j \in N^{(i)}} (\mathbf{x}_k - \mathbf{A}\hat{\mathbf{x}}_{k-1}^{(j)})^T \mathbf{P}_{|}^{(j)\,-1} (\mathbf{x}_k - \mathbf{A}\hat{\mathbf{x}}_{k-1}^{(j)}) \quad (8)$$

Our method is motivated from the DKF as follows.

First, equation (7) can be considered as the following optimization problem

$$\hat{\mathbf{x}}_k^{(i)} = \arg\min_{\mathbf{x}_k} (\mathbf{y}_k^{(i),a} - \mathbf{C}^{(i)}\mathbf{x}_k)^T \Sigma_v^{(i)\,-1} (\mathbf{y}_k^{(i),a} - \mathbf{C}^{(i)}\mathbf{x}_k) + \frac{1}{d_i} \sum_{j \in N^{(i)}} (\mathbf{x}_k - \mathbf{A}\hat{\mathbf{x}}_{k-1}^{(j)})^T \mathbf{P}_{|}^{(j)\,-1} (\mathbf{x}_k - \mathbf{A}\hat{\mathbf{x}}_{k-1}^{(j)}) \quad (9)$$

To make an optimization-based estimator more robust to attacks, a commonly used strategy is to use optimization with ℓ_1 norm on the terms affected by attack [4]. We apply a similar strategy, where we replace $(\mathbf{y}_k^{(i),a} - \mathbf{C}^{(i)}\mathbf{x}_k)^T \Sigma_v^{(i)\,-1} (\mathbf{y}_k^{(i),a} - \mathbf{C}^{(i)}\mathbf{x}_k)$ of equation (9) with its square root and a parameter, λ, which is similar to giving a smaller penalty on the attacked measurements, $\mathbf{y}_k^{(i),a}$. This procedure makes our algorithm more resilient to attacks. The optimization problem of equation (8) does not have an explicit solution, but it can be solved efficiently as it is in convex form.

The choice of λ is critical in our approach and it gives a balance between the terms, $\left\| \Sigma_v^{(i)\,-\frac{1}{2}} (\mathbf{y}_k^{(i),a} - \mathbf{C}^{(i)}\mathbf{x}_k) \right\|$ and $\sum_{j \in N^{(i)}} (\mathbf{x}_k - \mathbf{A}\hat{\mathbf{x}}_{k-1}^{(j)})^T \mathbf{P}_{|}^{(j)\,-1} (\mathbf{x}_k - \mathbf{A}\hat{\mathbf{x}}_{k-1}^{(j)})$. Large value of λ implies more weight is placed on $\mathbf{y}_k^{(i),a}$, which includes both true and corrupted sensor values. Although such a choice of λ makes the estimation error converge to zero quickly in the absence of attack, it will make the system unstable in the presence of attack. On the contrary, when λ is small, it will take longer for the estimation errors to converge to zero, but the method will be stable against attack.

As $\Sigma_v^{(i)\,-\frac{1}{2}}$ appears in the term $\lambda \left\| \Sigma_v^{(i)\,-\frac{1}{2}} (\mathbf{y}_k^{(i),a} - \mathbf{C}^{(i)}\mathbf{x}_k) \right\|$, Σ_v will have an impact different from λ on state estimates: when Σ_v is large, it will take longer for the estimation errors to converge to zero, but the method will be stable against attacks. Similarly, large value of Σ_w will result in large $\mathbf{P}_{|}$ and it will have an impact similar to Σ_v (and similar to that of λ) on state estimates. Furthermore, experimental results in Section IV demonstrate the impact of these three parameters on state estimation error.

To analyze convergence and resiliency of RDKF, we consider two scenarios:

1) All agents are benign and the system operates normally
2) Some agents are compromised of the distributed system

We provide the following theoretical guarantee (proof is in the Appendix) for the first scenario. It suggests that when the initial estimation errors $\hat{\mathbf{e}}_0^{(i)}$ are not too large, the algorithm obeys the "omniscience property" and the estimation error converges to zero. Now, the proof of this theorem is based on the structure of the proof of Theorem III.2, i.e., we first show that the estimation error do not increase over time and then, with some additional arguments, we show that the estimation error converges to zero.

Theorem III.3. (*Convergence of RDKF*) Under the assumptions of Theorem III.1, if the initial estimation errors $\{\hat{\mathbf{e}}_0^{(i)}\}_{i=1}^n$ satisfy the following condition: For any $\mathbf{x}$ that satisfies $\|\Sigma_v^{(i)\,-\frac{1}{2}} \mathbf{C}^{(i)}\mathbf{x}\| = 2\lambda$, it has the property that $\mathbf{x}^T \mathbf{P}^{(i)\,-1}\mathbf{x} \geq \hat{\mathbf{e}}_0^{(i)\,T} \mathbf{P}^{(i)\,-1} \hat{\mathbf{e}}_0^{(i)}$, then, for the first scenario without attack, the sequence produced by equation (8) converges to the correct solution i.e. for all $1 \leq i \leq n$, $\lim_{k\to\infty} \|\hat{\mathbf{x}}_k^{(i)} - \mathbf{x}_k\| \to 0$ and the convergence rate is linear.

We remark that, while this theorem makes the assumption that the initial estimation errors $\{\hat{\mathbf{e}}_0^{(i)}\}_{i=1}^n$ are not very large, in practice we notice that our algorithm converges even when initial estimations of $\mathbf{x}_0^{(i)}$ are bad.

For the second scenario, the following resiliency theorem (proof in the Appendix) states that no matter how large the magnitudes of the attack, the deviation of the state estimate of the algorithm is bounded. Consequently, even during worst-case attack scenario, the error of the state estimate is upper bounded. Compared to Theorem III.3, which states that the estimation errors converge to zero when there is no attack, this result suggests that the estimation errors are bounded during attack. This result separates our RDKF from the traditional DKF of Section III-A, where an unbounded attack could result in an unbounded estimation error. Furthermore, our analysis and results are different from the theoretical guarantees given for the resilient distributed estimator of [11]. We have made fewer assumptions on the eigenvalues of A and the graph structure of the network; and we only show that the estimation error is bounded (rather than convergence to zero result shown in [11]).

Theorem III.4. (*Resiliency of RDKF*) Consider the optimization problem, equation (8), whose solution $\hat{\mathbf{x}}_k^{(i)}$ is based on $\mathbf{y}_k^{(i),a}$ and $\{\hat{\mathbf{x}}_{k-1}^{(j)}\}_{j \in N^{(i)}}$. In this sense, we can write the

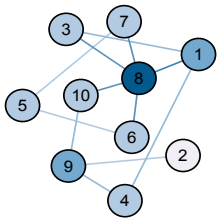

Fig. 1: Communication network of a distributed system of 10 agents and 12 edges

solution of equation (8) as a function $g : \mathbb{R}^q \times \mathbb{R}^{n\times|N^{(i)}|} \rightarrow \mathbb{R}^n$ as follows:

$$\hat{\mathbf{x}}_k^{(i)} = g(\mathbf{y}_k^{(i),a}, \{\hat{\mathbf{x}}_{k-1}^{(j)}\}_{j\in N(i)}).$$

The solution is resilient to the attack on $\mathbf{y}_k^{(i),a}$ as follows

$$\|g(\mathbf{y}_k^{(i),a}, \{\hat{\mathbf{x}}_{k-1}^{(j)}\}_{j\in N(i)}) - g(\mathbf{y}_k^{(i)}, \{\hat{\mathbf{x}}_{k-1}^{(j)}\}_{j\in N(i)})\|$$
$$\leq \lambda d_i \left\| \left(\sum_{j\in N^{(i)}} \mathbf{P}_{|}^{(j)\,-1} \right)^{-1} \Sigma_v^{(i)\,-\frac{1}{2}} \mathbf{C}^{(i)} \right\|. \quad (10)$$

This theorem implies that the disturbance on the state estimate caused by an arbitrary attack on $\mathbf{y}_k^{(i),a}$ is bounded. Although, the state estimates and the measurements are updated with time, the bound on error is independent of time. It also partially explains the observations made later in Section IV that large $\Sigma_v^{(i)}$ corresponds to more stable performance of the estimator during an attack, as from equation (10) (representing the maximum additional estimation error that can be caused by an attack), large $\Sigma_v^{(i)}$ gives a smaller upper bound for the additional estimation error.

However, we remark that this theorem only captures the impact of sporadic attack (an attack which do not occur continuously for long duration of time) on the estimation of $\hat{\mathbf{x}}_k^{(i)}$. Following this theorem, if the estimation error is small enough to satisfy the condition of Theorem III.3 after an attack, then we can consider such an estimation error as the "initial estimation error" in Theorem III.3 and use it to show that despite the attack, the estimation errors of our RDKF still converge to zero, provided we have attack-free measurements after the sporadic attack.

The long term impact of persistent attack is not considered in this paper and we leave it as possible future work.

IV. Experimental Results

In this section, we use a numerical example to demonstrate the effectiveness of our RDKF approach against sensor attacks on the nodes of distributed CPS. We represent the dynamics of the system using a linear time-invariant model given by equation (1). The sensor measurement of an agent with and without attacks are given by equation (2) & equation (3). We generate a random system matrix, $\mathbf{A}$, which is nondegenerate and random output matrices for all the agents, $\{C^i\}_{i=1}^n$. The dimension of the state $\mathbf{x}$ is 20 and the number of sensors per agent is 14.

The undirected communication graph of the system, as shown in Fig. 1, consists of 10 randomly located agents and 12 edges. Different shades of color are representative of nodes with different degrees. We simulate the distributed system dynamics, communication graph, and the sensor attack on nodes in MATLAB.

We first evaluate RDKF over the attack free scenario. Fig. 2 compares state estimation error of all the agents over a time frame of $50\,(0:50)$. Along the Y-axis is normalized estimation error of all the agents for one of the states of the agents and X-axis represent time in seconds. We observe that the state estimation error is less than 0.7 for all the agents and they converge to zero within 10 seconds.

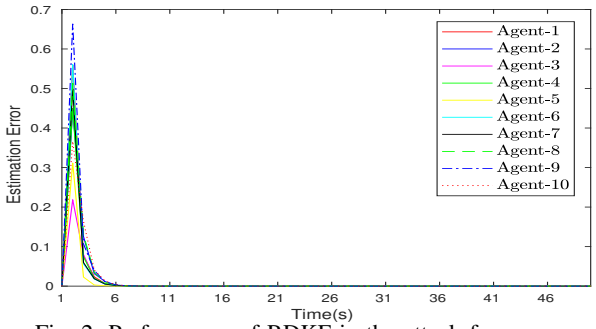

Fig. 2: Performance of RDKF in the attack free case.

Next, we consider the scenario where an adversary corrupts sensor measurements of nodes $1, 2, 3, 5, 8$ & 10 after a certain time point. In case of the attack, malicious data of random value are added to all the sensor output of the compromised nodes. We assume that the attack vector $\{\mathbf{a}_k^i\}_{i=1,2,3,5,8,10}$ is injected after time point $k = 20$ into the nodes and the probability of its occurrence at any time after initiation at the nodes is $p_a = 0.9$. For instance, in our experiment, the attack occurs at time points such as $22, 25, 32, 35, 37$ and 45. Fig. 3 provides comparison of estimation error among all nodes for all the normalized states. For the RDKF estimator, $(\lambda, \Sigma_v, \Sigma_w) = (1, \mathbf{I}, \mathbf{I})$ is used. We observe that the estimation errors are high for the neighbors,$\{3, 4, 8\}, \{9\}, \{1, 8\}, \{6, 7\}, \{1, 3, 6, 7, 10\}, \{8, 9\}$, of the compromised nodes $1, 2, 3, 5, 8, 10$, respectively. Also, estimations of the neighbors, $\{5, 8\}$ of nodes $6, 7$ and the neighbors, $\{2, 10\}$ of node 9 are corrupted. Note that the estimates obtained from our filter at attack time points does not get unbounded even when more than half of the neighbors of an agent/node are compromised. At all other time points (when there is no attack), our method performs as well as in the attack free case.

We also tried various values of parameters λ, Σ_v and Σ_w. In particular, we follow the setup of Fig. 3, with $(\lambda, \Sigma_v, \Sigma_w)$ replaced by $(0.1, \mathbf{I}, \mathbf{I})$, $(1, 10\mathbf{I}, \mathbf{I})$ and $(1, \mathbf{I}, 10\mathbf{I})$ respectively, and their results are shown in Fig. 4. As stated in Section III-B, smaller λ or larger Σ_v gives slower convergence at the beginning, but more stable performance to attacks; and larger Σ_w gives faster convergence at the beginning, but less stable performance to attacks.

We also check the performance of the DKF estimator, given by equation (9) during attacks, as is shown in Fig. 5. We see that by replacing $(\mathbf{y}_k^{(i),a} - \mathbf{C}^{(i)}\mathbf{x}_k)^T \Sigma_v^{(i)\,-1} (\mathbf{y}_k^{(i),a} - \mathbf{C}^{(i)}\mathbf{x}_k)$ of equation (9) with its square root (and with a

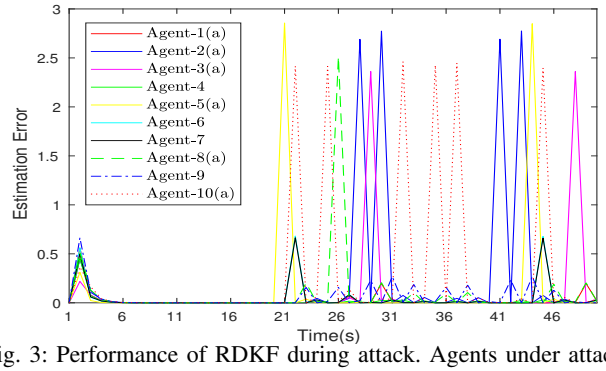

Fig. 3: Performance of RDKF during attack. Agents under attack are marked with (a). Estimation error is bounded and small.

scalar λ), RDKF becomes more resilient and its estimation errors becomes bounded during attacks.

V. Conclusion

In this paper, we have proposed RDKF, a novel attack resilient distributed state estimation algorithm that can recursively estimate states and bounds the disturbance on the state estimate caused by an attack. We prove and show using a numerical example that the estimation error of our method asymptotically convergence to zero when there is no attack and noise, and has an upper bound during attack. In future, we plan to improve our current analysis on estimation errors to stochastic systems and investigate the interplay between network connectivity, system stability, and convergence.

References

[1] D. U. Case, "Analysis of the cyber attack on the ukrainian power grid," *Electricity Information Sharing and Analysis Center (E-ISAC)*, 2016.

[2] R. M. Lee, M. J. Assante, and T. Conway, "German steel mill cyber attack," *Industrial Control Systems*, vol. 30, p. 62, 2014.

[3] S. Checkoway *et al.*, "Comprehensive experimental analyses of automotive attack surfaces.," in *USENIX Security Symposium*, pp. 77–92, San Francisco, 2011.

[4] Y. Mo and E. Garone, "Secure dynamic state estimation via local estimators," in *2016 IEEE 55th Conference on Decision and Control (CDC)*, pp. 5073–5078, Dec 2016.

[5] R. G. Dutta, X. Guo, T. Zhang, K. Kwiat, C. Kamhoua, L. Njilla, and Y. Jin, "Estimation of safe sensor measurements of autonomous system under attack," in *Proceedings of the 54th Annual Design Automation Conference 2017*, p. 46, ACM, 2017.

[6] R. Olfati-Saber, "Distributed kalman filtering for sensor networks," in *2007 46th IEEE Conference on Decision and Control*, pp. 5492–5498.

[7] S. Park and N. C. Martins, "Design of distributed lti observers for state omniscience," *IEEE Transactions on Automatic Control*, vol. 62, pp. 561–576, Feb 2017.

[8] D. Marelli, M. Zamani, and M. Fu, "Distributed Kalman Filter in a Network of Linear Dynamical Systems," *arXiv*, Nov 2017.

[9] U. A. Khan and A. M. Stanković, "Secure distributed estimation in cyber-physical systems," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 5209–5213, May 2013.

[10] I. Matei, J. S. Baras, and V. Srinivasan, "Trust-based multi-agent filtering for increased smart grid security," in *Control & Automation (MED), 2012 20th Mediterranean Conference on*, pp. 716–721, IEEE.

[11] A. Mitra and S. Sundaram, "Secure distributed observers for a class of linear time invariant systems in the presence of byzantine adversaries," in *2016 IEEE 55th Conference on Decision and Control (CDC)*.

[12] K. Rapp and P.-O. Nyman, "Stability Properties of the Discrete-Time Extended Kalman Filter," *2004 IFAC Proceedings Volumes*, vol. 37.

[13] S. Kar, S. Cui, H. V. Poor, and J. M. F. Moura, "Convergence results in distributed Kalman filtering," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2500–2503.

[14] D. Li, S. Kar, J. M. F. Moura, H. V. Poor, and S. Cui, "Distributed Kalman Filtering Over Massive Data Sets: Analysis Through Large Deviations of Random Riccati Equations," *IEEE Transactions on Information Theory*, vol. 61, pp. 1351–1372, March 2015.

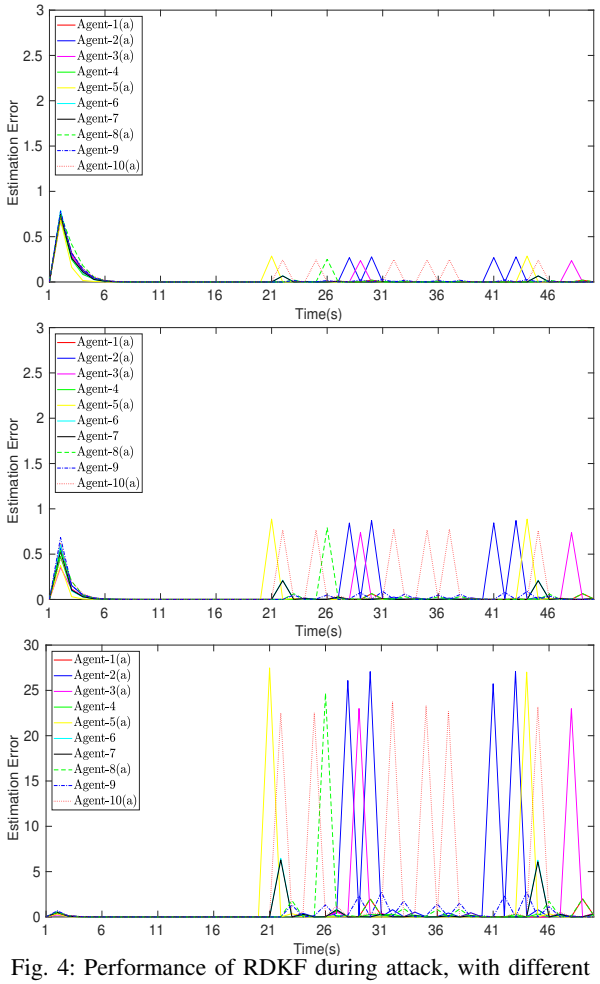

Fig. 4: Performance of RDKF during attack, with different parameter values. The three figures correspond to small value of λ, large value of Σ_v and large value of Σ_w respectively.

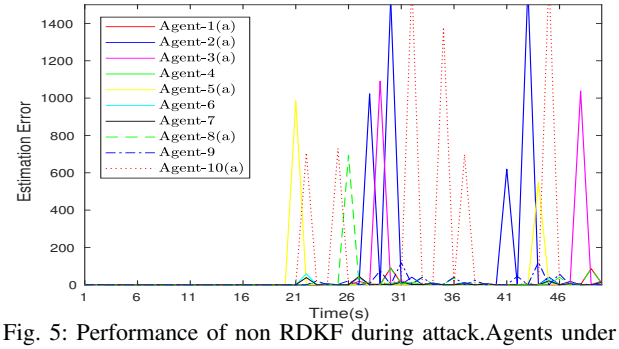

Fig. 5: Performance of non RDKF during attack.Agents under attack are marked with (a). Estimation error is large.

[15] J. Hespanha, *Linear Systems Theory: Second Edition*. Princeton University Press, 2018.

Appendix

Proof of Theorem III.1

In the proof, both $\mathbf{A} \succcurlyeq \mathbf{B}$ and $\mathbf{B} \preccurlyeq \mathbf{A}$ means that $\mathbf{A} - \mathbf{B}$ is positive semidefinite.

Here, we let $\mathbf{P}_0^{(i)} = \mathbf{0}$ for all $1 \leq i \leq n$ and show that for the sequence $\mathbf{P}_k^{(i)}$ generated by $\mathbf{P}_{k|k-1}^{(i)} = \mathbf{A}\mathbf{P}_{k-1}^{(i)}\mathbf{A}^T + \Sigma_w^{(i)}$ and $\mathbf{P}_k^{(i)} = \left(\frac{1}{d_i}\sum_{j\in N^{(i)}} \mathbf{P}_{k|k-1}^{(j)\,-1} + \mathbf{C}^{(i)\,T}\Sigma_v^{(i)\,-1}\mathbf{C}^{(i)}\right)^{-1}$, the limit, $\lim_{k\to\infty} \mathbf{P}_k^{(i)}$, exist and it is a positive definite

matrix for all $1 \leq i \leq n$. If this is true, then $\mathbf{P}^{(i)} = \lim_{k\to\infty} \mathbf{P}_k^{(i)}$ is a solution of equation (5).

We will first show that $\mathbf{P}_1^{(i)\,-1}$ is bounded below by a positive definite matrix. For $k = 1$, we have $\mathbf{P}_1^{(i)\,-1} \succcurlyeq \mathbf{C}^{(i)T}\Sigma_v^{(i)\,-1}\mathbf{C}^{(i)}$, which is positive semidefinite with range being the row space of $\mathbf{C}^{(i)}$, i.e., $\{\mathbf{C}^{(i)\,T}\mathbf{z} : \mathbf{z} \in \mathbb{R}^q\}$.

If $j \in N^{(i)}$ (and by definition $i \in N^{(i)}$), then $\mathbf{P}_2^{(i)\,-1} \succcurlyeq \frac{1}{d_i}(\mathbf{A}\mathbf{P}_1^{(j)}\mathbf{A}^T + \Sigma_w^{(j)})^{-1} + \mathbf{C}^{(i)T}\Sigma_v^{(i)\,-1}\mathbf{C}^{(i)} + \frac{1}{d_i}(\mathbf{A}\mathbf{P}_1^{(i)}\mathbf{A}^T + \Sigma_w^{(i)})^{-1}$, which is a positive semidefinite matrix of range $\{\mathbf{C}^{(i)\,T}\mathbf{z}_1 + \mathbf{A}\mathbf{C}^{(i)\,T}\mathbf{z}_2 + \mathbf{A}\mathbf{C}^{(i)\,T}\mathbf{z}_3 : \mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3 \in \mathbb{R}^q\}$. By applying the same procedure to time $k = 3, 4, \cdots$, we verify that for sufficiently large k, the range of $\mathbf{P}_k^{(i)\,-1}$ can be given by the linear combination of $\oplus_j\{\mathbf{A}^{l(j)}\mathbf{C}^{(j)T}\mathbf{z}\}$ $\forall j$ such that there exists a path from j to i of length $l(j)$. It can be shown that for sufficiently large k, the set contains the range of $\mathbf{A}^r\mathbf{C}^T, \mathbf{A}^{r+1}\mathbf{C}^T, \mathbf{A}^{r+2}\mathbf{C}^T, \cdots$ for some positive integer r. When $\mathbf{A}$ is full-rank and $(\mathbf{A}, \mathbf{C})$ is observable, this range is $\mathbb{R}^n$ and as a result, $\mathbf{P}_k^{(i)\,-1}$ is larger than a positive definite matrix with full rank. This suggests that $\mathbf{P}_k^{(i)}$ is bounded by a positive definite matrix from above.

In addition, by induction it can be shown that $\mathbf{P}_k^{(i)}$ is strictly increasing in the sense that $\mathbf{P}_0^{(i)} \preccurlyeq \mathbf{P}_1^{(i)} \preccurlyeq \mathbf{P}_2^{(i)} \preccurlyeq \cdots$. Since, the sequence is bounded above, its limit exist. In addition, $\mathbf{P}_k^{(i)} \succcurlyeq \left(\frac{1}{d_i}\sum_{j\in N^{(i)}} \Sigma_w^{(j)\,-1} + \mathbf{C}_i^T\Sigma_v^{(i)\,-1}\mathbf{C}_i\right)^{-1}$ Thus, its limit is positive definite. ■

Proof of Theorem III.2

We specify the estimation error as $\mathbf{e}_k^{(i)} = \hat{\mathbf{x}}_k^{(i)} - \mathbf{x}_k$ and show that,

$$\mathbf{e}_k^{(i)\,T}\mathbf{P}^{(i)\,-1}\mathbf{e}_k^{(i)} \leq \frac{1}{d_i}\sum_{j\in N^{(i)}} \mathbf{e}_{k-1}^{(j)\,T}\mathbf{P}^{(j)\,-1}\mathbf{e}_{k-1}^{(j)}. \quad (11)$$

Applying equation (9), we have $\mathbf{e}_k^{(i)} = \arg\min_{\mathbf{e}_k} f(\mathbf{e}_k)$, where

$$f(\mathbf{e}_k) = (\mathbf{C}^{(i)}\mathbf{e}_k)^T\Sigma_v^{(i)\,-1}(\mathbf{C}^{(i)}\mathbf{e}_k) + \frac{1}{d_i}\sum_{j\in N^{(i)}} (\mathbf{e}_k - \mathbf{A}\mathbf{e}_{k-1}^{(j)})^T\mathbf{P}_{|}^{(j)\,-1}(\mathbf{e}_k - \mathbf{A}\mathbf{e}_{k-1}^{(j)}) \quad (12)$$

Using the fact that $\nabla f(\mathbf{e}_k)|_{\mathbf{e}_k=\mathbf{e}_k^{(i)}} = 0$, we have

$$(\mathbf{C}^{(i)}\mathbf{e}_k^{(i)})^T\Sigma_v^{(i)\,-1}(\mathbf{C}^{(i)}\mathbf{e}_k^{(i)}) + \frac{1}{d_i}\sum_{j\in N^{(i)}} (\mathbf{e}_k^{(i)} - \mathbf{A}\mathbf{e}_{k-1}^{(j)})^T\mathbf{P}_{|}^{(j)\,-1}\mathbf{e}_k^{(i)} = 0. \quad (13)$$

Combining equation (13) with $f(\mathbf{e}_k^{(i)}) \geq 0$ gives,

$$\frac{1}{d_i}\sum_{j\in N^{(i)}} (\mathbf{A}\mathbf{e}_{k-1}^{(j)})^T\mathbf{P}_{|}^{(j)\,-1}\mathbf{A}\mathbf{e}_{k-1}^{(j)} \geq (\mathbf{C}^{(i)}\mathbf{e}_k^{(i)})^T\Sigma_v^{(i)\,-1}(\mathbf{C}^{(i)}\mathbf{e}_k^{(i)}) + \frac{1}{d_i}\sum_{j\in N^{(i)}} \mathbf{e}_k^{(i)\,T}\mathbf{P}_{|}^{(j)\,-1}\mathbf{e}_k^{(i)} = \mathbf{e}_k^{(i)\,T}\mathbf{P}^{(i)\,-1}\mathbf{e}_k^{(i)}. \quad (14)$$

Since, $\mathbf{P}^{(i)\,-1} - \mathbf{A}^T\mathbf{P}_{|}^{(j)\,-1}\mathbf{A} = \mathbf{P}^{(i)\,-1} - \mathbf{A}^T(\mathbf{A}\mathbf{P}^{(i)}\mathbf{A}^T + \Sigma_w^{(i)})^{-1}\mathbf{A} = \mathbf{P}^{(i)\,-1} - (\mathbf{P}^{(i)} + \mathbf{A}^{-1}\Sigma_w^{(i)}\mathbf{A}^{-1\,T})^{-1}$ is positive semidefinite, equation (14) implies equation (11), and equation (11) implies that $\max_{1\leq i\leq n} \mathbf{e}_k^{(i)}\mathbf{P}^{(i)\,-1}\mathbf{e}_k^{(i)}$ does not increase as a function of k and thus, it converges. However, it remains to be proven that it converges to zero. If this is not the case, then equation (14) achieves the equality $(\mathbf{A}\mathbf{e}_{k-1}^{(j)})^T\mathbf{P}_{|}^{(j)\,-1}\mathbf{A}\mathbf{e}_{k-1}^{(j)} = \mathbf{e}_k^{(j)\,T}\mathbf{P}^{(j)\,-1}\mathbf{e}_k^{(j)}$ and it implies that $\mathbf{e}_{k-1}^{(j)} = 0$ for all $j \in N^{(i)}$. Combining it with $\mathbf{A}\mathbf{e}_{k-1}^{(j)} = \mathbf{e}_k^{(i)}$ (which follows from the equality $f(\mathbf{e}_k^{(i)}) = 0$), we get $\mathbf{e}_k^{(i)} = 0$. This also suggests that the ratio of the two sides of equation (11) is strictly less than 1, implying that $\max_{1\leq i\leq n} \mathbf{e}_k^{(i)}\mathbf{P}^{(i)\,-1}\mathbf{e}_k^{(i)}$ converges linearly. ■

Proof of Theorem III.3

We follow the proof of Theorem III.2 and differentiate the objective function of equation (8), which gives us

$$\lambda\frac{(\mathbf{C}^{(i)}\mathbf{e}_k^{(i)})^T\Sigma_v^{(i)\,-1}(\mathbf{C}^{(i)}\mathbf{e}_k^{(i)})}{\|\Sigma_v^{(i)\,-\frac{1}{2}}(\mathbf{C}^{(i)}\mathbf{e}_k^{(i)})\|} + \frac{1}{d_i}\sum_{j\in N^{(i)}} (\mathbf{e}_k^{(i)} - \mathbf{A}\mathbf{e}_{k-1}^{(j)})^T\mathbf{P}_{|}^{(j)\,-1}\mathbf{e}_k^{(i)} = 0. \quad (15)$$

As a result, $\frac{1}{d_i}\sum_{j\in N^{(i)}}(\mathbf{e}_k^{(i)} - \mathbf{A}\mathbf{e}_{k-1}^{(j)})^T\mathbf{P}_{|}^{(j)\,-1}(\mathbf{e}_k^{(i)} - \mathbf{A}\mathbf{e}_{k-1}^{(j)}) \geq 0$ implies that $\frac{1}{d_i}\sum_{j\in N^{(i)}}(\mathbf{A}\mathbf{e}_{k-1}^{(j)})^T\mathbf{P}_{|}^{(j)\,-1}\mathbf{A}\mathbf{e}_{k-1}^{(j)} \geq 2\lambda\frac{(\mathbf{C}^{(i)}\mathbf{e}_k^{(i)})^T\Sigma_v^{(i)\,-1}(\mathbf{C}^{(i)}\mathbf{e}_k^{(i)})}{\|\Sigma_v^{(i)\,-\frac{1}{2}}(\mathbf{C}^{(i)}\mathbf{e}_k^{(i)})\|} + \frac{1}{d_i}\sum_{j\in N^{(i)}} \mathbf{e}_k^{(i)\,T}\mathbf{P}_{|}^{(j)\,-1}\mathbf{e}_k^{(i)} = \mathbf{e}_k^{(i)\,T}\mathbf{P}^{(i)\,-1}\mathbf{e}_k^{(i)}$, if $\|\Sigma_v^{(i)\,-\frac{1}{2}}(\mathbf{C}^{(i)}\mathbf{e}_k^{(i)})\| \leq 2\lambda$.

Using the assumption about the initial estimation errors $\mathbf{e}_i^{(0)}$ in Theorem III.3, it can be proved that $\max_{1\leq j\leq n}(\mathbf{A}\mathbf{e}_{k-1}^{(j)})^T\mathbf{P}_{|}^{(j)\,-1}\mathbf{A}\mathbf{e}_{k-1}^{(j)}$ is decreasing and following the proof of Theorem III.2, it converges to zero linearly. ■

Proof of Theorem III.4

First we introduce the following lemma.

Lemma .1. When $\mathbf{A}$ is a square matrix and $\mathbf{Q}$ is positive definite, then the minimizer of $\mathbf{x}^T\mathbf{Q}\mathbf{x} + \lambda\|\mathbf{A}\mathbf{x} - \mathbf{a}\|$, $\hat{\mathbf{x}}$, satisfies $\|\hat{\mathbf{x}}\| \leq \frac{\lambda}{2}\|\mathbf{Q}^{-1}\mathbf{A}\|$.

Proof. The gradient of the objective function $\mathbf{x}^T\mathbf{Q}\mathbf{x} + \lambda\|\mathbf{A}\mathbf{x} - \mathbf{a}\|$ of the minimizer should be zero, i.e. $2\mathbf{Q}\hat{\mathbf{x}} + \lambda\mathbf{A}\frac{\mathbf{A}\hat{\mathbf{x}}-\mathbf{a}}{\|\mathbf{A}\hat{\mathbf{x}}-\mathbf{a}\|} = 0$. So, $\hat{\mathbf{x}} = -\frac{\lambda}{2}\mathbf{Q}^{-1}\mathbf{A}\frac{\mathbf{A}\hat{\mathbf{x}}-\mathbf{a}}{\|\mathbf{A}\hat{\mathbf{x}}-\mathbf{a}\|}$ and $\|\hat{\mathbf{x}}\| \leq \frac{\lambda}{2}\|\mathbf{Q}^{-1}\mathbf{A}\|$. □

Based on Lemma .1, we have the following result:

For any $\mathbf{a}_1$, $\mathbf{a}_2$, the minimizers of $(\mathbf{x}-\mathbf{x}_0)^T\mathbf{Q}(\mathbf{x}-\mathbf{x}_0) + \lambda\|\mathbf{A}\mathbf{x} - \mathbf{a}_1\|$ and $(\mathbf{x}-\mathbf{x}_0)^T\mathbf{Q}(\mathbf{x}-\mathbf{x}_0) + \lambda\|\mathbf{A}\mathbf{x} - \mathbf{a}_2\|$ are at most $\lambda\|\mathbf{Q}^{-1}\mathbf{A}\|$ apart.

Now, we can prove the theorem. Note that the optimization problem of equation (8) leads to: $\hat{\mathbf{e}}_k^{(i)} = \arg\min_{\mathbf{e}_k} \lambda\left\|\Sigma_v^{(i)\,-\frac{1}{2}}(\mathbf{a}_k^{(i)} - \mathbf{C}^{(i)}\mathbf{e}_k)\right\| + \frac{1}{d_i}\sum_{j\in N^{(i)}}(\mathbf{e}_k - \mathbf{A}\mathbf{e}_{k-1}^{(j)})^T\mathbf{P}_{|}^{(j)\,-1}(\mathbf{e}_k - \mathbf{A}\mathbf{e}_{k-1}^{(j)})$. Thus, even for different attack vectors $\mathbf{a}_k^{(i)}$, the difference of their solutions are bounded above by equation (10). ■