Coordinated Data-Falsification Attacks in Consensus-based Distributed Kalman Filtering

Ashkan Moradi, Naveen K. D. Venkategowda, Stefan Werner
Department of Electronic Systems, NTNU, Trondheim, Norway
Email: {ashkan.moradi, naveen.dv, stefan.werner}@ntnu.no

***Abstract*—This paper considers consensus-based distributed Kalman filtering subject to data-falsification attack, where Byzantine agents share manipulated data with their neighboring agents. The attack is assumed to be coordinated among the Byzantine agents and follows a linear model. The goal of the Byzantine agents is to maximize the network-wide estimation error while evading false-data detectors at honest agents. To that end, we propose a joint selection of Byzantine agents and covariance matrices of attack sequences to maximize the network-wide estimation error subject to constraints on stealthiness and the number of Byzantine agents. The attack strategy is then obtained by employing block-coordinate descent method via Boolean relaxation and backward stepwise based subset selection method. Numerical results show the efficiency of the proposed attack strategy in comparison with other naive and uncoordinated attacks.**

I. Introduction

The adoption of internet of things (IoT) is rapidly growing with applications in security, environmental monitoring, and smart infrastructure [1]. IoT employs distributed signal processing algorithms in which an individual agent exchanges information with its neighboring agents for inference tasks such as event detection, tracking, and parameter estimation. Limited computational and energy resources at the IoT devices and the distributed nature of IoT render them vulnerable to cyber-security threats and malicious attacks from adversaries [2]. Thus, attack and defense mechanisms for secure distributed inference in IoT has garnered significant attention recently.

In data-falsification attacks, Byzantine agents inject malicious data or share manipulated information to decrease the system performance [3]. In such scenarios, the main challenges for distributed algorithms are trustworthiness of local information and resilient inference during attacks. To mitigate data-falsification attacks in distributed detection, adaptive design of local fusion rules to detect Byzantine agents was proposed in [4] and audit-bit based architecture where sensors transmit their decision via local groups in addition to direct communication with fusion center was presented in [5]. The authors in [6] proposed an attack detection procedure by employing reliable innovation data from the neighboring sensors in the distributed estimation process. In [7] weighted combination of local innovations and the information shared by neighbors is proposed for robust parameter estimation in presence of attacks. Joint attack detection and secure estimation methods have been proposed in [8] and [9]. The authors in [10] consider secure estimation for a networked cyber-physical system (CPS) under simultaneous false data injection and jamming attacks and propose a two-step attack detection mechanism and a measurement output model refinement to overcome the attacks.

On the other hand, knowledge of the optimal attack strategy and its impacts on the performance of IoT plays an important role in secure inference. It helps to understand the system behavior in presence of attacks, to identify critical links and agents, and to determine the regime in which the IoT no longer satisfies the operational goals. In this context, the trade-off between the detection performance with no attackers and the worst-case detection performance with an attacker was studied in [11] for hypothesis testing. For remote state estimation setting, optimal jamming policies for attacking the communication channels between sensors and fusion center to maximize the estimation error was proposed in [12] and optimal linear deception attack, which can successfully bypass a χ^2 false data detector, was presented in [13]. In [14] the mean square error (MSE) performance of single sensor Kalman filter with data-falsification attacks was analyzed considering the Kullback-Leibler (KL) divergence as a measure of attack stealthiness. Similarly, in [15] it was shown that with KL divergence as the stealth metric, the worst-case linear attack strategy that maximizes the estimation error covariance is a zero-mean Gaussian distributed attack sequence. In [16], the authors propose algorithms to design attack sequence to move the state of a CPS to a target state while satisfying the probability of detection constraints. These works [12]–[16] are limited to single sensor scenarios or centralized state estimation problems. Further, the performance and behavior of distributed state estimation with Byzantine agents are not addressed in the existing literature.

In this paper, we investigate the performance of consensus-based distributed Kalman filtering in presence of Byzantine agents. Assuming a linear attack model, we propose joint selection of Byzantine agents and their attack sequences that maximize the network-wide estimation error subject to constraints on stealthiness and the number of Byzantine agents. This results in an NP-hard optimization problem. Hence, we obtain suboptimal solutions by solving a sequence of semidefinite program (SDP) through the block-coordinate descent method and Boolean relaxation of the NP-hard optimization problem. To benchmark the proposed method, we present a backward stepwise subset selection based algorithm to determine the best set of Byzantine agents that maximizes the error.

Notations: Transpose and trace are denoted by $(\cdot)^T$ and $\mathsf{tr}(\cdot)$, the identity matrix of size n is represented by $\mathbf{I}_n$, the ones vector of length L is denoted by $\mathbf{1}_L$, whereas $\otimes$ denotes Kronecker product. Positive semidefinite matrix is represented by $\mathbf{A} \succeq 0$ and $\sup$ denotes the supremum. Matrices $\mathsf{diag}(\mathbf{a})$

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

and $\mathsf{diag}(\{\mathbf{A}_i\}_{i=1}^L)$ denote diagonal and block-diagonal matrices whose respective diagonals are the elements of vector $\mathbf{a}$ and matrices $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_L$.

II. System Model and Problem Formulation

Consider a connected multi-agent network of $L \in \mathbb{N}$ agents that collectively aim to estimate the state vector sequence $\{\mathbf{x}(k),\, k = 1, 2 \ldots\}$ from local observations $\{\mathbf{y}_i(k),\, k = 1, 2 \ldots,\, i = 1, 2 \ldots, L\}$ at the agents. The network is modeled as an undirected graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of all agents of the network with $|\mathcal{V}| = L$, and $\mathcal{E}$ is the edge set that represents the communication links between the agents. The neighbor set $\mathcal{N}_i$ comprises all the agents that are connected to i within one hop and excludes the agent itself. The network adjacency matrix is denoted by $\mathbf{E}$ and the graph Laplacian is defined as $\mathbf{L} = \mathsf{diag}(\{|\mathcal{N}_i|\}_{i=1}^L) - \mathbf{E}$.

A. Distributed Filtering

The state vector and observation sequences at the ith agent are characterized by the state-space model

$$\begin{aligned} \mathbf{x}(k+1) &= \mathbf{A}\mathbf{x}(k) + \mathbf{w}(k) \\ \mathbf{y}_i(k) &= \mathbf{H}_i\mathbf{x}(k) + \mathbf{v}_i(k), \end{aligned} \tag{1}$$

where $\mathbf{A} \in \mathbb{R}^{m\times m}$ is the state-transition matrix, $\mathbf{H}_i \in \mathbb{R}^{n\times m}$ is the observation matrix at agent i, whereas $\mathbf{w}(k)$ and $\mathbf{v}_i(k)$ are mutually independent zero-mean Gaussian processes with covariance matrices $\mathbf{Q} \in \mathbb{R}^{m\times m}$ and $\mathbf{R}_i \in \mathbb{R}^{n\times n}$, respectively.

The agents employ the consensus-based distributed Kalman filter to estimate $\mathbf{x}(k)$ in a collaborative manner [17]. The state estimate at agent i is given by

$$\begin{aligned} \hat{\mathbf{x}}_i(k+1) = \mathbf{A}\hat{\mathbf{x}}_i(k) + \mathbf{K}_i(k)\big(\mathbf{y}_i(k) - \mathbf{H}_i\hat{\mathbf{x}}_i(k)\big) \\ - \varepsilon\mathbf{A}\textstyle\sum_{j\in\mathcal{N}_i}\big(\hat{\mathbf{x}}_i(k) - \bar{\mathbf{x}}_j(k)\big), \end{aligned} \tag{2}$$

where $\mathbf{K}_i(k) \in \mathbb{R}^{m\times n}$ is the Kalman gain at agent i, ε is the consensus gain chosen as $0 \le \varepsilon \le 1/\max_i |\mathcal{N}_i|$, and $\{\bar{\mathbf{x}}_j(k)\}_{j\in\mathcal{N}_i}$ are estimates shared by the agents in the neighborhood set $\mathcal{N}_i$.

The optimal Kalman gain $\mathbf{K}_i(k)$ in (2) is found by minimizing the trace of the estimation error covariance $\mathbf{P}_i(k) \triangleq \mathbb{E}\{\mathbf{e}_i(k)\mathbf{e}_i^T(k)\}$, where the estimation error $\mathbf{e}_i(k)$ at agent i evolves as

$$\begin{aligned} \mathbf{e}_i(k+1) \triangleq \hat{\mathbf{x}}_i(k) - \mathbf{x}(k) = (\mathbf{A} - \mathbf{K}_i(k)\mathbf{H}_i)\mathbf{e}_i(k) - \mathbf{w}(k) \\ + \mathbf{K}_i(k)\mathbf{v}_i(k) - \varepsilon\mathbf{A}\textstyle\sum_{j\in\mathcal{N}_i}\big(\mathbf{e}_i(k) - \mathbf{e}_j(k)\big). \end{aligned} \tag{3}$$

After some calculations, the estimation error covariance can be expressed as

$$\begin{aligned} \mathbf{P}_i(k+1) &= \mathbf{F}_i(k)\mathbf{P}_i(k)\mathbf{F}_i^T(k) + \mathbf{Q} + \mathbf{K}_i(k)\mathbf{R}_i\mathbf{K}_i^T(k) \\ &- \varepsilon\mathbf{F}_i(k)\textstyle\sum_{s\in\mathcal{N}_i}\big(\mathbf{P}_i(k) - \mathbf{P}_{is}(k)\big)\mathbf{A}^T \\ &- \varepsilon\mathbf{A}\textstyle\sum_{r\in\mathcal{N}_i}\big(\mathbf{P}_i(k) - \mathbf{P}_{ri}(k)\big)\mathbf{F}_i^T(k) \\ &+ \varepsilon^2\sum_{r\in\mathcal{N}_i}\sum_{s\in\mathcal{N}_i}\mathbf{A}\Big(\mathbf{P}_i(k) - \mathbf{P}_{is}(k) - \mathbf{P}_{ri}(k) + \mathbf{P}_{rs}(k)\Big)\mathbf{A}^T, \end{aligned} \tag{4}$$

where $\mathbf{P}_{ij}(k) \triangleq \mathbb{E}\{\mathbf{e}_i(k)\mathbf{e}_j^T(k)\}$ and $\mathbf{F}_i(k) = \mathbf{A} - \mathbf{K}_i(k)\mathbf{H}_i$. Thus, the optimal Kalman gain, which is found by differentiating the trace of (4) with respect to $\mathbf{K}_i(k)$, is given by

$$\mathbf{K}_i^*(k) = \mathbf{A}\Big(\mathbf{P}_i(k) - \varepsilon\sum_{j\in\mathcal{N}_i}\big(\mathbf{P}_i(k) - \mathbf{P}_{ji}(k)\big)\Big)\mathbf{H}_i^T\mathbf{M}_i^{-1}(k), \tag{5}$$

where $\mathbf{M}_i(k) = \mathbf{H}_i\mathbf{P}_i(k)\mathbf{H}_i^T + \mathbf{R}_i$.

B. Attack Model

In the following it is assumed that a subset $\mathcal{B} \subseteq \mathcal{V}$ with $|\mathcal{B}| \le L$ are Byzantine agents. In contrast to the "honest agents", Byzantines share a falsified version of their state estimate with their neighbors to deteriorate the network-wide estimation performance [3]. Byzantine agent $j \in \mathcal{B}$ shares a modified state estimate $\hat{\mathbf{x}}_j(k) + \boldsymbol{\delta}_j(k)$ instead of $\hat{\mathbf{x}}_j(k)$ for $k \ge k_0$, where k_0 is the time instant when attack is initiated. Consequently, for $k \ge k_0$, the local estimates used for consensus building in (2) can be expressed as

$$\bar{\mathbf{x}}_j(k) = \begin{cases} \hat{\mathbf{x}}_j(k) + \boldsymbol{\delta}_j(k) & j \in \mathcal{B} \\ \hat{\mathbf{x}}_j(k) & j \notin \mathcal{B}, \end{cases} \tag{6}$$

where $\boldsymbol{\delta}_j(k) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_j)$ denotes the *data-falsification sequence*. Assuming a coordinated attack by the Byzantine agents, the augmented attack sequence across the network is given by

$$\boldsymbol{\delta}(k) \triangleq [\boldsymbol{\delta}_1^T(k), \boldsymbol{\delta}_2^T(k), \ldots \boldsymbol{\delta}_L^T(k)]^T \tag{7}$$

and its covariance matrix is denoted by $\boldsymbol{\Sigma} = \mathbb{E}\{\boldsymbol{\delta}(k)\boldsymbol{\delta}^T(k)\}$. It assumed that the Byzantine agents have knowledge of the network and observation matrices. To maximize the attack stealthiness, $\boldsymbol{\delta}(k)$ is chosen as a zero-mean Gaussian sequence with covariance $\boldsymbol{\Sigma}$ [14]–[16]. The probability of attack-detection is proportional to the covariance of the attack sequence [14]–[16]. Therefore, $\boldsymbol{\Sigma}$ is limited to $\mathsf{tr}(\boldsymbol{\Sigma}) \le \eta$, where η captures the stealthiness of the attack.

C. Problem Statement

The main objective of the Byzantine attack is to maximize the network-wide mean squared error (NMSE) defined as

$$\mathsf{NMSE} = \limsup_{K\to\infty}\frac{1}{K}\textstyle\sum_{k=1}^K\sum_{i=1}^L\mathsf{tr}\big(\mathbf{P}_i(k)\big) \tag{8}$$

while still maintaining a desired level of stealthiness. Due to limited resources only a subset of agents can be Byzantines, which is denoted by $\mathcal{B}$. We need to decide the subset of agents that participate in the attack and determine the covariance matrices $\boldsymbol{\Sigma}_j$, $j \in \mathcal{B}$, of the corresponding falsification sequences. To that end, we introduce the Boolean variable $z_j = 1$ if $j \in \mathcal{B}$ and zero otherwise, and define the selection vector $\mathbf{z} = [z_1, z_2, \ldots, z_L]^T$ [18]. The optimal attack strategy can be expressed as an optimization problem given by

$$\begin{aligned} \underset{\boldsymbol{\Sigma},\,\mathbf{z}}{\max.} \quad & \mathsf{NMSE} \\ \text{s. t.} \quad & \textstyle\sum_{j\in\mathcal{B}}\mathsf{tr}(\boldsymbol{\Sigma}_j) \le \eta, \\ & \boldsymbol{\Sigma} \succeq 0, \\ & \mathbf{z} \in \{0,\,1\}^L, \quad \mathbf{1}^T\mathbf{z} = B, \end{aligned} \tag{9}$$

where the first constraint is related to the stealthiness, i.e., the ability to evade detection and the last constraint limits

the number of Byzantine agents to $|\mathcal{B}| = B$. The parameter η is employed to restrict the total power of the falsification sequences and satisfy the detection-avoidance target.

In the next section, we compute the network-wide mean squared error as a function of the attack sequence covariance matrices and propose different methods for joint design of the attack sequence and subset of Byzantines with the aim of maximizing the error.

III. Joint Selection of Byzantine Agents and Design of Attack Sequences

To solve the problem in (9), we first derive the expression for the objective function to capture the NMSE. To that end, define the network-wide estimation error in presence of Byzantine attack after $k \geq k_0$ as

$$\bar{\mathbf{e}}(k) \triangleq [\bar{\mathbf{e}}_1^T(k), \bar{\mathbf{e}}_2^T(k), \ldots, \bar{\mathbf{e}}_L^T(k)]^T, \tag{10}$$

where the error at the ith agent is given by

$$\begin{aligned}\bar{\mathbf{e}}_i(k+1) = (\mathbf{A} - \mathbf{K}_i(k)\mathbf{H}_i)\bar{\mathbf{e}}_i(k) - \mathbf{w}(k) + \mathbf{K}_i(k)\mathbf{v}_i(k) \\ - \varepsilon\mathbf{A}\textstyle\sum_{j\in\mathcal{N}_i}\big(\bar{\mathbf{e}}_i(k) - \bar{\mathbf{e}}_j(k) - \boldsymbol{\delta}_j(k)\big).\end{aligned} \tag{11}$$

Defining $\boldsymbol{\Gamma} = \mathbf{E}\ \mathsf{diag}(\mathbf{z}) \otimes \mathbf{A}$, the evolution of the network estimation error can be expressed as

$$\bar{\mathbf{e}}(k+1) = \bar{\mathbf{A}}(k)\bar{\mathbf{e}}(k) + \bar{\mathbf{b}}(k) + \varepsilon\boldsymbol{\Gamma}\boldsymbol{\delta}(k), \tag{12}$$

where $\bar{\mathbf{A}}(k) = (\mathbf{I}_L - \varepsilon\mathbf{L}) \otimes \mathbf{A} - \mathsf{diag}(\{\hat{\mathbf{K}}_i(k)\mathbf{H}_i\}_{i=1}^L)$, $\hat{\mathbf{K}}_i(k)$ is the Kalman gain assuming the statistics of the attack sequence is known, and

$$\bar{\mathbf{b}}(k) = \mathsf{diag}(\{\hat{\mathbf{K}}_i(k)\mathbf{v}_i(k)\}_{i=1}^L) - \mathbf{1}_L \otimes \mathbf{w}(k).$$

From (12), the covariance matrix of the error $\hat{\mathbf{P}}(k+1) \triangleq \mathbb{E}\{\mathbf{e}(k+1)\mathbf{e}^T(k+1)\}$ is given by

$$\hat{\mathbf{P}}(k+1) = \bar{\mathbf{A}}(k)\hat{\mathbf{P}}(k)\bar{\mathbf{A}}^T(k) + \bar{\mathbf{Q}}(k) + \varepsilon^2\boldsymbol{\Gamma}\boldsymbol{\Sigma}\boldsymbol{\Gamma}^T, \tag{13}$$

where $\bar{\mathbf{Q}}(k) = \mathsf{diag}(\{\hat{\mathbf{K}}_i(k)\mathbf{R}_i\hat{\mathbf{K}}_i^T(k)\}_{i=1}^L + \mathbf{1}_L\mathbf{1}_L^T \otimes \mathbf{Q}$. The optimal Kalman gain that minimizes $\mathsf{tr}(\hat{\mathbf{P}}(k))$ in (13) can obtained as

$$\hat{\mathbf{K}}_i(k) = \mathbf{A}\Big(\hat{\mathbf{P}}_i(k) - \varepsilon\sum_{j\in\mathcal{N}_i}\big(\hat{\mathbf{P}}_i(k) - \hat{\mathbf{P}}_{ji}(k)\big)\Big)\mathbf{H}_i^T\hat{\mathbf{M}}_i^{-1}(k), \tag{14}$$

where $\hat{\mathbf{M}}_i(k) = \mathbf{H}_i\hat{\mathbf{P}}_i(k)\mathbf{H}_i^T + \mathbf{R}_i$. In contrast to (4) and (5), (13) and (14) capture the error dynamics in presence of a Byzantine attack.

Assuming that the network is connected, $(\mathbf{A}, \bar{\mathbf{Q}}^{1/2})$ is controllable, and $(\mathbf{A}, \mathbf{H}_i)$ is observable, it can be shown that $\lim_{k\to\infty}\hat{\mathbf{P}}(k) = \hat{\mathbf{P}}$ i.e., $\hat{\mathbf{P}}(k)$ converges to a bounded value. In other words, there exists a matrix $\hat{\mathbf{K}}_i(k)$ such that $\hat{\mathbf{P}}(k)$ is bounded and converges to a unique positive definite matrix for all k and any initial non-negative symmetric matrix. Since obtaining a closed form expression for the covariance matrix of the actual error in (11) induced by the attack is intractable, we employ $\mathsf{tr}(\hat{\mathbf{P}})$ as a proxy to the objective function. Here $\mathsf{tr}(\hat{\mathbf{P}})$ is a lower bound for the actual NMSE.

The solution to the Riccati equation in (13) can be obtained from an SDP [19]. Motivated by this fact and substituting $\mathsf{NMSE} = \mathsf{tr}(\hat{\mathbf{P}})$ in (9), we express the joint Byzantine agent selection and attack design optimization problem as

$$\begin{aligned}\mathcal{P}: \ \max_{\mathbf{X},\boldsymbol{\Sigma},\mathbf{z}} \ & \mathsf{tr}(\mathbf{X}) \\ \text{s. t.} \ & \mathbf{X} \succeq \bar{\mathbf{A}}\mathbf{X}\bar{\mathbf{A}}^T + \bar{\mathbf{Q}} + \varepsilon^2\boldsymbol{\Gamma}\boldsymbol{\Sigma}\boldsymbol{\Gamma}^T, \\ & \boldsymbol{\Gamma} = \mathbf{E}\ \mathsf{diag}(\mathbf{z}) \otimes \mathbf{A}, \\ & \mathbf{X} \succeq 0 \\ & \textstyle\sum_{j\in\mathcal{B}}\mathsf{tr}(\boldsymbol{\Sigma}_j) \leq \eta, \quad \boldsymbol{\Sigma} \succeq 0, \\ & \mathbf{1}^T\mathbf{z} \leq B, \quad z_i \in \{0,1\}, \quad i = 1, \ldots L.\end{aligned} \tag{15}$$

The above problem is NP-hard [20], and difficult to solve due to the non-convex quadratic terms in the first constraint. In the subsequent sections we propose different methods to find a suboptimal solution to the above problem.

A. Block-Coordinate Descent (BCD) based Approach

The problem in (15) is non-convex due to the Boolean variables. To circumvent this, we relax the Boolean constraint $z_i \in \{0,1\}$ to a linear inequality constraint $0 \leq z_i \leq 1$. We see that for a given $\mathbf{z}$ or $\boldsymbol{\Sigma}$ the problem (15) is an SDP, as its first constraint is convex. Therefore, we employ the block-coordinate descent (BCD) method where $(\mathbf{X}, \boldsymbol{\Sigma})$ and $(\mathbf{X}, \mathbf{z})$ are alternately optimized with the other variable fixed. Applying the trace operator on both sides of the convergence constraint leads to a linear approximation with respect to $\mathbf{z}$ and $\boldsymbol{\Sigma}$. The proposed approach starts with an arbitrary $\mathbf{z}_0$ as initial condition and its first step is given by

$$\begin{aligned}\mathcal{P}_1: \ \max_{\mathbf{X},\boldsymbol{\Sigma}} \ & \mathsf{tr}(\mathbf{X}) \\ \text{s. t.} \ & \mathsf{tr}(\mathbf{X}) \succeq \mathsf{tr}(\bar{\mathbf{A}}\mathbf{X}\bar{\mathbf{A}}^T + \bar{\mathbf{Q}}) + \varepsilon^2\mathsf{tr}(\boldsymbol{\Gamma}\boldsymbol{\Sigma}\boldsymbol{\Gamma}^T), \\ & \mathbf{X} \succeq 0, \\ & \textstyle\sum_{j\in\mathcal{B}}\mathsf{tr}(\boldsymbol{\Sigma}_j) \leq \eta, \quad \boldsymbol{\Sigma} \succeq 0.\end{aligned} \tag{16}$$

The second step of the BCD approach is to determine the Byzantine agents by solving

$$\begin{aligned}\mathcal{P}_2: \ \max_{\mathbf{X},\mathbf{z}} \ & \mathsf{tr}(\mathbf{X}) \\ \text{s. t.} \ & \mathsf{tr}(\mathbf{X}) \succeq \mathsf{tr}(\bar{\mathbf{A}}\mathbf{X}\bar{\mathbf{A}}^T + \bar{\mathbf{Q}}) + \varepsilon^2\mathsf{tr}(\boldsymbol{\Gamma}\boldsymbol{\Sigma}\boldsymbol{\Gamma}^T), \\ & \boldsymbol{\Gamma} = \mathbf{E}\ \mathsf{diag}(\mathbf{z}) \otimes \mathbf{A}, \\ & \mathbf{X} \succeq 0, \\ & \mathbf{1}^T\mathbf{z} \leq B, \quad 0 \leq z_i \leq 1, \quad i = 1, \ldots L.\end{aligned} \tag{17}$$

The subproblems (16) and (17) are convex and (16) has an unique solution for a given $\mathbf{z}$. Hence from [21, Theorem 1], we conclude that the proposed algorithm converges to a stationary point. The steps in (16) and (17) reduce the problem in (15) to that of solving a sequence of SDPs, which can be efficiently solved by interior-point methods.

The optimal $\mathbf{z}^* \in [0,1]^L$ is not Boolean due to the relaxation in (17). Hence, we recover a feasible solution $\mathbf{z}'$ of (15) by sorting the elements of $\mathbf{z}^*$ in descending order and set $z_i' = 1$ for the agents corresponding to the $|\mathcal{B}| = B$ largest elements.

Algorithm 1 Backward Stepwise Selection based Attack

Initialize: $\mathcal{B}_L = \mathcal{V}$,

1: **for** $j = L$ downto $B+1$ **do**
2: Determine $l_j^* = \arg\max_{l\in\mathcal{B}_j} U(\mathcal{B}_j\backslash\{l\})$.
3: Update $\mathcal{B}_{j-1} = \mathcal{B}_j\backslash\{l_j^*\}$.
4: **end for**
5: Set attack strategy $z_i = 1$ if $i \in \mathcal{B}_B$ else $z_i = 0$.
6: Find optimal attack sequence covariance matrix from (16).

B. Backward Stepwise Selection based Attack Strategy

For a given attack selection vector $\mathbf{z}$, the problem in (15) is an SDP. Hence, instead of relaxing the Boolean constraints, we employ an improved greedy search based method to determine the set of Byzantine agents and then find the corresponding optimal covariance matrices from (16). To select the Byzantine agents, we adopt the backward stepwise selection algorithm [22]. In this method, the algorithm begins by considering all agents as Byzantine i.e., $\mathcal{B} = \mathcal{V}$, and then iteratively removes the agent that contributes least to the overall objective. The algorithm stops when only B most effective agents are remaining. At iteration index j, let $\mathcal{B}_j$ denote the set of Byzantine agents with $|\mathcal{B}_j| = j$ and the corresponding performance of the network is defined as

$$U(\mathcal{B}_j) = \sum_{i\in\mathcal{V}\backslash\mathcal{B}_j} \mathsf{tr}(\hat{\mathbf{P}}_i) + \sum_{i\in\mathcal{B}_j} \mathsf{tr}(\hat{\mathbf{P}}_i), \tag{18}$$

which is computed from (16). The agent l_j^* that contributes lowest to the overall objective $U(\mathcal{B}_j)$ is removed from $\mathcal{B}_j$ at iteration j by determining l_j^* from

$$l_j^* = \arg\max_{l\in\mathcal{B}_j} U(\mathcal{B}_j\backslash\{l\}).$$

The algorithm is terminated when $\mathcal{B}_j$ consists of B agents and the attack sequence covariance matrix is determined from (16) with $z_i = 1$ if $i \in \mathcal{B}_B$ else $z_i = 0$. The proposed backward stepwise selection based attack strategy is summarized in Algorithm 1.

IV. Simulation Results

We consider a randomly generated undirected connected network with $L = 25$ sensor agents, maximum degree of $\Delta = 11$ and consensus gain $\varepsilon = 0.08$. The discrete time system and agent parameters are considered to be $\mathbf{A} = [0.6,\ 0.005;\ 0.25,\ 0.6]$, $\mathbf{Q} = \mathbf{I}_2$, $\mathbf{R}_i = \mathbf{I}_2$ and $\mathbf{H}_i = \mu_i\mathbf{I}_2$, $i = 1, 2, \cdots, L$ with $\mu_i \sim \mathcal{U}(0, 1)$. We set $N = 10$ iterations for the BCD method. The Byzantine agents start falsifying data at time index $k_0 = 20$ and the stealthiness parameter is set to $\eta' = \eta/|\mathcal{B}| = 15$ per Byzantine agent.

The proposed attack strategies are compared with two naive strategies, namely, random selection attack and uniform perturbation attack. The former strategy randomly selects the Byzantine agents, while the associate covariance matrices are obtained from (16). The latter strategy, choose the attack sequence covariance matrices as $\Sigma_j = \frac{P}{m}\mathbf{I}_m$ for all $j \in \mathcal{B}$ and the set of Byzantines are determined from (17). Fig. 1, illustrates the steady-state NMSE for the considered strategies, with $|\mathcal{B}| = 5$. It shows that the proposed methods significantly outperform the naive random and uniform attack strategies.The BCD based approach is computationally less intensive and performs close to the greedy search based method. It can be inferred that the covariance design influences the overall performance more in comparison with Byzantine agent selection.

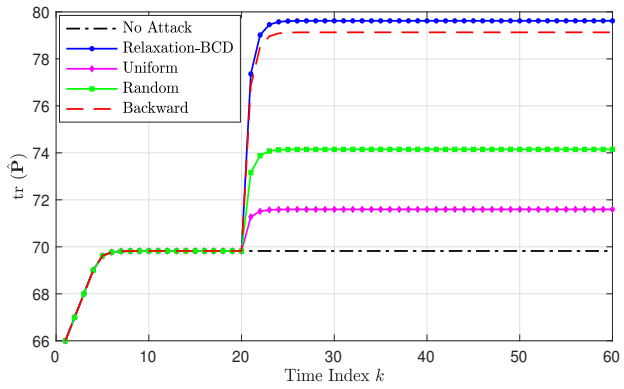

Fig. 1. NMSE for different attack strategies in a network with $L = 25$ agents, $B = |\mathcal{B}| = 5$ Byzantine agents, and stealthiness parameter $\eta' = \eta/B = 15$.

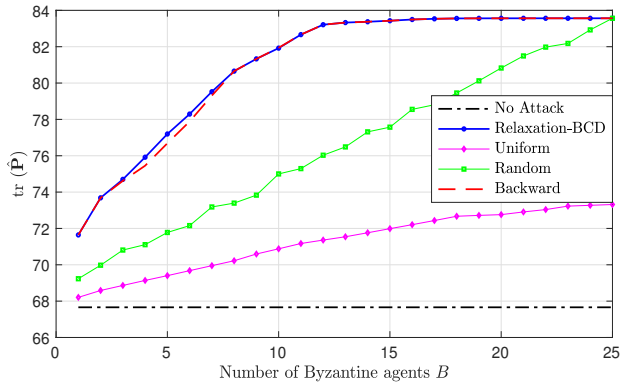

Fig. 2. NMSE versus number of Byzantine agents for a network with $L = 25$ agents and stealthiness parameter $\eta' = \eta/B = 15$.

Fig. 2 shows the NMSE versus the number of Byzantine agents for fixed stealthiness parameter $\eta' = \eta/|\mathcal{B}| = 15$ per Byzantine agent. We observe that the joint attack strategy performs close to the backward stepwise selection based method. When compared with random and uniform attack strategies, the proposed methods cause larger degradation in the NMSE for a fixed number of Byzantine agents.

V. Conclusion

This paper considered a distributed Kalman filter in presence of a coordinated data-falsification attack with Byzantine agents. It has been shown that the optimal set of Byzantine agents and covariance matrices of the falsification data that maximize the network-wide estimation error can be obtained by solving a sequence of semidefinite programs. Further, a greedy strategy for the Byzantine agent selection problem has been presented as an alternative to the Boolean relaxation based block-coordinate descent method. Simulation results demonstrate the efficacy of the proposed attack strategies in comparison with the naive approaches.

REFERENCES

[1] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of things: A survey on enabling technologies, protocols, and applications," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 2347–2376, Fourthquarter 2015.

[2] Y. Chen, S. Kar, and J. M. F. Moura, "The internet of things: Secure distributed inference," *IEEE Signal Process. Mag.*, vol. 35, no. 5, pp. 64–75, Sep. 2018.

[3] 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, Sep. 2013.

[4] 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, Mar. 2017.

[5] W. Hashlamoun, S. Brahma, and P. K. Varshney, "Audit bit based distributed bayesian detection in the presence of Byzantines," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 4, no. 4, pp. 643–655, Dec. 2018.

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

[7] Y. Chen, S. Kar, and J. M. F. Moura, "Resilient distributed estimation: Sensor attacks," *IEEE Trans. Autom. Control*, pp. 1–1, 2019.

[8] N. Forti, G. Battistelli, L. Chisci, S. Li, B. Wang, and B. Sinopoli, "Distributed joint attack detection and secure state estimation," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 4, no. 1, pp. 96–110, Mar. 2018.

[9] Y. Chen, S. Kar, and J. M. F. Moura, "Resilient distributed estimation through adversary detection," *IEEE Trans. Signal Process.*, vol. 66, no. 9, pp. 2455–2469, May 2018.

[10] Y. Guan and X. Ge, "Distributed attack detection and secure estimation of networked cyber-physical systems against false data injection attacks and jamming attacks," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 4, no. 1, pp. 48–59, Mar. 2018.

[11] X. Ren, J. Yan, and Y. Mo, "Binary hypothesis testing with Byzantine sensors: Fundamental tradeoff between security and efficiency," *IEEE Trans. Signal Process.*, vol. 66, no. 6, pp. 1454–1468, Mar. 2018.

[12] X. Ren, J. Wu, S. Dey, and L. Shi, "Attack allocation on remote state estimation in multi-systems: Structural results and asymptotic solution," *Automatica*, vol. 87, pp. 184 – 194, 2018.

[13] 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, Mar. 2017.

[14] C. 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.

[15] Z. Guo, D. Shi, K. H. Johansson, and L. Shi, "Worst-case stealthy innovation-based linear attack on remote state estimation," *Automatica*, vol. 89, pp. 117 – 124, 2018.

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

[17] R. Olfati-Saber, "Kalman-consensus filter : Optimality, stability, and performance," in *Proc. of the 48th IEEE Conference on Decision and Control (CDC)*, Dec. 2009, pp. 7036–7042.

[18] S. Joshi and S. Boyd, "Sensor selection via convex optimization," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 451–462, Feb. 2009.

[19] W. Yang, C. Yang, H. Shi, L. Shi, and G. Chen, "Stochastic link activation for distributed filtering under sensor power constraint," *Automatica*, vol. 75, pp. 109 – 118, 2017.

[20] H. Zhang, R. Ayoub, and S. Sundaram, "Sensor selection for Kalman filtering of linear dynamical systems: Complexity, limitations and greedy algorithms," *Automatica*, vol. 78, pp. 202 – 210, 2017.

[21] M. Hong, M. Razaviyayn, Z.-Q. Luo, and J.-S. Pang, "A unified algorithmic framework for block-structured optimization involving big data: With applications in machine learning and signal processing," *IEEE Signal Process. Mag.*, vol. 33, no. 1, pp. 57–77, Jan. 2016.

[22] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning: With Applications in R*. Springer Publishing Company, Incorporated, 2014, vol. 112.