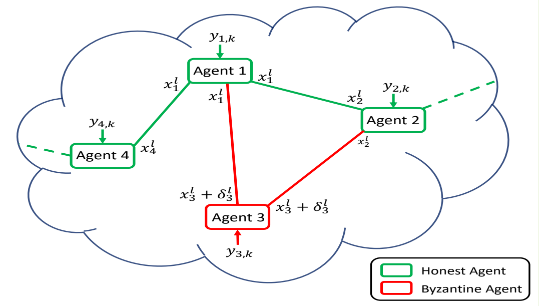
# Total Variation based Distributed Kalman Filtering for Resiliency Against Byzantines

Ashkan Moradi[1], Naveen K. D. Venkategowda[2], *Member, IEEE*, Stefan Werner[1], *Fellow, IEEE*

***Abstract*— This paper proposes a distributed Kalman filter (DKF) with enhanced robustness against Byzantine adversaries. A Byzantine agent is a legitimate network agent that, unlike an honest agent, manipulates information before sharing it with neighbors to impair the overall system performance. In contrast to the literature, the DKF is modeled as a distributed optimization problem where resiliency against Byzantine agents is accomplished by employing a total variation (TV) penalty term. We utilize a distributed subgradient algorithm to compute the state estimate and error covariance matrix updates of the DKF. Additionally, we prove that the proposed suboptimal solution converges to a neighborhood of the optimal centralized solution of the Kalman filter (KF) with a bounded radius when Byzantine agents are present. Numerical simulations corroborate the theoretical findings and demonstrate the robustness of the proposed DKF against Byzantine attacks.**



***Index Terms*— Multiagent network, Kalman filtering, Distributed optimization, Byzantine attack, attack robustness**

## I. Introduction

DISTRIBUTED filtering techniques have found widespread use in diverse applications such as environmental monitoring, smart grids, and state estimation [1]–[4]. Due to the lack of a fusion center in distributed Kalman filtering scenarios, agents rely on local interactions to complete a common task across the network [5], [6]. However, the local collaboration renders distributed Kalman filtering susceptible to security attacks.

Attacks on multi-agent networks can be classified as either active or passive; for example, a passive attack can be an eavesdropper intercepting a communication link between agents in order to obtain information [7]. On the other hand, active attacks include denial-of-service attacks (DoS) and data falsification attacks. During DoS attacks, agents cannot exchange information due to link blockages [8]. In contrast, in data falsification attacks, false information is injected into the network [9] by either external adversaries or malicious agents, also termed Byzantine agents, to degrade the overall system performance. Data falsification attacks can be performed independently by each Byzantine agent or designed cooperatively in order to maximize system degradation [9].

Data falsification attacks, in general, have been extensively studied to analyze the impact of malicious behaviors on distributed filtering and estimation [10]–[18]. One approach to reducing the impact of malicious adversaries on the network performance is to detect them and counteract their actions by implementing correction measures [19]–[21]. For example, [22] proposed a defense strategy for a distributed recursive filter by detecting adversarial attacks based on changes in innovation signals of agents and redesigning their gains. Several studies have been proposed in the literature to design an optimal data falsification attack from the perspective of an adversary that evades detection [23]–[26]. For example, the authors in [23] and [24] propose stealthy linear data falsification attacks in remote state estimation scenarios assuming K-L divergence and $\chi^2$ detectors, respectively. Furthermore, the integrity attack also includes stealthy attack strategies, which inject false data into the network without being detected [27]–[29]. In contrast, [25], [26] mainly focus on designing attacks to ensure that the probability of detection does not exceed a given threshold. These have shown that relying on attack detection to limit the impact of adversaries has limited utility in the presence of stealth attacks. Hence, there is a need for a robust algorithm that can operate effectively even when unidentified attacks occur [30].

To that end, works in [31], [32] propose using the statistics of innovation signals to re-design the consensus weights of agents in distributed signal detection and filtering scenarios to minimize the impact of Byzantine agents. A Byzantine-resilient distributed state estimation algorithm is proposed in [33], which allows agents to update state estimates locally by selecting the best subset of neighbors to be effective in updating the state estimate. To reduce computational resources, in [34], [35], distributed state estimation approaches provide

*This work was supported in part by the Research Council of Norway.

[1] A. Moradi and S. Werner are with the Department of Electronic Systems, Norwegian University of Science and Technology, Trondheim, Norway, E-mail: {ashkan.moradi, stefan.werner}@ntnu.no. S. Werner is also with the Department of Signal Processing and Acoustics, Aalto University, Finland

[2] N. K. D. Venkategowda is with Linköping University, Norrköping, Sweden, E-mail: naveen.venkategowda@liu.se.

resilience against measurement attacks by assigning adaptive weights to received measurements from neighbors. By assigning smaller weights to measurements whose norm exceeds a certain threshold, they would have a smaller impact on updating state estimates. The studies in [36], [37] investigate the problem of multi-sensor estimation under undetectable attacks. From the perspective of an adversary, authors in [36], [37] design the attack to maximize the estimation error of the network. Moreover, the gains of the estimator are re-designed in order to mitigate the impact of the designed optimal attack. In addition, a secure state estimation strategy with triple-loop local state observers is proposed in [38], while in [39], the secure state estimation problem is solved by a local observer that achieves robustness against sensor attacks by employing the median of its local estimates.

Homomorphic encryption schemes have been proposed to further ensure the confidentiality of the signals sent over the network [40]. In [41], the authors propose employing additively homomorphic encryption, which enables the cloud server and security module to integrate the information of multiple parties while maintaining data privacy. However, the authors in [42] propose a modified encoding and decoding scheme that, unlike the previous work in [43], does not negatively affect estimation performance in the absence of attacks and further protects data integrity in multi-sensor networks. Moreover, utilizing randomization-based methods to disrupt and mislead attackers in their malicious activities is a less resource-intensive method to mitigate the impact of adversarial attacks in the network [44]. Furthermore, to improve network resistance in the presence of adversarial attacks, [45], [46] introduced a redundancy-based approach for CPSs at different levels of communication, channels, software, and hardware. Redundant subsystems serve as backups or parallel integrity verification units to reduce the effect of malfunctioning behaviors in the network [47]. However, an approach based on redundancy demands strict network requirements and can only tolerate a few Byzantine adversaries. Accordingly, the authors in [48] reduce the stringent requirements of redundancy to only a group of agents and make them resistant to attacks. Generally, these approaches reduce the impact of adversarial attacks on the network. Still, they require more local computations and information transfer in the network, which is undesirable in resource-constrained situations.

The Kalman filtering algorithm has been modeled as an optimization problem. However, this optimization-based approach has not been analyzed in adversarial situations or adapted for robustness in the presence of Byzantine agents. Therefore, contrary to the literature, we propose a distributed Kalman filtering algorithm modeled as an optimization problem with total variation-based constraints that provides robustness to coordinated Byzantine attacks. First, we design the filtering algorithm by adapting the framework proposed in [49] to model the Kalman filtering operations as a solution to an optimization problem and using the TV-norm penalty in the objective function to enforce resiliency against data-falsification attacks in [50]–[52]. Then, we solve the TV-norm-penalized optimization problem using a distributed subgradient algorithm that updates the state estimate for all agents through local collaborations. Furthermore, we model the error covariance update of agents as a TV-norm-penalized optimization problem, which is solved by a similar subgradient approach in the presence of Byzantine agents. Moreover, we show that the proposed TV-norm penalized optimization problem corresponding to the state estimate update results in the same solution as the centralized Kalman filter (CKF). In addition, in the presence of Byzantine agents, we show that the proposed suboptimal solution for the state estimate update, obtained by the subgradient algorithm, converges to a bounded neighborhood of the optimal solution. Finally, we provide numerical simulations to demonstrate the resiliency against Byzantine behavior by obtaining lower filtering mean square error (MSE).

The remainder of this article is organized as follows. Section II presents the problem formulation and provides background information. Section III proposes a DKF with a TV-norm penalized objective function that is robust against Byzantine agents. Section IV presents the convergence of the proposed TV-norm-penalized distributed optimization problem to a bounded neighborhood of the CKF solution. Finally, numerical results are provided in Section V to demonstrate the resiliency against Byzantines, and Section VI concludes the article.

***Mathematical Notation:*** Scalars are denoted by lowercase letters, column vectors by bold lowercase, and matrices by bold uppercase. Superscripts $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{-1}$ denote the transpose and inverse operators, respectively. The symbol $\mathbf{1}_m$ represents the $m \times 1$ column vector with all entries equal to one, and $\mathbf{I}_m$ is the $m \times m$ identity matrix. The trace operator is denoted as $\mathsf{tr}(\cdot)$, whereas the greater than and less than symbols in the scalar inequalities are represented by $>$ and $<$, respectively. A positive semidefinite matrix $\mathbf{A}$ is denoted by $\mathbf{A} \succcurlyeq 0$ and $\mathbf{A} \succcurlyeq \mathbf{B}$ indicates that $\mathbf{A} - \mathbf{B}$ is a positive semidefinite matrix. The element-wise sign function is represented by $\mathsf{sign}(\cdot)$ where given $x > 0$, $\mathsf{sign}(x) = 1$ and $\mathsf{sign}(x) = -1$ when $x < 0$. In case of $x = 0$, the value of $\mathsf{sign}(x)$ can be any arbitrary value within $[-1, 1]$. The half vectorization of a symmetric matrix $\mathbf{M} \in \mathbb{R}^{m \times m}$is denoted by $\mathsf{vec}_h(\mathbf{M}) \in \mathbb{R}^{m(m+1)/2}$, where $\mathsf{vec}_h(\mathbf{M}) = [M_{1,1}, \cdots, M_{1,m}, M_{2,2}, \cdots, M_{2,m}, \cdots, M_{m,m}]^{\mathrm{T}}$ with $M_{ij}$ as the $ij$th element of $\mathbf{M}$. The operator of $\mathsf{vec}_h^{-1}(\cdot)$ denotes the inverse function of $\mathsf{vec}_h(\cdot)$, i.e., $\mathsf{vec}_h^{-1}(\mathsf{vec}_h(\mathbf{M})) = \mathbf{M}$. The stacked vector $\mathbf{x} = [\mathbf{a}]_{i=1}^{N} \in \mathbb{R}^{Nm}$ corresponds to $N$ times stacking the smaller vector $\mathbf{a} \in \mathbb{R}^m$ together.

## II. Background and Problem Formulation

Consider a network modeled as a connected graph $\mathcal{G} = \{\mathcal{N}, \mathcal{E}\}$, where the node set $\mathcal{N}$ represents agents of the network and $\mathcal{E}$ is the set of edges that represent communication links between agents, i.e., $(i, j) \in \mathcal{E}$ if nodes $i$ and $j$ are connected. Additionally, the set $\mathcal{N}_i$ specifies the neighborhood of node $i$ and does not include the node itself. The cardinality of the set $\mathcal{N}_i$ is denoted by $|\mathcal{N}_i|$, while $N = |\mathcal{N}|$ is the number of agents in the network.

### A. Distributed Kalman Filter (DKF)

We revisit the DKF problem that is modeled as a maximum likelihood estimation problem and represents the relationship between a KF [5] and an optimization problem [49]. The state-space model characterizes the state vector evolution and observation vectors and is given by

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

$$\mathbf{y}_{i,k} = \mathbf{H}_i\mathbf{x}_k + \mathbf{v}_{i,k} \tag{2}$$

where for time instant $k$, $\mathbf{F} \in \mathbb{R}^{m\times m}$ denotes the state transition matrix, $\mathbf{H} = [\mathbf{H}_1^{\mathrm{T}}, \cdots, \mathbf{H}_N^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{Nn\times m}$ denotes the network observation matrix, $\mathbf{y}_k = [\mathbf{y}_{1,k}^{\mathrm{T}}, \cdots, \mathbf{y}_{N,k}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{Nn}$ is the network observation vector, and $\mathbf{w}_k \in \mathbb{R}^{m}$ and $\mathbf{v}_k = [\mathbf{v}_{1,k}^{\mathrm{T}}, \cdots, \mathbf{v}_{N,k}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{Nn}$, are process and observation noises, respectively. The process noise $\mathbf{w}_k$ and observation noise $\mathbf{v}_k$ are zero-mean white Gaussian noise processes with a covariance matrices $\mathbf{Q} \in \mathbb{R}^{m\times m}$ and $\mathbf{R} \in \mathbb{R}^{Nn\times Nn}$, respectively, where $\mathbf{R} \triangleq \mathsf{diag}(\{\mathbf{R}_i\}_{i=1}^N)$ and $\mathbf{R}_i = \mathbb{E}\{\mathbf{v}_{i,k}\mathbf{v}_{i,k}^{\mathrm{T}}\} \in \mathbb{R}^{n\times n}$. We assume that the pair $(\mathbf{F}, \mathbf{H})$ is observable and observation noise sequences are uncorrelated. Every agent estimates the state of the network by processing its local and neighboring information. A local estimate for each agent must be provided in a way that the local mean squared error of the agent is minimized.

### B. Byzantine Attack Strategy

We assume a distributed setting in which a subset of agents $\mathcal{B}$ are Byzantines, i.e., $\mathcal{B} \subset \mathcal{N}$, and unlike honest agents, they share the manipulated version of their local estimates. In order to update the *a posteriori* state estimate, agents need information exchange with neighbors; we, therefore, assume that Byzantine agents falsify their state estimate before sharing it with neighbors at each iteration. The shared state estimate can be modeled as

$$\tilde{\mathbf{x}}_{i,k}^l = \begin{cases} \mathbf{x}_{i,k}^l + \boldsymbol{\delta}_i^l & i \in \mathcal{B} \\ \mathbf{x}_{i,k}^l & i \notin \mathcal{B} \end{cases} \tag{3}$$

where at agent $i$ and iteration $l$, $\mathbf{x}_{i,k}^l$ denotes the state estimate and $\boldsymbol{\delta}_i^l \in \mathbb{R}^m$ is the perturbation sequence of the Byzantine agent. To maximize the attack stealthiness, as shown in [53], [54], we consider the perturbation sequence to be a zero-mean Gaussian sequence with covariance matrix $\boldsymbol{\Sigma}_i \in \mathbb{R}^{m\times m}$. Moreover, in order to maximize the damage caused by the Byzantine attack, we assume that Byzantines design a coordinated attack with covariance matrix $\boldsymbol{\Sigma} = \mathbb{E}\{\boldsymbol{\delta}^l(\boldsymbol{\delta}^l)^{\mathrm{T}}\} \in \mathbb{R}^{Nm\times Nm}$ where $\boldsymbol{\delta}^l = [(\boldsymbol{\delta}_1^l)^{\mathrm{T}}, \cdots, (\boldsymbol{\delta}_N^l)^{\mathrm{T}}]^{\mathrm{T}}$ is the network-wide perturbation sequence and $\boldsymbol{\delta}_i^l = \mathbf{0}$ if $i \notin \mathcal{B}$.

## III. Byzantine-Robust Distributed Kalman Filter (BR-DKF)

We consider a network of $N$ agents and assume each agent runs a local KF without sending information to a fusion center. Instead, agents exchange information with their neighbors to develop their optimal estimates. The communication network is considered as graph $\mathcal{G}$ with adjacency and Laplacian matrices $\mathbf{E}$ and $\mathbf{L}$, respectively. Each agent $i \in \mathcal{N}$ updates its local estimate by employing the local observation vector in (2). Similar to the centralized case in [49], the DKF also requires two steps of prediction and correction, where for each agent $i$ and time instant $k$, the prediction updates are modeled as

$$\hat{\mathbf{x}}_{i,k|k-1} = \mathbf{F}\hat{\mathbf{x}}_{i,k-1} \tag{4}$$

$$\mathbf{P}_{i,k|k-1} = \mathbf{F}\mathbf{P}_{i,k-1}\mathbf{F}^{\mathrm{T}} + \mathbf{Q} \tag{5}$$

with $\hat{\mathbf{x}}_{i,k-1}$ and $\mathbf{P}_{i,k-1} = \mathbb{E}\{\mathbf{e}_{i,k-1}\mathbf{e}_{i,k-1}^{\mathrm{T}}\}$ being the state estimate and error covariance matrix at time instant $k-1$, and $\mathbf{e}_{i,k-1} = \mathbf{x}_{k-1} - \hat{\mathbf{x}}_{i,k-1}$. The *a priori* state estimate and error covariance are denoted by $\hat{\mathbf{x}}_{i,k|k-1}$ and $\mathbf{P}_{i,k|k-1} = \mathbb{E}\{\mathbf{e}_{i,k|k-1}\mathbf{e}_{i,k|k-1}^{\mathrm{T}}\}$, respectively, with $\mathbf{e}_{i,k|k-1} = \mathbf{x}_k - \hat{\mathbf{x}}_{i,k|k-1}$.

The correction steps of the DKF can be modeled as the solution of a constrained optimization problem [49]; in particular, the *a posteriori* state estimates can be obtained by solving the optimization problem

$$\begin{aligned} &\min_{\{\mathbf{x}_{i,k}\}_{i=1}^N} \sum_{i=1}^N f_i(\mathbf{x}_{i,k}) \\ &\quad \text{s. t.} \;\; \mathbf{x}_{i,k} = \mathbf{x}_{j,k}, \;\; \forall j \in \mathcal{N}_i, i = 1, 2, \cdots, N \end{aligned} \tag{6}$$

where the local objective function $f_i(\mathbf{x}_{i,k})$ is given by

$$\begin{aligned} f_i(\mathbf{x}_{i,k}) = &\frac{1}{2}\Big( (\mathbf{y}_{i,k} - \mathbf{H}_i\mathbf{x}_{i,k})^{\mathrm{T}}\mathbf{R}_i^{-1}(\mathbf{y}_{i,k} - \mathbf{H}_i\mathbf{x}_{i,k}) \\ &+ \frac{1}{N}(\mathbf{x}_{i,k} - \hat{\mathbf{x}}_{i,k|k-1})^{\mathrm{T}}\mathbf{P}_{i,k|k-1}^{-1}(\mathbf{x}_{i,k} - \hat{\mathbf{x}}_{i,k|k-1}) \Big) \end{aligned} \tag{7}$$

and the constraints enforce consensus across all the agents in the network. The distributed Kalman filtering problem can be solved by any distributed algorithm that finds the optimal solutions in (6), i.e., $\mathbf{x}_{i,k}^*$ for each $i \in \mathcal{N}$. Subsequently, the *a posteriori* state estimates of agents are obtained as $\hat{\mathbf{x}}_k = [\hat{\mathbf{x}}_{1,k}^{\mathrm{T}}, \cdots, \hat{\mathbf{x}}_{N,k}^{\mathrm{T}}]^{\mathrm{T}}$ where $\hat{\mathbf{x}}_{i,k} = \mathbf{x}_{i,k}^*$.

Motivated by [50], [51], the constraints in (6) can be approximated by a TV-norm penalty which also endows robustness to data falsification attacks. In the absence of a Byzantine agent in the network, the TV-norm-penalized problem of (6) can be formulated as

$$\mathbf{x}_{c_k}^* = \min_{\{\mathbf{x}_{i,k}\}_{i=1}^N} \sum_{i=1}^N \left( f_i(\mathbf{x}_{i,k}) + \frac{\lambda_{\text{tv}}}{2} \sum_{j\in\mathcal{N}_i} \|\mathbf{x}_{i,k} - \mathbf{x}_{j,k}\|_1 \right) \tag{8}$$

where $\mathbf{x}_{c_k}^* = [(\mathbf{x}_{1,k}^*)^{\mathrm{T}}, \cdots, (\mathbf{x}_{N,k}^*)^{\mathrm{T}}]^{\mathrm{T}}$ and $\lambda_{\text{tv}}$ is a penalty parameter. Due to the penalty parameter $\lambda_{\text{tv}}$, estimates $\mathbf{x}_{i,k}$ and $\mathbf{x}_{j,k}$ are forced to be close. The larger the $\lambda_{\text{tv}}$, the closer $\mathbf{x}_{i,k}$ and $\mathbf{x}_{j,k}$ become. However, the TV-norm penalty allows for some pairs of $\mathbf{x}_{i,k}$ and $\mathbf{x}_{j,k}$ to be different, which is crucial when Byzantine agents are present in the network.

We solve the optimization problem in (8) with a subgradient method [51], and derive the state estimate update at each agent $i \in \mathcal{N}$ as

$$\mathbf{x}_{i,k}^{l+1} = \mathbf{x}_{i,k}^l - \alpha_k \left( \nabla_{\mathbf{x}_{i,k}} f_i(\mathbf{x}_{i,k}^l) + \lambda_{\text{tv}} \sum_{j\in\mathcal{N}_i} \mathsf{sign}(\mathbf{x}_{i,k}^l - \mathbf{x}_{j,k}^l) \right) \tag{9}$$

where $\alpha_k > 0$ denotes the step size and $\mathbf{x}_{i,k}^l$ is the state estimate of the subgradient method at agent $i$ and iteration

$l$. Assuming that a group of agents is conducting Byzantine attacks on the network, i.e., $\mathcal{B} \subset \mathcal{N}$, and by substituting the gradient $\nabla_{\mathbf{x}_{i,k}} f_i(\mathbf{x}_{i,k})$, we obtain

$$\mathbf{x}_{i,k}^{l+1} = \mathbf{x}_{i,k}^{l} - \alpha_k \Bigg( \boldsymbol{\Omega}_{i,k}\mathbf{x}_{i,k}^{l} - \boldsymbol{\theta}_{i,k} + \lambda_{\text{tv}} \sum_{j\in\mathcal{R}_i} \mathsf{sign}(\mathbf{x}_{i,k}^{l} - \mathbf{x}_{j,k}^{l}) + \lambda_{\text{tv}} \sum_{j\in\mathcal{B}_i} \mathsf{sign}(\mathbf{x}_{i,k}^{l} - \tilde{\mathbf{x}}_{j,k}^{l}) \Bigg) \tag{10}$$

where $\tilde{\mathbf{x}}_{j,k}^{l} = \mathbf{x}_{j,k}^{l} + \boldsymbol{\delta}_j^l$ is the state estimate received from the $j$th Byzantine neighbor, $\mathcal{R}_i$ and $\mathcal{B}_i$ include honest and Byzantine members of $\mathcal{N}_i$, and

$$\begin{aligned} \boldsymbol{\Omega}_{i,k} &= \mathbf{H}_i^{\mathrm{T}}\mathbf{R}_i^{-1}\mathbf{H}_i + \frac{1}{N}\mathbf{P}_{i,k|k-1}^{-1} \\ \boldsymbol{\theta}_{i,k} &= \mathbf{H}_i^{\mathrm{T}}\mathbf{R}_i^{-1}\mathbf{y}_{i,k} + \frac{1}{N}\boldsymbol{\Omega}_{i,k|k-1}\hat{\mathbf{x}}_{i,k|k-1} \end{aligned} \tag{11}$$

with $\boldsymbol{\Omega}_{i,k|k-1} = \mathbf{P}_{i,k|k-1}^{-1}$. Regardless of the state estimate received from neighbors, the value of $\mathsf{sign}(\mathbf{x}_{i,k}^{l} - \tilde{\mathbf{x}}_{j,k}^{l})$ is restricted to $[-1, 1]$. Thus, the last term in (10) limits the effect of perturbed data received from a Byzantine agent, so that the state estimate update is more resistant to Byzantine attacks.

Similarly, the error covariance update also requires designing an optimization problem to obtain the average consensus of the information matrices $N\boldsymbol{\Omega}_{i,k}$ throughout the network. To this end, we propose the following optimization problem that derives the error covariance update

$$\begin{aligned} &\min_{\{\boldsymbol{\zeta}_i\}_{i=1}^N} \sum_{i=1}^{N} \|\boldsymbol{\zeta}_i - \mathsf{vec}_h(N\boldsymbol{\Omega}_{i,k})\|_F^2 \\ &\text{s. t.} \quad \boldsymbol{\zeta}_i = \boldsymbol{\zeta}_j, \ \ \forall j \in \mathcal{N}_i, i = 1, 2, \cdots, N. \end{aligned} \tag{12}$$

The optimal solution of (12) is denoted by $\boldsymbol{\zeta}^* = [(\boldsymbol{\zeta}_1^*)^{\mathrm{T}}, \cdots, (\boldsymbol{\zeta}_N^*)^{\mathrm{T}}]^{\mathrm{T}}$ which returns the average of $\mathsf{vec}_h(N\boldsymbol{\Omega}_{i,k})$ throughout the entire network. Subsequently, the error covariance matrix can be updated as $\mathbf{P}_{i,k} = (\mathsf{vec}_h^{-1}(\boldsymbol{\zeta}_i^*))^{-1}$. Motivated by the TV-norm-penalized optimization problem in (8), we modify the optimization problem in (12) as

$$\boldsymbol{\zeta}^* = \min_{\{\boldsymbol{\zeta}_i\}_{i=1}^N} \sum_{i=1}^{N} \Bigg( g_i(\boldsymbol{\zeta}_i) + \frac{\lambda_{\text{tv}}}{2} \sum_{j\in\mathcal{N}_i} \|\boldsymbol{\zeta}_i - \boldsymbol{\zeta}_j\|_1 \Bigg) \tag{13}$$

where $g_i(\boldsymbol{\zeta}_i) = \|\boldsymbol{\zeta}_i - \mathsf{vec}_h(N\boldsymbol{\Omega}_{i,k})\|_F^2$. Employing a similar subgradient approach as in (9), results in

$$\boldsymbol{\zeta}_i^{l+1} = \boldsymbol{\zeta}_i^{l} - \gamma_k \Bigg( \nabla_{\boldsymbol{\zeta}_i} g_i(\boldsymbol{\zeta}_i^l) + \lambda_{\text{tv}} \sum_{j\in\mathcal{N}_i} \mathsf{sign}(\boldsymbol{\zeta}_i^l - \boldsymbol{\zeta}_j^l) \Bigg) \tag{14}$$

where $\gamma_k > 0$ denotes the step size and the update equation in (14) is simplified as

$$\boldsymbol{\zeta}_i^{l+1} = \boldsymbol{\zeta}_i^{l} - \gamma_k \Bigg( \boldsymbol{\zeta}_i^l - \mathsf{vec}_h(N\boldsymbol{\Omega}_{i,k}) + \lambda_{\text{tv}} \sum_{j\in\mathcal{N}_i} \mathsf{sign}(\boldsymbol{\zeta}_i^l - \boldsymbol{\zeta}_j^l) \Bigg) \tag{15}$$

After a large enough number of iterations, say $l^*$, the suboptimal solutions in (10) and (15) converge to $(\mathbf{x}_{i,k}^{l^*}, \boldsymbol{\zeta}_i^{l^*})$ and the filtering *a posteriori* state estimate and error covariance matrix can be updated as

**Algorithm 1** Byzantine-Robust DKF (BR-DKF)

```
• For each agent $i \in \mathcal{N}$
• Initialize $\hat{\mathbf{x}}_{i,0}$ and $\mathbf{P}_{i,0}$
for all $k > 0$ do
    $\hat{\mathbf{x}}_{i,k|k-1} = \mathbf{F}\hat{\mathbf{x}}_{i,k-1}$
    $\mathbf{P}_{i,k|k-1} = \mathbf{F}\mathbf{P}_{i,k-1}\mathbf{F}^{\mathrm{T}} + \mathbf{Q}$
    $\boldsymbol{\Omega}_{i,k|k-1} = \mathbf{P}_{i,k|k-1}^{-1}$
    $\boldsymbol{\Omega}_{i,k} = \mathbf{H}_i^{\mathrm{T}}\mathbf{R}_i^{-1}\mathbf{H}_i + \frac{1}{N}\boldsymbol{\Omega}_{i,k|k-1}$
    $\boldsymbol{\theta}_{i,k} = \mathbf{H}_i^{\mathrm{T}}\mathbf{R}_i^{-1}\mathbf{y}_{i,k} + \frac{1}{N}\boldsymbol{\Omega}_{i,k|k-1}\hat{\mathbf{x}}_{i,k|k-1}$
    Set $\mathbf{x}_{i,k}^1 = \mathbf{0}$ and $\boldsymbol{\zeta}_i^1 = \mathbf{0}$
    for $l = 1$ to $l^*$ do
        Share $\mathbf{x}_{i,k}^l + \boldsymbol{\delta}_i^l$ with neighbors if $i \in \mathcal{B}$
        $\mathbf{x}_{i,k}^{l+1} = \mathbf{x}_{i,k}^l - \alpha_k \left(\boldsymbol{\Omega}_{i,k}\mathbf{x}_{i,k}^l - \boldsymbol{\theta}_{i,k} + \lambda_{\text{tv}} \sum_{j\in\mathcal{N}_i} \mathsf{sign}(\mathbf{x}_{i,k}^l - \tilde{\mathbf{x}}_{j,k}^l)\right)$
        $\boldsymbol{\zeta}_i^{l+1} = \boldsymbol{\zeta}_i^l - \gamma_k \left(\boldsymbol{\zeta}_i^l - \mathsf{vec}_h(N\boldsymbol{\Omega}_{i,k}) + \lambda_{\text{tv}} \sum_{j\in\mathcal{N}_i} \mathsf{sign}(\boldsymbol{\zeta}_i^l - \boldsymbol{\zeta}_j^l)\right)$
    end for
    $\hat{\mathbf{x}}_{i,k} = \mathbf{x}_{i,k}^{l^*}$
    $\mathbf{P}_{i,k} = (\mathsf{vec}_h^{-1}(\boldsymbol{\zeta}_i^{l^*}))^{-1}$
end for
```

$$\begin{aligned} \hat{\mathbf{x}}_{i,k} &= \mathbf{x}_{i,k}^{l^*} \\ \mathbf{P}_{i,k} &= (\mathsf{vec}_h^{-1}(\boldsymbol{\zeta}_i^*))^{-1}. \end{aligned}$$

Assuming that Byzantine agents manipulate only state estimates, i.e., falsify the state estimate $\mathbf{x}_{i,k}^l$ at each iteration $l$, Algorithm 1 summarizes detailed operations of the proposed BR-DKF. It can be seen in Algorithm 1 that two additional $\mathsf{sign}(\cdot)$ operations are computed at each iteration $l$, compared to conventional consensus-based DKFs. The complexity of $\mathsf{sign}(\cdot)$ operator is dominated by the complexity of $O(m^2)$ for the multiplication of $\boldsymbol{\Omega}_{i,k}\mathbf{x}_{i,k}^l$ at each iteration $l$. As a result, the local computational complexity of the proposed method is the same as that of the conventional consensus-based DKF algorithms.

## IV. Performance Analysis

In this section, we demonstrate that the TV-norm-penalized problem in (8) yields a feasible solution when the penalty parameter $\lambda_{\text{tv}}$ is sufficiently large. We also show that the suboptimal solution in (10) converges to a neighborhood of the optimal solution of the problem in (8) with a bounded radius when Byzantine agents are in the network. To assist in future calculations, we define $\mathbf{A} = [a_{ij}] \in \mathbb{R}^{N\times|\mathcal{E}|}$ as the node-edge incidence matrix where for each edge $e = (i, j) \in \mathcal{E}$ with $i < j$, we set $a_{ei} = 1$ and $a_{je} = -1$, otherwise, the elements of $\mathbf{A}$ remain zero. In the following Theorem, we establish the optimality of the proposed solution in (8) to yield the same solution as the centralized solution $\hat{\mathbf{x}}_k^*$ in [49]. We provide a lower bound threshold for the penalty parameter $\lambda_{\text{tv}}$ that guarantees convergence of the solution in (8) to the centralized solution in [49].

*Theorem 1:* Given that the network topology is connected, if $\lambda_{\text{tv}} \geq \lambda_0$ where

$$\lambda_0 = \frac{\sqrt{N}}{\sigma_{\min}(\mathbf{A})} \max_{\forall k} \max_{i\in\mathcal{N}} \|\boldsymbol{\Omega}_{i,k}\mathbf{x}_{i,k}^* - \boldsymbol{\theta}_{i,k}\|_\infty \tag{16}$$

with $\sigma_{\min}(\mathbf{A})$ being the minimum non-zero singular value of $\mathbf{A}$, $\boldsymbol{\Omega}_{i,k}$ and $\boldsymbol{\theta}_{i,k}$ defined in (11); then, for the optimal solution $\mathbf{x}^*_{c_k}$ in (8) and the optimal solution of the CKF problem $\hat{\mathbf{x}}^*_k$ in [49], we have $\mathbf{x}^*_{c_k} = [\hat{\mathbf{x}}^*_k]_{i=1}^N$.

*Proof:* The proof begins with stating the fact that for each $i \in \mathcal{N}$, the optimal solution $\mathbf{x}^*_{c_k} = [\mathbf{x}^*_{i,k}]_{i=1}^N$ satisfies the optimality condition

$$\nabla_{\mathbf{x}_{i,k}} f_i(\mathbf{x}^*_{i,k}) + \lambda_{\text{tv}} \sum_{j \in \mathcal{N}_i} \mathsf{sign}(\mathbf{x}^*_{i,k} - \mathbf{x}^*_{j,k}) = \mathbf{0}\cdot \tag{17}$$

Let us assume $\mathbf{s}_{ij} = \mathsf{sign}(\mathbf{x}^*_{i,k} - \mathbf{x}^*_{j,k})$ and $\boldsymbol{\nu}_{i,k} = \nabla_{\mathbf{x}_{i,k}} f_i(\mathbf{x}^*_{i,k})$;[1] then knowing that $\mathbf{s}_{ji} = -\mathbf{s}_{ij}$, we have

$$\boldsymbol{\nu}_{i,k} + \lambda_{\text{tv}} \sum_{j \in \mathcal{N}_i, i<j} \mathbf{s}_{ij} - \lambda_{\text{tv}} \sum_{j \in \mathcal{N}_i, i>j} \mathbf{s}_{ij} = \mathbf{0}\cdot \tag{18}$$

Assuming $\boldsymbol{\nu}_k = [\boldsymbol{\nu}^{\mathrm{T}}_{1,k}, \cdots, \boldsymbol{\nu}^{\mathrm{T}}_{N,k}]^{\mathrm{T}}$ and $\mathbf{s} = [\mathbf{s}^{\mathrm{T}}_1, \cdots, \mathbf{s}^{\mathrm{T}}_{|\mathcal{E}|}]^{\mathrm{T}}$ with $\mathbf{s}_t = \mathbf{s}_{ij}$ for each $(i,j) \in \mathcal{E}$, we have

$$\boldsymbol{\nu}_k + \lambda_{\text{tv}} \mathbf{A}\mathbf{s} = \mathbf{0}\cdot \tag{19}$$

Now the problem is to show that (19) has at least one solution $\mathbf{s}^*$ and due to the structure of $\mathbf{s}$ its elements are within $[-1, 1]$ or $\|\mathbf{s}\|_\infty \leq 1$. The rank of $\mathbf{A}$ is $N-1$ with the column null space of one vector, i.e., $\mathbf{1}_N$, since $\mathcal{G} = \{\mathcal{N}, \mathcal{E}\}$ is bidirectionally connected. In addition, the optimality condition of the centralized solution in [49] satisfies

$$\sum_{i \in \mathcal{N}} \boldsymbol{\nu}_{i,k} = \sum_{i \in \mathcal{N}} \nabla_{\mathbf{x}_{i,k}} f_i(\mathbf{x}^*_{i,k}) = \mathbf{0} \tag{20}$$

which means $\lambda_{\text{tv}}\mathbf{A}$ and $\boldsymbol{\nu}_k$ share the same null space and have the same rank that consequently, states that we will have at least one solution for (19). In order to find the solution that satisfies $\|\mathbf{s}\|_\infty \leq 1$, we consider the least-squares solution as $\mathbf{s} = -\frac{1}{\lambda_{\text{tv}}} \mathbf{A}^\dagger \boldsymbol{\nu}_k$ where † denotes the pseudo inverse. The least-squares solution is bounded as

$$\|\mathbf{s}\|_2 = \frac{1}{\lambda_{\text{tv}}} \|\mathbf{A}^\dagger \boldsymbol{\nu}_k\|_2\cdot \tag{21}$$

Then, we have

$$\|\mathbf{s}\|_2 \leq \frac{1}{\lambda_{\text{tv}}} \sigma_{\max}(\mathbf{A}^\dagger) \|\boldsymbol{\nu}_k\|_2 \leq \frac{1}{\lambda_{\text{tv}} \sigma_{\min}(\mathbf{A})} \|\boldsymbol{\nu}_k\|_2 \tag{22}$$

where $\sigma_{\max}(\cdot)$ and $\sigma_{\min}(\cdot)$ represent the maximum and minimum non-zero singular values of the argument matrix, respectively. Since $\|\mathbf{s}\|_\infty \leq \|\mathbf{s}\|_2$ and $\|\boldsymbol{\nu}_k\|_2 \leq \sqrt{N} \|\boldsymbol{\nu}_k\|_\infty$, we have

$$\|\mathbf{s}\|_\infty \leq \frac{\sqrt{N}}{\lambda_{\text{tv}} \sigma_{\min}(\mathbf{A})} \|\boldsymbol{\nu}_k\|_\infty = \frac{\sqrt{N}}{\lambda_{\text{tv}} \sigma_{\min}(\mathbf{A})} \max_{i \in \mathcal{N}} |\boldsymbol{\nu}_{i,k}|\cdot \tag{23}$$

Thus, $\|\mathbf{s}\|_\infty \leq 1$ if $\lambda_{\text{tv}} \geq \frac{\sqrt{N}}{\sigma_{\min}(\mathbf{A})} \max_{i \in \mathcal{N}} |\boldsymbol{\nu}_{i,k}|$ for each $k$, which results in the requirement of $\lambda_{\text{tv}} \geq \lambda_0$ where

$$\begin{aligned} \lambda_0 &= \frac{\sqrt{N}}{\sigma_{\min}(\mathbf{A})} \max_{\forall k} \max_{i \in \mathcal{N}} \|\nabla_{\mathbf{x}^*_{c_i}} f_i(\mathbf{x}^*_{c_i})\|_\infty \\ &= \frac{\sqrt{N}}{\sigma_{\min}(\mathbf{A})} \max_{\forall k} \max_{i \in \mathcal{N}} \|\boldsymbol{\Omega}_{i,k} \mathbf{x}^*_{i,k} - \boldsymbol{\theta}_{i,k}\|_\infty \end{aligned}$$

that completes the proof. ■

After showing the convergence of the proposed method to the desired centralized case, we need to theoretically analyze the performance of the proposed solution in the presence of Byzantines. The following theorem shows that the proposed suboptimal solution in (10) converges to a neighborhood of the optimal centralized solution within a bounded radius despite the presence of Byzantine agents.

*Theorem 2:* Given the assumptions in Theorem 1 and $\lambda_{\text{tv}} \geq \lambda_0$, at each agent $i \in \mathcal{N}$ and the presence of Byzantine agents, the solution proposed in (10) stays in the neighborhood of the optimal solution $\mathbf{x}^*_{c_k} = [\mathbf{x}^*_{i,k}]_{i=1}^N$ in (8) with radius

$$\lim_{l \to \infty} \mathbb{E}_l\{\|\mathbf{x}^{l+1}_{i,k} - \mathbf{x}^*_{i,k}\|^2\} \leq \frac{\Delta_0}{1 - \|\boldsymbol{\Delta}\|} \tag{24}$$

where $\boldsymbol{\Delta} = \left(1 + 2\alpha_k^2 \|\boldsymbol{\Omega}_{i,k}\|^2 + 2\varepsilon\alpha_k\right)\mathbf{I} - 2\alpha_k \boldsymbol{\Omega}_{i,k}$, $\Delta_0 = \lambda^2_{\text{tv}} \alpha_k (4\alpha_k + \frac{1}{\varepsilon})(4|\mathcal{R}_i|^2 + |\mathcal{B}_i|^2) m$, and the step size $\alpha_k$ satisfies

$$\alpha_k \leq \min_{i \in \mathcal{N}} \left\{ \frac{\lambda_{\min}(\boldsymbol{\Omega}_{i,k}) - \varepsilon}{\|\boldsymbol{\Omega}_{i,k}\|^2} \right\}\cdot \tag{25}$$

*Proof:* The proof begins by computing the gap between the optimal solution in (8) and the proposed solution in (10) after $l$ iterations as follows

$$\begin{aligned} \mathbb{E}_l\{\|\mathbf{x}^{l+1}_{i,k} - \mathbf{x}^*_{i,k}\|^2\} = \mathbb{E}_l\{\|\mathbf{x}^l_{i,k} - \mathbf{x}^*_{i,k} \\ - \alpha_k \big(\boldsymbol{\Omega}_{i,k} \mathbf{x}^l_{i,k} - \boldsymbol{\theta}_{i,k} + \lambda_{\text{tv}} \sum_{j \in \mathcal{N}_i} \mathsf{sign}(\mathbf{x}^l_{i,k} - \mathbf{x}^l_{j,k})\big)\|^2\}\cdot \end{aligned} \tag{26}$$

In this case, (26) can be further simplified as

$$\begin{aligned} &\mathbb{E}_l\{\|\mathbf{x}^{l+1}_{i,k} - \mathbf{x}^*_{i,k}\|^2\} = \mathbb{E}_l\{\|\mathbf{x}^l_{i,k} - \mathbf{x}^*_{i,k}\|^2\} \\ &+ \underbrace{\alpha_k^2\, \mathbb{E}_l\{\|\boldsymbol{\Omega}_{i,k} \mathbf{x}^l_{i,k} - \boldsymbol{\theta}_{i,k} + \lambda_{\text{tv}} \sum_{j \in \mathcal{N}_i} \mathsf{sign}(\mathbf{x}^l_{i,k} - \mathbf{x}^l_{j,k})\|^2\}}_{\beta_1} \\ &- 2\alpha_k \underbrace{< \mathbf{x}^l_{i,k} - \mathbf{x}^*_{i,k}, \boldsymbol{\Omega}_{i,k} \mathbf{x}^l_{i,k} - \boldsymbol{\theta}_{i,k} + \lambda_{\text{tv}} \sum_{j \in \mathcal{N}_i} \mathsf{sign}(\mathbf{x}^l_{i,k} - \mathbf{x}^l_{j,k}) >}_{\beta_2} \end{aligned} \tag{27}$$

Considering the optimality condition for the optimal solution $\mathbf{x}^*_{c_k}$ in (8) as

$$\boldsymbol{\Omega}_{i,k} \mathbf{x}^*_{i,k} - \boldsymbol{\theta}_{i,k} + \lambda_{\text{tv}} \sum_{j \in \mathcal{R}_i} \mathsf{sign}(\mathbf{x}^*_{i,k} - \mathbf{x}^*_{j,k}) = \mathbf{0}, \tag{28}$$

we have

$$\begin{aligned} \beta_1 =& \mathbb{E}_l\{\|\boldsymbol{\Omega}_{i,k} \mathbf{x}^l_{i,k} - \boldsymbol{\theta}_{i,k} + \lambda_{\text{tv}} \sum_{j \in \mathcal{R}_i} \mathsf{sign}(\mathbf{x}^l_{i,k} - \mathbf{x}^l_{j,k}) \\ &+ \lambda_{\text{tv}} \sum_{j \in \mathcal{B}_i} \mathsf{sign}(\mathbf{x}^l_{i,k} - \tilde{\mathbf{x}}^l_{j,k}) - \boldsymbol{\Omega}_{i,k} \mathbf{x}^*_{i,k} + \boldsymbol{\theta}_{i,k} \\ &- \lambda_{\text{tv}} \sum_{j \in \mathcal{R}_i} \mathsf{sign}(\mathbf{x}^*_{i,k} - \mathbf{x}^*_{j,k})\|^2\} \qquad (29) \\ =& \mathbb{E}_l\{\|\boldsymbol{\Omega}_{i,k}(\mathbf{x}^l_{i,k} - \mathbf{x}^*_{i,k}) + \lambda_{\text{tv}} \sum_{j \in \mathcal{B}_i} \mathsf{sign}(\mathbf{x}^l_{i,k} - \tilde{\mathbf{x}}^l_{j,k}) \\ &+ \lambda_{\text{tv}} \sum_{j \in \mathcal{R}_i} \big(\mathsf{sign}(\mathbf{x}^l_{i,k} - \mathbf{x}^l_{j,k}) - \mathsf{sign}(\mathbf{x}^*_{i,k} - \mathbf{x}^*_{j,k})\big)\|^2\}\cdot \end{aligned}$$

[1] Throughout the article, we remove the index $k$ from $\mathbf{s}_{ij}$ in order to simplify the notation.

Due to the inequality $(a+b)^2 \leq 2a^2 + 2b^2$, we have

$$\begin{aligned}\beta_1 \leq\ & 2\mathbb{E}_l\{\|\mathbf{\Omega}_{i,k}(\mathbf{x}^l_{i,k} - \mathbf{x}^*_{i,k})\|^2\} \\ & + 2\lambda_{\text{tv}}^2 \mathbb{E}_l\{\| \sum_{j \in \mathcal{R}_i} \big(\mathsf{sign}(\mathbf{x}^l_{i,k} - \mathbf{x}^l_{j,k}) - \mathsf{sign}(\mathbf{x}^*_{i,k} - \mathbf{x}^*_{j,k})\big) \\ & + \sum_{j \in \mathcal{B}_i} \mathsf{sign}(\mathbf{x}^l_{i,k} - \tilde{\mathbf{x}}^l_{j,k})\|^2\} \qquad (30) \\ \leq & 2\mathbb{E}_l\{\|\mathbf{\Omega}_{i,k}(\mathbf{x}^l_{i,k} - \mathbf{x}^*_{i,k})\|^2\} \\ & + 4\lambda_{\text{tv}}^2 \underbrace{\mathbb{E}_l\{\| \sum_{j \in \mathcal{B}_i} \mathsf{sign}(\mathbf{x}^l_{i,k} - \tilde{\mathbf{x}}^l_{j,k})\|^2\}}_{\leq |\mathcal{B}_i|^2 m} \\ & + 4\lambda_{\text{tv}}^2 \underbrace{\mathbb{E}_l\{\| \sum_{j \in \mathcal{R}_i} \big(\mathsf{sign}(\mathbf{x}^l_{i,k} - \mathbf{x}^l_{j,k}) - \mathsf{sign}(\mathbf{x}^*_{i,k} - \mathbf{x}^*_{j,k})\big)\|^2\}}_{\leq 4|\mathcal{R}_i|^2 m} \cdot\end{aligned}$$

Since for the matrix norm $\|\cdot\|$, we have[2]

$$\mathsf{tr}(\mathbf{AB}) \leq \min\{\|\mathbf{A}\|\mathsf{tr}(\mathbf{B}), \|\mathbf{B}\|\mathsf{tr}(\mathbf{A})\} \qquad (31)$$

where $\mathbf{A}$ and $\mathbf{B}$ are positive semi-definite and $\|\mathbf{AB}\| \leq \|\mathbf{A}\|\|\mathbf{B}\|$, we can show that

$$\|\mathbf{\Omega}_{i,k}(\mathbf{x}^l_{i,k} - \mathbf{x}^*_{i,k})\|^2 \leq \|\mathbf{\Omega}_{i,k}\|^2\|\mathbf{x}^l_{i,k} - \mathbf{x}^*_{i,k}\|^2,$$

and subsequently

$$\beta_1 \leq 2\|\mathbf{\Omega}_{i,k}\|^2\|\mathbf{x}^l_{i,k} - \mathbf{x}^*_{i,k}\|^2 + 4\lambda_{\text{tv}}^2(4|\mathcal{R}_i|^2 + |\mathcal{B}_i|^2)m\cdot \qquad (32)$$

Additionally, we have

$$\begin{aligned}-2 \times \beta_2 = & -2 < \mathbf{x}^l_{i,k} - \mathbf{x}^*_{i,k}, \mathbf{\Omega}_{i,k}\mathbf{x}^l_{i,k} - \boldsymbol{\theta}_{i,k} \\ & + \lambda_{\text{tv}} \sum_{j \in \mathcal{N}_i} \mathsf{sign}(\mathbf{x}^l_{i,k} - \mathbf{x}^l_{j,k}) \\ & - \mathbf{\Omega}_{i,k}\mathbf{x}^*_{i,k} + \boldsymbol{\theta}_{i,k} - \lambda_{\text{tv}} \sum_{j \in \mathcal{R}_i} \mathsf{sign}(\mathbf{x}^*_{i,k} - \mathbf{x}^*_{j,k}) > \\ = & -2 < \mathbf{x}^l_{i,k} - \mathbf{x}^*_{i,k}, \mathbf{\Omega}_{i,k}(\mathbf{x}^l_{i,k} - \mathbf{x}^*_{i,k}) > \\ & - 2 < \mathbf{x}^l_{i,k} - \mathbf{x}^*_{i,k}, \lambda_{\text{tv}} \sum_{j \in \mathcal{B}_i} \mathsf{sign}(\mathbf{x}^l_{i,k} - \tilde{\mathbf{x}}^l_{j,k}) > \\ & - 2 < \mathbf{x}^l_{i,k} - \mathbf{x}^*_{i,k}, \lambda_{\text{tv}} \sum_{j \in \mathcal{R}_i} \big(\mathsf{sign}(\mathbf{x}^l_{i,k} - \mathbf{x}^l_{j,k}) \\ & \qquad - \mathsf{sign}(\mathbf{x}^*_{i,k} - \mathbf{x}^*_{j,k})\big) > \cdot \qquad (33)\end{aligned}$$

The inequality $-2ab \leq \varepsilon a^2 + \frac{b^2}{\varepsilon}$ for each $\varepsilon \geq 0$ gives

$$\begin{aligned}-2 <& \mathbf{x}^l_{i,k} - \mathbf{x}^*_{i,k}, \lambda_{\text{tv}} \sum_{j \in \mathcal{B}_i} \mathsf{sign}(\mathbf{x}^l_{i,k} - \tilde{\mathbf{x}}^l_{j,k}) > \\ & \leq \varepsilon\|\mathbf{x}^l_{i,k} - \mathbf{x}^*_{i,k}\|^2 + \frac{\lambda_{\text{tv}}^2}{\varepsilon}|\mathcal{B}_i|^2 m \qquad (34)\end{aligned}$$

and

$$\begin{aligned}-2 <& \mathbf{x}^l_{i,k} - \mathbf{x}^*_{i,k}, \lambda_{\text{tv}}\sum_{j \in \mathcal{R}_i} \big(\mathsf{sign}(\mathbf{x}^l_{i,k} - \mathbf{x}^l_{j,k}) \\ & - \mathsf{sign}(\mathbf{x}^*_{i,k} - \mathbf{x}^*_{j,k})\big) > \\ & \leq \varepsilon\|\mathbf{x}^l_{i,k} - \mathbf{x}^*_{i,k}\|^2 + \frac{4\lambda_{\text{tv}}^2}{\varepsilon}|\mathcal{R}_i|^2 m\cdot \qquad (35)\end{aligned}$$

After substituting (29) and (33) in (27), we get

$$\begin{aligned}& \mathbb{E}_l\{\|\mathbf{x}^{l+1}_{i,k} - \mathbf{x}^*_{i,k}\|^2\} \\ & \leq \big(1 + 2\alpha_k^2\|\mathbf{\Omega}_{i,k}\|^2 + 2\varepsilon\alpha_k\big)\,\mathbb{E}_l\{\|\mathbf{x}^l_{i,k} - \mathbf{x}^*_{i,k}\|^2\} \\ & \quad - 2\alpha_k\mathbb{E}_l\{(\mathbf{x}^l_{i,k} - \mathbf{x}^*_{i,k})^{\mathrm{T}}\mathbf{\Omega}_{i,k}(\mathbf{x}^l_{i,k} - \mathbf{x}^*_{i,k})\} \\ & \quad + \lambda_{\text{tv}}^2\alpha_k(4\alpha_k + \frac{1}{\varepsilon})(4|\mathcal{R}_i|^2 + |\mathcal{B}_i|^2)m \qquad (36) \\ & = \mathbb{E}_l\{(\mathbf{x}^l_{i,k} - \mathbf{x}^*_{i,k})^{\mathrm{T}}\Big(\big(1 + 2\alpha_k^2\|\mathbf{\Omega}_{i,k}\|^2 + 2\varepsilon\alpha_k\big)\mathbf{I} - 2\alpha_k\mathbf{\Omega}_{i,k}\Big) \\ & \quad (\mathbf{x}^l_{i,k} - \mathbf{x}^*_{i,k})\} + \lambda_{\text{tv}}^2\alpha_k(4\alpha_k + \frac{1}{\varepsilon})(4|\mathcal{R}_i|^2 + |\mathcal{B}_i|^2)m\cdot\end{aligned}$$

To guarantee that the error is decreasing with each iteration, we must have

$$\big(1 + 2\alpha_k^2\|\mathbf{\Omega}_{i,k}\|^2 + 2\varepsilon\alpha_k\big)\,\mathbf{I} - 2\alpha_k\mathbf{\Omega}_{i,k} \preccurlyeq \mathbf{I} \qquad (37)$$

that yields

$$2\alpha_k\,\big(\alpha_k\|\mathbf{\Omega}_{i,k}\|^2\mathbf{I} + \varepsilon\mathbf{I} - \mathbf{\Omega}_{i,k}\big) \preccurlyeq 0 \qquad (38)$$

Since $\alpha_k \geq 0$ and by assuming $\bar{\alpha}_k = \alpha_k\|\mathbf{\Omega}_{i,k}\|^2 + \varepsilon$, we only need to have

$$\mathbf{\Omega}_{i,k} - \bar{\alpha}_k\mathbf{I} \succcurlyeq 0 \qquad (39)$$

which requires $\bar{\alpha}_k \leq \lambda_j(\mathbf{\Omega}_{i,k})$ for each $j = 1, 2, \cdots, m$ that means

$$\alpha_k \leq \frac{\lambda_{\min}(\mathbf{\Omega}_{i,k}) - \varepsilon}{\|\mathbf{\Omega}_{i,k}\|^2}\cdot$$

Thus, to ensure that the error gap $\mathbb{E}_l\{\|\mathbf{x}^{l+1}_i - \mathbf{x}^*_i\|^2\}$ is bounded for all agents, the step size must satisfy

$$0 \leq \alpha_k \leq \min_{i \in \mathcal{N}}\left\{\frac{\lambda_{\min}(\mathbf{\Omega}_{i,k}) - \varepsilon}{\|\mathbf{\Omega}_{i,k}\|^2}\right\} \qquad (40)$$

where $0 \leq \varepsilon \leq \lambda_{\min}(\mathbf{\Omega}_{i,k})$. Defining

$$\begin{aligned}\mathbf{\Delta} &= \big(1 + 2\alpha_k^2\|\mathbf{\Omega}_{i,k}\|^2 + 2\varepsilon\alpha_k\big)\,\mathbf{I} - 2\alpha_k\mathbf{\Omega}_{i,k} \\ \Delta_0 &= \lambda_{\text{tv}}^2\alpha_k(4\alpha_k + \frac{1}{\varepsilon})(4|\mathcal{R}_i|^2 + |\mathcal{B}_i|^2)m\end{aligned}$$

and assuming that $\alpha_k$ satisfies (40), we get $\|\mathbf{\Delta}\| \leq 1$. Now, employing (31), the error gap in (36) turns into

$$\mathbb{E}_l\{\|\mathbf{x}^{l+1}_{i,k} - \mathbf{x}^*_{i,k}\|^2\} \leq \|\mathbf{\Delta}\|\,\mathbb{E}_l\{\|\mathbf{x}^l_{i,k} - \mathbf{x}^*_{i,k}\|^2\} + \Delta_0, \qquad (41)$$

which simplifies as

$$\mathbb{E}_l\{\|\mathbf{x}^{l+1}_{i,k} - \mathbf{x}^*_{i,k}\|^2\} \leq \|\mathbf{\Delta}\|^{l+1}\,\mathbb{E}_l\{\|\mathbf{x}^0_{i,k} - \mathbf{x}^*_{i,k}\|^2\} + \Delta_0\sum_{s=0}^{l}\|\mathbf{\Delta}\|^s\cdot \qquad (42)$$

As a result of $\|\mathbf{\Delta}\| \leq 1$, the error gap becomes

$$\lim_{l \to \infty} \mathbb{E}_l\{\|\mathbf{x}^{l+1}_{i,k} - \mathbf{x}^*_{i,k}\|^2\} \leq \frac{\Delta_0}{1 - \|\mathbf{\Delta}\|} \qquad (43)$$

asymptotically, which completes the proof. ■

*Remark 1:* The error gap in (43) illustrates that the BR-DKF restricts the impact of attack amplitude completely due to the $\mathsf{sign}(\cdot)$ terms; however, the number of Byzantine agents in the network still affects the error bound in (43) by altering $\Delta_0$.

*Remark 2:* This work provides the analysis for an undirected graph topology, and analyzing the algorithm with a directed graph topology requires a new analysis, which is beyond the scope of this work.

[2] The matrix norm $\|\cdot\|$ is defined as $\|\mathbf{A}\| \triangleq \sigma_{\max}(\mathbf{A})$ with $\sigma_{\max}(\cdot)$ representing the largest singular value of the argument matrix.

## V. Simulation Results

The performance of the proposed Byzantine-Robust DKF (BR-DKF) is illustrated by considering two network topologies, including a network of $N = 5$ agents with the edge set of $\mathcal{E} = \{(1,2),(2,3),(3,5),(4,5),(5,1)\}$, same as [15], shown in Fig. 1, and a randomly generated undirected connected network with $N = 25$ agents with the topology shown in Fig. 6. The discrete-time system and agent parameters are considered similar to the work in [49], and are given by

$$\mathbf{x}_{k+1} = \begin{bmatrix} 0.4 & 0.9 & 0 & 0 \\ -0.9 & 0.4 & 0 & 0 \\ 0 & 0 & 0.5 & 0.8 \\ 0 & 0 & -0.8 & 0.5 \end{bmatrix} \mathbf{x}_k + \mathbf{w}_k,$$

$$\mathbf{y}_{i,k} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix} \mathbf{x}_k + \mathbf{v}_{i,k},$$

where the state noise covariance $\mathbf{Q} = 0.1\mathbf{I}$, and the observation noise covariance $\mathbf{R}_i = \mathsf{diag}(0.1, 0.2, 0.3, 0.1)$. To benchmark our proposed algorithm, we evaluate the following scenarios: the centralized KF (CKF), distributed KF (DKF) [49], DKF subject to Byzantine attack (B-DKF), and the proposed BR-DKF subject to Byzantine attack.

Considering Byzantines as $B$ agents with the largest node degree in the graph topology, the corresponding perturbation covariances are designed following the optimization problem $\mathcal{P}_1$ in [16]. In problem $\mathcal{P}_1$, the steady-state network mean squared error (NMSE) is maximized by designing the covariance of the perturbation sequences at the Byzantine agents. The NMSE is defined as

$$\text{NMSE} \triangleq \limsup_{K\to\infty} \frac{1}{K} \sum_{k=1}^{K} \sum_{i=1}^{N} \mathsf{tr}(\mathbf{P}_{i,k}),$$

where $\mathbf{P}_{i,k}$ is the error covariance of the DKF in [16] at agent $i$ and time instant $k$. Accordingly, the optimization problem to design the perturbation covariances is modeled as

$$\begin{aligned} \max_{\boldsymbol{\Sigma}}.\quad & \text{NMSE} \\ \text{s. t.}\quad & \sum_{j\in\mathcal{B}} \mathsf{tr}(\boldsymbol{\Sigma}_j) \leq \eta, \\ & \boldsymbol{\Sigma} \succcurlyeq 0, \end{aligned}$$

where the first constraint limits the total power of the falsification sequences and satisfies the detection-avoidance target with parameter $\eta$. The second constraint ensures that the perturbation covariance $\boldsymbol{\Sigma}$ is positive semidefinite. As a result, the proposed algorithm is examined under the worst-case scenario of an attack that maximizes the network MSE.

In the first scenario, we consider the network in Fig. 1 comprising $N = 5$ agents, of which $B = 2$ are Byzantine agents, taken as the agents with the highest node degree. We plot the average MSE across agents, i.e.,

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^{N} (\mathbf{x}_k - \hat{\mathbf{x}}_{i,k})^{\mathrm{T}} (\mathbf{x}_k - \hat{\mathbf{x}}_{i,k}) \cdot \tag{44}$$

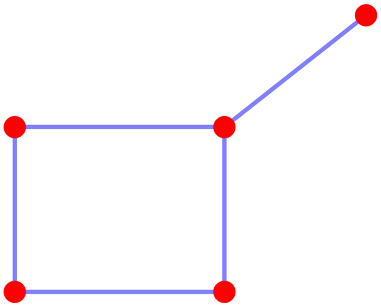


Fig. 1. Network topology with $\boldsymbol{N = 5}$ agents.

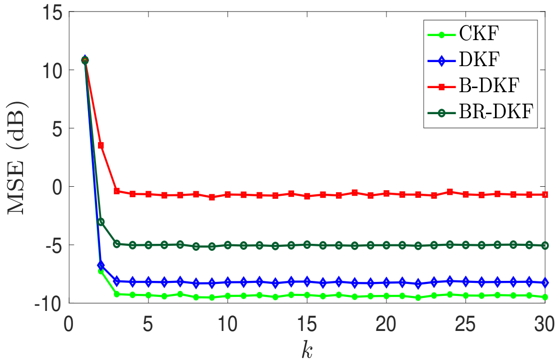


Fig. 2. MSE versus filtering time index $\boldsymbol{k}$ in the network with $\boldsymbol{N = 5}$ agents.

In the absence of Byzantines, the parameters of $\alpha_k$, $\gamma_k$, and $\lambda_{\text{tv}}$ of the BR-DKF are tuned to obtain the nearest possible MSE to the DKF algorithm. Even without a Byzantine attack, the BR-DKF does not reach the same performance as the DKF method; this is because the $\mathsf{sign}(\cdot)$ terms in the updating process restrict the actual values of the state estimate. Here, Byzantine agents conduct a coordinated data-falsification attack where $\boldsymbol{\Sigma}_i$ denotes the covariance matrix of perturbation sequences of agent $i \in \mathcal{B}$.

Fig. 2 shows the MSE in (44) versus the filtering time index $k$ in a network of $N = 5$ agents. The number of iterations for the state estimate and the error covariance updates is set to $l^* = 25$ and the results are averaged over 2000 Monte Carlo experiments. The BR-DKF achieves lower MSE than the B-DKF under the same Byzantine attack, which demonstrates its robustness. There is a performance gap between centralized and distributed Kalman filters, even without Byzantine agents, which is due to the number of iterations in the subgradient solution. By increasing the number of $l^*$, the performance of the DKF will approach the CKF asymptotically.

Fig. 3 shows how the actual state of the network, with $m = 4$, is closely estimated by various filtering methods. Tracking performance for different filtering settings is illustrated in shaded colors for all agents in the network, and the average of the estimate for all agents is shown as a solid line. We see that the proposed BR-DKF method estimates the actual state elements with a smaller variance than the B-DKF method.

Fig. 4 shows the MSE versus the percentage of Byzantine agents in the network. The BR-DKF method is significantly

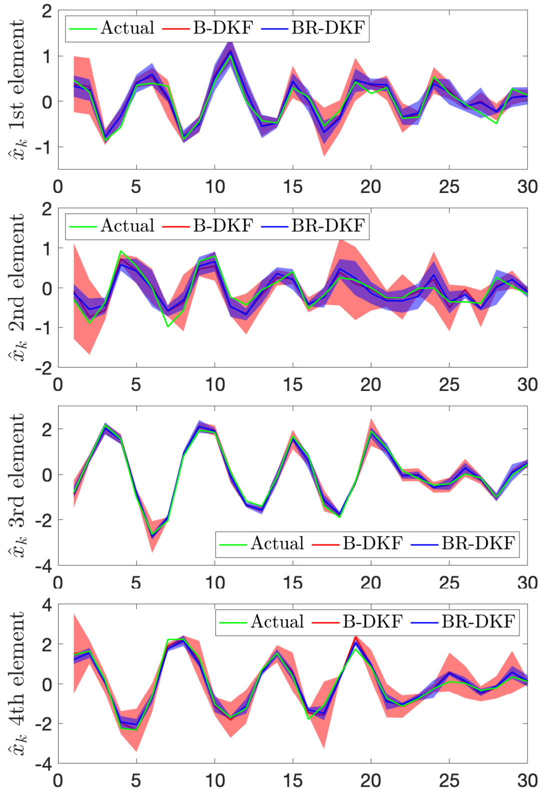


Fig. 3. State estimation accuracy for the different elements of the state in a network of $\boldsymbol{N = 5}$.

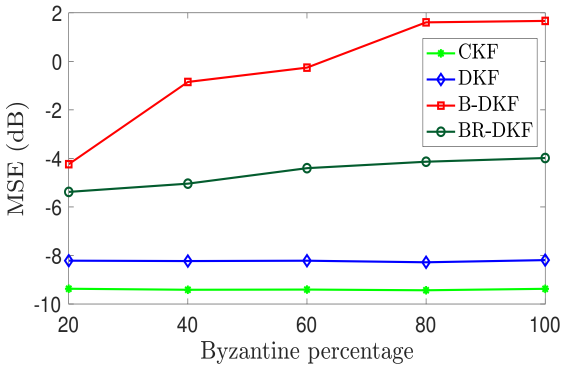


Fig. 4. Steady-state MSE versus percentage of the Byzantine agents in the network with $\boldsymbol{N = 5}$ agents.

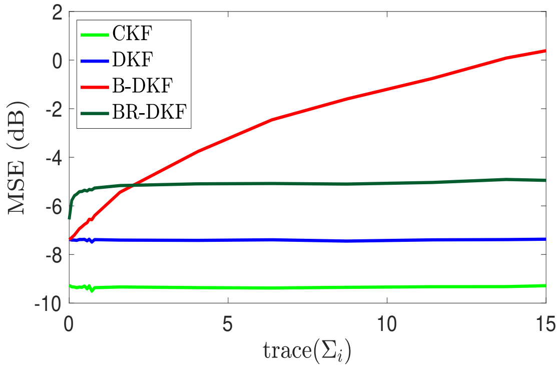


Fig. 5. Steady-state MSE versus trace of the Byzantine agent attack covariance in the network with $\boldsymbol{N = 5}$ agents.

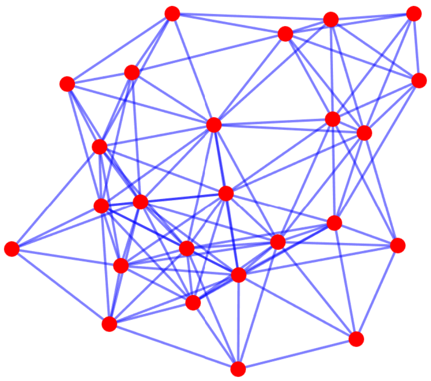


Fig. 6. Network topology with $\boldsymbol{N = 25}$ agents.

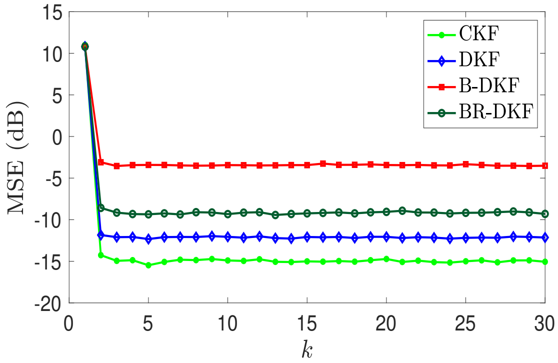


Fig. 7. MSE versus filtering time index $\boldsymbol{k}$ in a network $\boldsymbol{N = 25}$ agents.

less sensitive to the number of Byzantines in the networks than the B-DKF method. Fig. 5 shows the MSE versus the trace of perturbation sequence covariance of individual Byzantine agents. As shown, even without injecting any noise by the Byzantine agent, the MSE in BR-DKF does not reach the DKF method; this is because the sign$(\cdot)$ terms in the updating process limit the actual value of the state estimates. Upon starting the Byzantine attack, the obtained MSE under the B-DKF increases dramatically as more noise is injected, but the obtained MSE under the BR-DKF does not change. This is due to the restriction that the sign$(\cdot)$ term provides, and as stated in *Remark 1*, the number of Byzantine agents is the only factor impacting the steady-state MSE in BR-DKF.

In the second scenario, we consider a network of $N = 25$ agents as in Fig. 6, including $B = 5$ Byzantine agents that are chosen as network agents with the highest node degree. A similar tuning is made to the step size parameters in order to ensure the smallest difference in MSE for DKF and BR-DKF algorithms in the absence of an attack. In Fig. 7, the MSE in (44) is plotted versus the filtering time index $k$ for different filtering approaches. The subgradient solution for the state and error covariance are iterated for $l^* = 25$ iterations. Under the same Byzantine attack, the proposed BR-DKF obtains a lower MSE than the B-DKF, which verifies its robustness against Byzantine behaviors.

Similar to the previous scenario, the estimation accuracy for

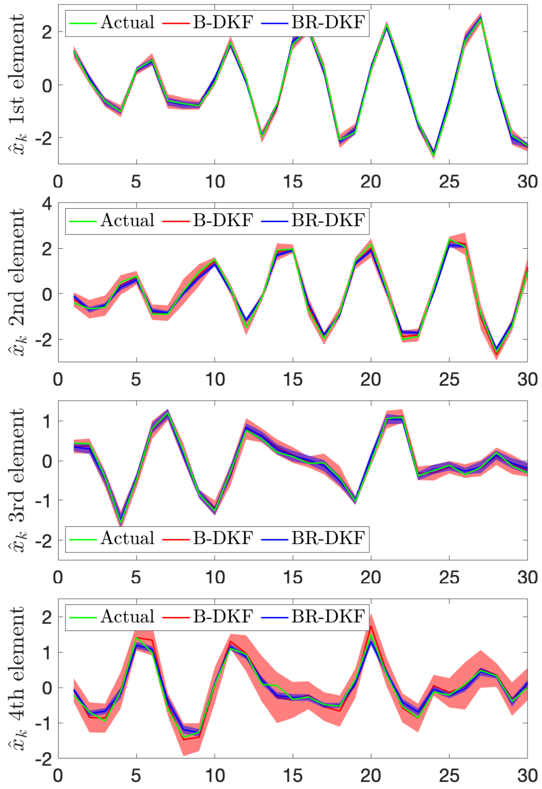


Fig. 8. State estimation accuracy for the different elements of the state in a network of $N = 25$.

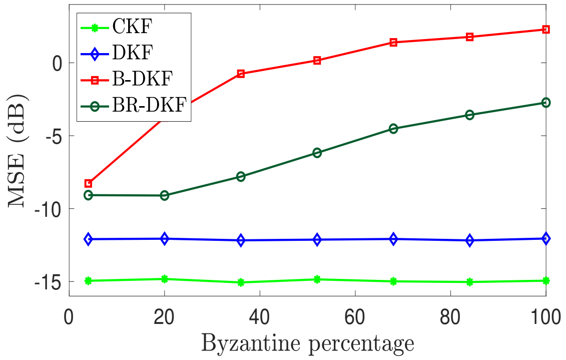


Fig. 9. Steady-state MSE versus percentage of the Byzantine agents in the network with $N = 25$ agents.

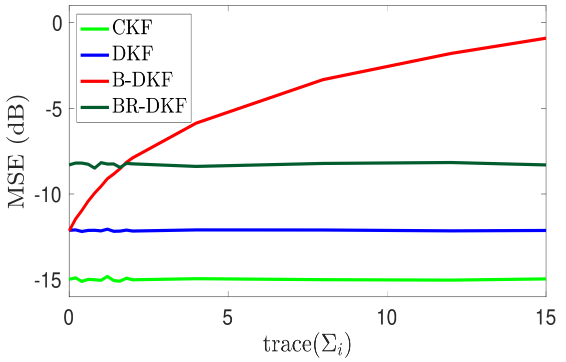


Fig. 10. Steady-state MSE versus trace of the Byzantine agent attack covariance in the network with $N = 25$ agents.

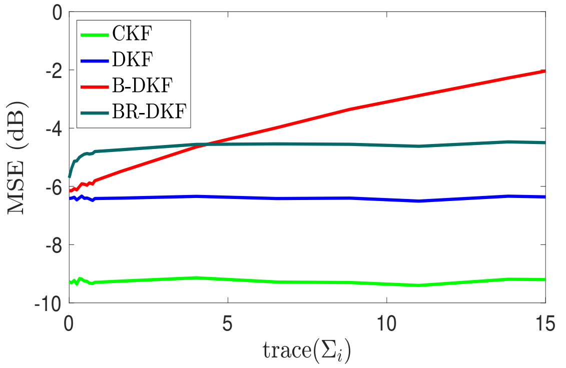


Fig. 11. Steady-state MSE versus trace of the Byzantine agent attack covariance, with unstable state matrix, in the network with $N = 5$ agents.

different state vector elements, with $m = 4$, is shown in Fig. 8. The estimated values of agents are plotted in shaded colors and their average of the estimated values in solid colors. It can be seen that the proposed BR-DKF reduces the variance of the estimated values and can robustly track the actual state of the network with higher accuracy than the B-DKF algorithm.

Fig. 9 illustrates the obtained MSE versus the percentage of Byzantine agents in the network for different algorithms. A similar trend is observed, showing that the greater the percentage of Byzantine agents in the network, the higher the MSE, while the BR-DKF sensitivity to the Byzantine percentage is significantly less than the B-DKF. In Fig. 10, the MSE is illustrated versus the trace of the perturbation covariance of individual Byzantine agents, which shows that under the BR-DKF, as the trace of attack covariance is low, sign$(\cdot)$ terms in the state estimate update equations constrain the actual values and degrade the MSE compared to the DKF. When Byzantines inject more noise, the performance of the BR-DKF is not degraded, while under the B-DKF algorithm, the MSE increases significantly as more noise is injected. This confirms the resilience of the BR-DKF to the coordinated data falsification attack.

Simulation results are provided for a stable state matrix $\mathbf{F}$, spectral radius less than one, while the algorithm also performs efficiently for unstable state matrices. To verify the stability of the proposed algorithm using an unstable state matrix, in Fig. 11 and Fig. 12, we plot the MSE versus the trace of perturbation covariance for the case where only $\mathbf{F}$ is different and is considered as

$$\mathbf{F} = \begin{bmatrix} 0.6 & 0.9 & 0 & 0 \\ -0.9 & 0.6 & 0 & 0 \\ 0 & 0 & 0.7 & 0.8 \\ 0 & 0 & -0.8 & 0.7 \end{bmatrix}. \tag{45}$$

It can be seen that the trend of changing MSE versus trace of the perturbation covariance in different algorithms remains the same.

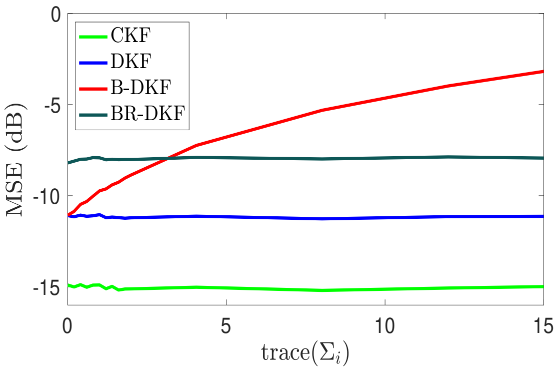


Fig. 12. Steady-state MSE versus trace of the Byzantine agent attack covariance, with unstable state matrix, in the network with $N = 25$ agents.

## VI. Conclusion

This paper proposed a distributed Kalman filter (DKF) with resiliency against Byzantine attacks. Considering the Byzantine agent as a network member that alters information before exchanging it with neighbors, we investigated DKF operations from the perspective of distributed optimization. The resulting optimization-based DKF solution improved the robustness of the filtering operations against Byzantine behaviors by employing a TV-norm penalty term for the objective function. We utilized a distributed subgradient algorithm to derive a suboptimal solution to update the state estimate and error covariance matrix of the proposed Byzantine robust DKF (BR-DKF). Furthermore, we demonstrated that the proposed suboptimal solution converges to a neighborhood of the optimal centralized solution with a bounded radius in the presence of the Byzantine agents. Numerical simulations corroborated the theoretical findings and demonstrated the robustness of the proposed BR-DKF against Byzantine behaviors. In future research, the impact of time-varying and directed graph topologies on the performance of the proposed algorithm will be investigated.

## References

[1] Y. Liu, B. Wang, W. Ye, X. Ning, and B. Gu, "Global estimation method based on spatial–temporal Kalman filter for dpos," *IEEE Sensors J.*, vol. 21, no. 3, pp. 3748–3756, Feb. 2021.

[2] C. Li and H. Wang, "Distributed frequency estimation over sensor network," *IEEE Sensors J.*, vol. 15, no. 7, pp. 3973–3983, July 2015.

[3] Y. Yu, "Consensus-based distributed linear filter for target tracking with uncertain noise statistics," *IEEE Sensors J.*, vol. 17, no. 15, pp. 4875–4885, Aug. 2017.

[4] Y. Chen, Q. Zhao, Z. An, P. Lv, and L. Zhao, "Distributed multi-target tracking based on the K-MTSCF algorithm in camera networks," *IEEE Sensors J.*, vol. 16, no. 13, pp. 5481–5490, July 2016.

[5] R. Olfati, "Kalman-consensus filter: Optimality, stability, and performance," in *Proc. 48th IEEE Conf. Decis. and Control*, 2009, pp. 7036–7042.

[6] F. S. Cattivelli and A. H. Sayed, "Diffusion strategies for distributed Kalman filtering and smoothing," *IEEE Trans. Autom. Control*, vol. 55, no. 9, pp. 2069–2084, Sept. 2010.

[7] D. Kapetanovic, G. Zheng, and F. Rusek, "Physical layer security for massive MIMO: An overview on passive eavesdropping and active attacks," *IEEE Commun. Mag.*, vol. 53, no. 6, pp. 21–27, June 2015.

[8] R. K. Chang, "Defending against flooding-based distributed denial-of-service attacks: A tutorial," *IEEE Commun. Mag.*, vol. 40, no. 10, pp. 42–51, Oct. 2002.

[9] A. Vempaty, L. Tong, and P. K. Varshney, "Distributed inference with byzantine data: State-of-the-art review on data falsification attacks," *IEEE Signal Process. Mag.*, vol. 30, no. 5, pp. 65–75, Sept. 2013.

[10] W. Yang, W. Luo, and X. Zhang, "Distributed secure state estimation under stochastic linear attacks," *IEEE Trans. Netw. Sci. Eng.*, vol. 8, no. 3, pp. 2036–2047, July-Sept. 1 2021.

[11] W. Yang, Y. Zhang, G. Chen, C. Yang, and L. Shi, "Distributed filtering under false data injection attacks," *Elsevier Automatica*, vol. 102, pp. 34–44, April 2019.

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

[13] H. Wu, B. Zhou, and C. Zhang, "Secure distributed estimation against data integrity attacks in internet-of-things systems," *IEEE Trans. Autom. Sci. Eng.*, pp. 1–14, 2021.

[14] H. Song, D. Ding, H. Dong, and Q.-L. Han, "Distributed maximum correntropy filtering for stochastic nonlinear systems under deception attacks," *IEEE Trans. Cybern.*, vol. 52, no. 5, pp. 3733–3744, May 2022.

[15] A. Moradi, N. K. D. Venkategowda, S. P. Talebi, and S. Werner, "Privacy-preserving distributed Kalman filtering," *IEEE Trans. Signal Process.*, vol. 70, pp. 3074–3089, June 2022.

[16] A. Moradi, N. K. Venkategowda, and S. Werner, "Coordinated data-falsification attacks in consensus-based distributed Kalman filtering," in *Proc. 8th IEEE Int. Workshop Comput. Advances Multi-Sensor Adaptive Process.*, 2019, pp. 495–499.

[17] A. Moradi, N. K. Venkategowda, S. P. Talebi, and S. Werner, "Distributed Kalman filtering with privacy against honest-but-curious adversaries," in *Proc. 55th IEEE Asilomar Conf. Signals, Syst., Comput.*, 2021, pp. 790–794.

[18] A. Moradi, N. K. D. Venkategowda, S. P. Talebi, and S. Werner, "Securing the distributed Kalman filter against curious agents," in *Proc. 24th IEEE Int. Conf. Inf. Fusion*, 2021, pp. 1–7.

[19] M. N. Kurt, Y. Yılmaz, and X. Wang, "Distributed quickest detection of cyber-attacks in smart grid," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 8, pp. 2015–2030, Aug. 2018.

[20] M. N. Kurt, Y. Yilmaz, and X. Wang, "Real-time detection of hybrid and stealthy cyber-attacks in smart grid," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 2, pp. 498–513, Feb. 2019.

[21] M. Aktukmak, Y. Yilmaz, and I. Uysal, "Sequential attack detection in recommender systems," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 3285–3298, Apr. 2021.

[22] D. Ding, Q.-L. Han, Z. Wang, and X. Ge, "Recursive filtering of distributed cyber-physical systems with attack detection," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 51, no. 10, pp. 6466–6476, Oct. 2021.

[23] C.-Z. Bai, V. Gupta, and F. Pasqualetti, "On Kalman filtering with compromised sensors: Attack stealthiness and performance bounds," *IEEE Trans. Autom. Control*, vol. 62, no. 12, pp. 6641–6648, Dec. 2017.

[24] Z. Guo, D. Shi, K. H. Johansson, and L. Shi, "Optimal linear cyber-attack on remote state estimation," *IEEE Control Netw. Syst.*, vol. 4, no. 1, pp. 4–13, March 2016.

[25] Y. Chen, S. Kar, and J. M. Moura, "Optimal attack strategies subject to detection constraints against cyber-physical systems," *IEEE Control Netw. Syst.*, vol. 5, no. 3, pp. 1157–1168, Sept. 2017.

[26] X.-X. Ren, G. Yang, and X.-G. Zhang, "Statistical-based optimal-stealthy attack under stochastic communication protocol: An application to networked pmsm systems," *IEEE Trans. Ind. Electron.*, 2022.

[27] Y. Mo and B. Sinopoli, "On the performance degradation of cyber-physical systems under stealthy integrity attacks," *IEEE Trans. Autom. Control*, vol. 61, no. 9, pp. 2618–2624, Sept. 2016.

[28] R. Deng, G. Xiao, R. Lu, H. Liang, and A. V. Vasilakos, "False data injection on state estimation in power systems—attacks, impacts, and defense: A survey," *IEEE Trans. Ind. Informat.*, vol. 13, no. 2, pp. 411–423, Apr. 2017.

[29] F. Li and Y. Tang, "False data injection attack for cyber-physical systems with resource constraint," *IEEE Trans. Cybern.*, vol. 50, no. 2, pp. 729–738, Feb. 2020.

[30] V. Kekatos and G. B. Giannakis, "Distributed robust power system state estimation," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 1617–1626, May 2013.

[31] B. Kailkhura, S. Brahma, and P. K. Varshney, "Data falsification attacks on consensus-based detection systems," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 3, no. 1, pp. 145–158, March 2016.

[32] X. He, X. Ren, H. Sandberg, and K. H. Johansson, "How to secure distributed filters under sensor attacks," *IEEE Trans. Autom. Control*, vol. 67, no. 6, pp. 2843–2856, 2022.

[33] L. An and G.-H. Yang, "Byzantine-resilient distributed state estimation: A min-switching approach," *Elsevier Automatica*, vol. 129, p. 109664, July 2021.

[34] Y. Chen, S. Kar, and J. M. Moura, "Resilient distributed estimation: Sensor attacks," *IEEE Trans. Autom. Control*, vol. 64, no. 9, pp. 3772–3779, Sept. 2018.

[35] Y. Chen, S. Kar, and J. M. F. Moura, "Resilient distributed parameter estimation with heterogeneous data," *IEEE Trans. Signal Process.*, vol. 67, no. 19, pp. 4918–4933, Oct. 2019.

[36] H. Song, P. Shi, C.-C. Lim, W.-A. Zhang, and L. Yu, "Attack and estimator design for multi-sensor systems with undetectable adversary," *Elsevier Automatica*, vol. 109, p. 108545, Nov. 2019.

[37] H. Song, P. Shi, W.-A. Zhang, C.-C. Lim, and L. Yu, "Distributed $h_\infty$ estimation in sensor networks with two-channel stochastic attacks," *IEEE Trans. Cybern.*, vol. 50, no. 2, pp. 465–475, Feb. 2020.

[38] Y. Shi and Y. Wang, "Online secure state estimation of multiagent systems using average consensus," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 52, no. 5, pp. 3174–3186, May 2022.

[39] J. G. Lee, J. Kim, and H. Shim, "Fully distributed resilient state estimation based on distributed median solver," *IEEE Trans. Autom. Control*, vol. 65, no. 9, pp. 3935–3942, Sept. 2020.

[40] M. Fauser and P. Zhang, "Resilient homomorphic encryption scheme for cyber-physical systems," in *Proc. 60th IEEE Conf. Decis. and Control (CDC)*, 2021, pp. 5634–5639.

[41] Y. Ni, J. Wu, L. Li, and L. Shi, "Multi-party dynamic state estimation that preserves data and model privacy," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 2288–2299, Jan. 2021.

[42] J. Zhou, W. Ding, and W. Yang, "A secure encoding mechanism against deception attacks on multi-sensor remote state estimation," *IEEE Trans. Inf. Forensics Security*, vol. 17, pp. 1959–1969, 2022.

[43] M. Fauser and P. Zhang, "Resilience of cyber-physical systems to covert attacks by exploiting an improved encryption scheme," in *Proc. 59th IEEE Conf. Decis. and Control (CDC)*, 2020, pp. 5489–5494.

[44] H. Lin, Z. T. Kalbarczyk, and R. K. Iyer, "Raincoat: Randomization of network communication in power grid cyber infrastructure to mislead attackers," *IEEE Trans. Smart Grid*, vol. 10, no. 5, pp. 4893–4906, Sept. 2019.

[45] A. Mitra and S. Sundaram, "Byzantine-resilient distributed observers for lti systems," *Elsevier Automatica*, vol. 108, p. 108487, Oct. 2019.

[46] S. Rajput, H. Wang, Z. Charles, and D. Papailiopoulos, "DETOX: A redundancy-based framework for faster and more robust gradient aggregation," in *Proc. NIPS*, vol. 32, 2019, p. 10320–10330.

[47] P. Krishnamurthy and F. Khorrami, "Resilient redundancy-based control of cyber–physical systems through adaptive randomized switching," *Systems & Control Letters*, vol. 158, p. 105066, Dec. 2021.

[48] A. Mitra, F. Ghawash, S. Sundaram, and W. Abbas, "On the impacts of redundancy, diversity, and trust in resilient distributed state estimation," *IEEE Trans. Control Netw. Syst.*, vol. 8, no. 2, pp. 713–724, June 2021.

[49] K. Ryu and J. Back, "Distributed Kalman-filtering: Distributed optimization viewpoint," in *Proc. 58th IEEE Conf. Decis. and Control*, 2019, pp. 2640–2645.

[50] W. Ben-Ameur, P. Bianchi, and J. Jakubowicz, "Robust distributed consensus using total variation," *IEEE Trans. Autom. Control*, vol. 61, no. 6, pp. 1550–1564, June 2016.

[51] J. Peng, W. Li, and Q. Ling, "Byzantine-robust decentralized stochastic optimization over static and time-varying networks," *Elsevier Signal Process.*, vol. 183, p. 108020, June 2021.

[52] J. Peng, W. . Li, and Q. Ling, "Variance reduction-boosted Byzantine robustness in decentralized stochastic optimization," in *Proc. 47th IEEE Int. Conf. Acoust., Speech and Signal Process.*, 2022, pp. 4283–4287.

[53] Y. Chen, S. Kar, and J. M. Moura, "Optimal attack strategies subject to detection constraints against cyber-physical systems," *IEEE Control Netw. Syst.*, vol. 5, no. 3, pp. 1157–1168, Sept. 2018.

[54] C.-Z. Bai, V. Gupta, and F. Pasqualetti, "On Kalman filtering with compromised sensors: Attack stealthiness and performance bounds," *IEEE Trans. Autom. Control*, vol. 62, no. 12, pp. 6641–6648, Dec. 2017.

**Ashkan Moradi** received the M.Sc. degree in Telecommunication Networks from University of Tehran, Iran, in 2016. He is currently pursuing a Ph.D. degree at the Department of Electronic Systems at the Norwegian University of Science and Technology (NTNU). His expertise and research interests include distributed learning and estimation algorithms in resource-constrained networks, with an emphasis on agent privacy and data security. From June 2022 to Aug. 2022, he was visiting researcher at the Technical University of Munich in Germany.

**Naveen K. D. Venkategowda** (S'12–M'17) received the B.E. degree in electronics and communication engineering from Bangalore University, Bengaluru, India, in 2008, and the Ph.D. degree in electrical engineering from Indian Institute of Technology, Kanpur, India, in 2016. He is currently an Universitetslektor at the Department of Science and Technology, Linköping University, Sweden. From Oct. 2017 to Feb. 2021, he was postdoctoral researcher at the Department of Electronic Systems, Norwegian University of Science and Technology, Trondheim, Norway. He was a Research Professor at the School of Electrical Engineering, Korea University, South Korea from Aug. 2016 to Sep. 2017. He was a recipient of the TCS Research Fellowship (2011-15) from TCS for graduate studies in computing sciences and the ERCIM Alain Bensoussan Fellowship in 2017.

**Stefan Werner** (F'23) received the M.Sc. degree in electrical engineering from the Royal Institute of Technology, Stockholm, Sweden, in 1998, and the D.Sc. degree (Hons.) in electrical engineering from the Signal Processing Laboratory, Helsinki University of Technology, Espoo, Finland, in 2002. He is currently a Professor at the Department of Electronic Systems, Norwegian University of Science and Technology (NTNU), Director of IoT@NTNU, and Adjunct Professor at Aalto University in Finland. He was a visiting Melchor Professor with the University of Notre Dame during the summer of 2019 and an Adjunct Senior Research Fellow with the Institute for Telecommunications Research, University of South Australia, from 2014 to 2020. He held an Academy Research Fellowship, funded by the Academy of Finland, from 2009 to 2014. His research interests include adaptive and statistical signal processing, wireless communications, and security and privacy in cyber-physical systems. He is a member of the editorial boards for the EURASIP Journal of Signal Processing and the IEEE Transactions on Signal and Information Processing over Networks. Dr. Werner is a Fellow of the IEEE.