Distributed Secure State Estimation Using Diffusion Kalman Filters and Reachability Analysis

Amr Alanwar, Hazem Said, and Matthias Althoff

***Abstract*— The tight coupling of information technology with physical sensing and actuation in cyber-physical systems (CPS) has given rise to new security vulnerabilities and attacks with potentially life-threatening consequences. These attacks are designed to transfer the physical system into unstable and insecure states by providing corrupted sensor readings. In this work, we present an approach for distributed secure linear state estimation in the presence of modeling and measurement noise between a network of nodes with pairwise measurements. We provide security against measurement attacks and simplify the traditional distributed secure state estimation problem. Reachability analysis is utilized to establish a security layer providing secure estimate shares for the distributed diffusion Kalman filter. Furthermore, we consider not only attacks on the link level but also on the sensor level. The proposed combined filter protects against measurement and diffusion attacks without requiring specialized hardware or cryptographic techniques. The effectiveness of the approach is demonstrated by a localization example of a rotating target.**

I. Introduction

In the last decade, secure state estimation has attracted attention due to the rise of new security vulnerabilities and attacks at the sensor level with potentially life-threatening consequences. The Office of Science and Technology Policy (OSTP) assigns a high priority to cyber-physical systems (CPS) security, since the recent attacks launched in the cyber domain led to calamitous consensuses during the past decades [1]. For instance, the Maroochy Water Breach [2] made it possible to attack the underlying infrastructure at Maroochy Water Services in Queensland. Also, one popular attack is the Stuxnet attack on Supervisory Control and Data Acquisition (SCADA) systems, which are used in industrial process control [3], [4]; other security issues on SCADA networks are shown in [5]. Attacks on analog sensors which have increasingly become an indispensable part of many modern systems are shown in [6]. Thus, researchers came up with techniques that address the problem of secure state estimation under the sensor, actuators, and communication network attacks. Secure state estimation allows the estimation of the state of the CPS from corrupted/attacked measurements. We review literature focusing on centralized and distributed secure state estimation.

Centralized Secure State Estimation: Different techniques have addressed the problem of secure state estimation against sensor attacks in centralized dynamical systems. Fawzi et al. show the impossibility of accurately reconstructing the state of a system if more than half of the sensors are attacked [7]. The presence of process and measurements noise offers attackers an additional possibility to tamper with CPS sensors, thereby making the detection task more challenging. Another work in [8] uses brute force search for studying the observability of linear systems under adversarial attacks; however, this approach is not applicable to large-scale systems. A practical solution is proposed in [9] that considers jitter, latency and synchronization errors. Graph-theoretic conditions for the detectability of attacks for a noiseless system are shown in [10]. Also, a measure of the stealthiness of attacks in stochastic control systems is proposed in [11].

Observing and recording sensor readings and replying them afterward while carrying out an attack is commonly defined as a replay attack. This kind of attack is considered in [12] where all sensors were attacked and the attacker does not have any model knowledge. Another work considered a stochastic game for detecting replay attacks [13]. Also, a solution for denial of service attacks under Gaussian noise is proposed in [14]. Furthermore, false data injection is solved by proposing an ellipsoidal algorithm where the strategy of the attacker is formulated as a constrained control problem [15].

Distributed Secure State Estimation: Distributed processing mitigates the computation load by getting rid of the fusion center in centralized fusion and estimation. Pasqualetti et al. [16] propose a fully decentralized solution for attack identification. However, they only consider noiseless systems. Distributed secure controllers based on a virtual fractional dynamic surface are designed in [17]. Also, a consensus-based protocol is utilized for distributed secure state estimation in [18].

Contributions: We propose an approach for distributed linear secure state estimation in the presence of measurement noise and modeling errors. By combining the diffusion Kalman filter [19] with reachability analysis [20], we provide a new algorithm for distributed secure state estimation between a network of nodes.

Outline: The paper is organized as follows. After we introduce the problem and the proposed solution in Sec. II, the secure measurement update is presented in Sec. III and secure diffusion in Sec. IV. The applicability of the algorithm is demonstrated in Sec. V. This is followed by a discussion of the algorithm and a conclusion in Sec. VI.

Amr Alanwar and Matthias Althoff are with the Department of Computer Science, Technical University of Munich, D-85748 Garching, Germany. {alanwar,althoff}@tum.de

Hazem Said is with the Department of Computer Engineering, Ain Shams University, 11535 Cairo, Egypt. hazem.said@eng.asu.edu.eg

II. Distributed Secure State Estimation and Proposed Solution

We aim to estimate the full state vector of a system in a distributed fashion by observing physical signals through sensory devices which are under attack. In order to deviate the system from its correct operation, an attacker endeavors to either a) physically attack the sensor environment, b) attack the sensor hardware, c) break the communication links in the CPS, or d) modify the sensor readings (e.g., by delaying packets in time-of-flight based localization). We first discuss our threat model and preliminaries followed by a mathematical formulation of the distributed secure state estimation problem.

A. Threat Model

We consider attackers that directly compromise the readings of various sensor groups and man-in-the-middle attackers that endeavor to modify the data transfer between sensors, as shown in Figure 1.a. Our assumptions are:

- The adversary can corrupt all sensors and has unlimited computational power.
- The selection of attacked sensors is unknown to the system and can change dynamically over time.
- The adversary can commit to unbounded attack values.
- The adversary additionally has no prior knowledge of the system parameters.

B. Preliminaries

We state some preliminaries around the proposed solution.

Definition 1: **(Zonotope)** A zonotope $\mathscr{Z} = \langle c, G \rangle \subset \mathbb{R}^n$ consists of a center $c \in \mathbb{R}^n$ and generator matrix $G \in \mathbb{R}^{n \times e}$. We define G as e generators $g^{(i)} \in \mathbb{R}^n$ $(i = \{1, .., e\})$, where $G = [g^1, ..., g^e]$ [20]. A zonotope is a set

$$\mathscr{Z} = \Big\{ c + \sum_{i=1}^{e} \beta_i g^{(i)} \Big| -1 \leq \beta_i \leq 1 \Big\}. \tag{1}$$

□

Given two zonotopes $\mathscr{Z}_1 = \langle c_1, G_1 \rangle$ and $\mathscr{Z}_2 = \langle c_2, G_2 \rangle$, we define [20]:

1) Minkowski sum:

$$\mathscr{Z}_1 \oplus \mathscr{Z}_2 = \Big\langle c_1 + c_2, [G_1, G_2] \Big\rangle \tag{2}$$

2) Linear map:

$$L\mathscr{Z}_1 = \Big\langle Lc_1, LG_1 \Big\rangle \tag{3}$$

We define the reachable set as the set of possible solution x_i which can be reached at each time step. In this work, reachable sets are represented by zonotopes due to their favorable computational complexity as discussed in [20].

C. System Model

Consider a set of N nodes indexed by $k \in \{0, \ldots, N-1\}$ distributed geographically over some region. We denote the neighborhood of a given node by the set $\mathscr{N}_k$ containing the nodes connected to node k; the size of $\mathscr{N}_k$ is m_k. Every node is interested in estimating the state $\tilde{x}$ of the network securely. We assume that network connectivity is fixed with time and the measurements trace follows a predefined sequence. We consider a discrete-time, linear system model with pairwise measurements taken per time step i.

$$\begin{aligned} \tilde{x}_{i+1}^k &= \tilde{F}_i \tilde{x}_i^k + \tilde{n}_i^k \\ y_i^{k,j} &= \tilde{H}_i^{k,j} \tilde{x}_i^k + \tilde{v}_i^k + a_i^{k,j}, \end{aligned} \tag{4}$$

where $\tilde{x}_i^k \in \mathbb{R}^{n_k}$ is the state of node k at time $i \in \mathbb{N}$ and $y_i^{k,j} \in \mathbb{R}^{m_k}$ is the measurement sent to node k from the neighboring node $j \in \mathscr{N}_k$. The process and measurement noises are denoted by $\tilde{n}_i^k$ and $\tilde{v}_i^k$, respectively. All vectors and matrices are real-valued and have proper dimensions. The attack vector $a_i^{k,j}$ is a vector which models how an attacker corrupts the sensor measurements between node k and node j at time i. A non-zero element in the vector a corresponds to the attacked values on the corresponding sensor, otherwise the measurement is not attacked. Thereby, the additive attack vector $a_i^{k,j}$ can account for both a malicious node k and a corrupted link (k, j). Moreover, the attack values can be constant or time-varying. The modeling noise $\tilde{n}_i$ and measurement noise $\tilde{v}_i$ are assumed to be unknown but bounded by zonotopes: $\tilde{n}_i \in I_{Q_i} = \langle 0, Q_i \rangle$ and $\tilde{v}_i \in I_{R_i} = \langle 0, R_i \rangle$.

D. Proposed Solution

Our proposed solution is based mainly on a diffusion Kalman filter [21] integrated within a secure state estimation concept [22] and combined with reachability analysis [20]. The original non-secure diffusion Kalman filter consists of three main steps, namely, measurement update, time update, and diffusion update. By ensuring the security of all the steps of the diffusion Kalman filter, we can obtain a secure distributed state estimator. Thus, our solution consists of:

1) Secure measurement update: Every node shares its measurements with its neighbors and does some internal processing. Protecting the measurement update is achieved by extending secure state estimation in [22] for the distributed case.
2) Secure diffusion: Every node shares its network state estimate with its neighbors and combines the estimates in a convex way. We protect the diffusion step by accepting the shared estimate if and only if it is within the accepted region of the reachability analysis.
3) Time update: Every node updates its state, which is trivially protected as no data is exchanged.

We will describe the protection of the measurement update and secure diffusion in more detail in the following sections.

III. Secure Measurement Update

Sensor nodes have one radio for one type of measurement. Thus, each node has one active link at each time step for performing measurements (like calculating the pairwise distances between nodes). For instance, node_1 performs measurement with node_3 at t_1 in Figure 1.b. Then, it will perform measurement with node_4 at time t_2 in Figure 1.c. At the same time, node_2 performs measurements with node_4 then with node_3 as shown in Figures 1.b and 1.c. We have

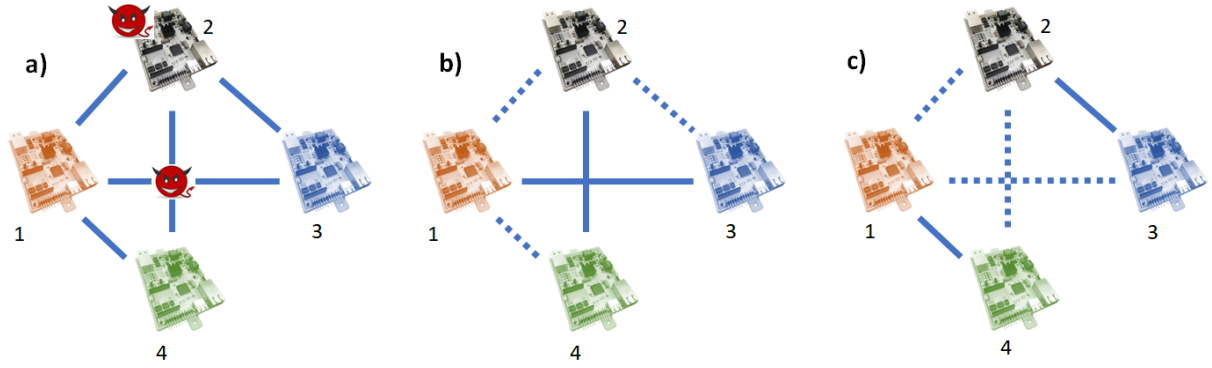


Fig. 1: Attacks are on links and sensors as shown in sub-figure a. Links are divided into passive (dashed) and active (bold) links at each time step. Active links carry a measurement between two nodes. The two sub-figures b and c show the active and passive links at two-time steps.

five links with five attacks to be mitigated in Figure 1 according to (4). Related work typically considers collecting the measurements from all the links and performs the estimation at once. This results in a very complex problem with the number of unknown attacks to be equal to the number of links.

In contrast, we propose solving the problem from another perspective: We define the links between node_2 and node_4 and between node_1 and node_3 in Figure 1.b as "active" links (bold) which carry measurements. The other links (dashed) in Figure 1.b are called "passive" links. The measurements' time horizon is divided into time steps where each node performs a measurement at each time step. We do not estimate the attack on passive links at the corresponding time steps. So, for example at time t_1 in Figure 1.b, why do we trouble ourselves to estimate $a_1^{2,3}$? At each time instant i, each node performs measurement with one neighbor. Note that due to this principle, it is possible to denote the attack $a_i^{k,j}$ only by a_i^k, i.e., neighbor j is uniquely defined by time step i and node k. With this idea, the number of attacks at each time instant equals the number of nodes, instead of the number of links. Therefore, we drastically simplify the secure state estimation problem. All the links for each node are modeled using one variable at different time steps. This even works while attacking all the links with time-varying attacks as we will show in Sec. V. This concept is repeated at each node k as every node is interested in estimating the network state. In short, we consider the attack variables attached to nodes instead of links at different time steps.

We utilize our idea to change the general model in (4) by including the attack value in the state of the node initiating the measurement. Following this procedure, the state of node k is extended to $x_i^k = [\tilde{x}_i^{k\,\mathrm{T}}, a_i^{k,j\,\mathrm{T}}]^{\mathrm{T}}$ yielding a modified system model

$$\begin{aligned} x_{i+1}^k &= F_i x_i^k + n_i^k, \\ y_i^{k,j} &= H_i^{k,j} x_i^k + v_i^k. \end{aligned} \tag{5}$$

Proposition 1: System model (4) is equivalent to system model (5) under the assumption that the measurements processing is done for each node with one neighbor at each time step for a network with pairwise measurements.

Proof: We get rid of j in $a_i^{k,j}$ in (4) by choosing $x_i^k = [\tilde{x}_i^{k\,\mathrm{T}}, a_i^{k,j\,\mathrm{T}}]^{\mathrm{T}}$ where the variations in j would be presented by changing the time step i, i.e., the couple i and k uniquely defines the neighbor j because node k can only communicate with one node at time i. The matrices $\tilde{F}_i$ and $\tilde{H}_i^{k,j}$ are changed accordingly to matrices F_i and $H_i^{k,j}$, respectively. We assume that the network connectivity is fixed with time and the sequence in measurements trace is predefined. □

We wish to highlight that our proposed solution still works if nodes perform multiple measurements at the same time step. This would be done by dividing the time horizon and processing one measurement for each node at one time step, as shown in Figures 1.b and 1.c.

A valid question would be how to model the time-evolution for time-varying attacks a_i^k? We choose to set $a_{i+1}^k = a_i^k + n_{a_i}^k$, and the variance of the modeling noise of the attack entries $n_{a_i}^k$ accounts for the time-varying aspect of the attack and for changing the links between time steps. This lets us move from specifying specific dynamics for the attacks. Another question would be how to obtain bounds on $n_{a_i}^k$. Our proposed solution is to use reachability analysis [20] and only accept the measurements that let the state stay inside the expected reachable set. Also, it should be noted that, since high attack values can be easily detected by threshold methods (e.g., if the reported distance is far beyond the range of the area where the network is employed), critical attacks can occur within a limited interval, which can be represented by the modeling covariance. This concept - of modeling a time-varying signal using the modeling noise - has been applied before in localization [23] where the authors use a stationary model for process updates of a flying quadrotor. As long as the quadrotor moves in the range of the modeling noise, the work in [23] would be able to localize it correctly.

To move forward with utilizing reachability analysis in our solution, we need to define the following sets:

Definition 2: **(Predicted State Set)** Given system (5) with

initial state $x_0 \in \langle c_0, G_0 \rangle$, the reachable state set $\mathscr{Z}_i^k$ of node k is defined as the set of all possible solutions x_i which can be reached given x_{i-1}. I_{Q_i} is the zonotope which bounds modeling noise [24, p.4]

$$\mathscr{Z}_i^k = F_i \mathscr{Z}_{i-1}^k \oplus I_{Q_i}. \tag{6}$$

□

Definition 3: **(Measurement State Set)** Given system (5), the measurement state set $\mathscr{S}_i^{k,j}$ of node k is defined as the set of all possible solutions x_i which can be reached given y_i and v_i. This measurement set is a strip [24, p.4]:

$$\mathscr{S}_i^{k,j} = \left\{ x_i \middle| |H_i^{k,j} x_i - y_i^{k,j}| \leq R_i^j \right\}. \tag{7}$$

□

Definition 4: **(Corrected State Set)** Given system (5) with initial state $x_0 \in \langle c_0, G_0 \rangle$, the reachable corrected state set $\mathscr{Z}_{\psi_{i|i}}^k$ of node k is defined as the intersection between $\mathscr{Z}_i^k$ and $\mathscr{S}_i^{k,j}$ [24, p.4]:

$$\mathscr{Z}_{\psi_{i|i}}^k = \mathscr{Z}_i^k \cap \mathscr{S}_i^{k,j}. \tag{8}$$

□

We denote the predicted and the filtered estimates of x_i at time step i obtained by node k as $\hat{x}_{i+1|i}^k$ and $\hat{x}_{i|i}^k$, respectively. The main algorithm is summarized in Algorithm 1. We start with zonotope $\mathscr{Z}_0 = \langle c_0, G_0 \rangle$ where the center c_0 equals the expected initial estimates x_0.

Every measurement $y_i^{k,j}$ restricts the state to be in a strip $\mathscr{S}_i^{k,j}$ as shown in (7). Every node corrects the reachable set (zonotope $\mathscr{Z}_{i|i-1}^k$) by determining the set of consistent states with the model and the measurements received from each neighbor. Therefore, we need to find the intersection between the family of strips in (7) and the zonotope $\mathscr{Z}_{i|i-1}^k$. This results in calculating the corrected over-approximated zonotope $\mathscr{Z}_{\psi_{i|i}}^k$ for node k. We extend the work [25] in the following theorem to find the required intersection.

Theorem 1: The zonotope $\mathscr{Z} = \langle c_0, G_0 \rangle$, the family of m strips $\mathscr{S}_i^{k,j}$ (7), and the vectors $\lambda_i^{k,j} \in \mathbb{R}^n$ are given. The intersection between the zonotope and the strips can be over-approximated by a zonotope $\mathscr{Z}_{\psi_{i|i}}^k = \langle c(\lambda), G(\lambda) \rangle$, where

$$c(\lambda) = c_0 + \sum_{j \in \mathscr{N}_k} \lambda_i^{k,j} (y_j - H_i^{k,j} c_0) \tag{9}$$

$$G(\lambda) = \left[(I - \sum_{j \in \mathscr{N}_k} \lambda_i^{k,j} H_i^{k,j}) G_0, \ \lambda_i^{k,1} R_i^1, ..., \lambda_i^{k,m_k} R_i^{m_k} \right]. \tag{10}$$

Proof: Let $x \in (\mathscr{Z} \cap \mathscr{S}_1^{k,j} \cap ... \cap \mathscr{S}_m^{k,j})$, then there is a z, where

$$x = c_0 + G_0 z, \tag{11}$$

where G_0 has full rank. We would like to highlight that the over-approximation comes from choosing $x \in (\mathscr{Z} \cap \mathscr{S}_1^{k,j} \cap ... \cap \mathscr{S}_m^{k,j})$. Then, adding and subtracting $\sum_{j \in \mathscr{N}_k} \lambda_i^{k,j} H_i^{k,j} G_0 z$ results in

$$x = c_0 + \sum_{j \in \mathscr{N}_k} \lambda_i^{k,j} H_i^{k,j} G_0 z + (I - \sum_{j \in \mathscr{N}_k} \lambda_i^{k,j} H_i^{k,j}) G_0 z. \tag{12}$$

Given that x is inside the intersection of the zonotope $\mathscr{Z}$ and the family of strips, then $x \in \mathscr{S}_i^{k,j}, \forall j \in \mathscr{N}_k$, i.e., there exists a $b_j \in [-1, 1]$ in (7) for the j^{th} strip so that:

$$H_i^{k,j} x - y_j = R_i^j b_j. \tag{13}$$

Inserting (11) in (13) results in

$$H_i^{k,j} G_0 z = y_j - H_i^{k,j} c_0 + R_i^j b_j. \tag{14}$$

Inserting (14) in (12) results in

$$\begin{aligned}
x = \ & c_0 + \sum_{j \in \mathscr{N}_k} \lambda_i^{k,j} (y_j - H_i^{k,j} c_0 + R_i^j b_j) \\
+ \ & (I - \sum_{j \in \mathscr{N}_k} \lambda_i^{k,j} H_i^{k,j}) G_0 z \\
= \ & c_0 + \sum_{j \in \mathscr{N}_k} \lambda_i^{k,j} (y_j - H_i^{k,j} c_0) \\
+ \ & (I - \sum_{j \in \mathscr{N}_k} \lambda_i^{k,j} H_i^{k,j}) G_0 z + \sum_{j \in \mathscr{N}_k} \lambda_i^{k,j} R_i^j b_j \\
= \ & \underbrace{c_0 + \sum_{j \in \mathscr{N}_k} \lambda_i^{k,j} (y_j - H_i^{k,j} c_0)}_{c(\lambda)} \\
& + \underbrace{\left[(I - \sum_{j \in \mathscr{N}_k} \lambda_i^{k,j} H_i^{k,j}) G_0, \lambda_i^{k,1} R_i^1, ..., \lambda_i^{k,m_k} R_i^{m_k} \right]}_{G(\lambda)} \begin{bmatrix} z \\ b_1 \\ .. \\ b_{m_k} \end{bmatrix} \\
= \ & c(\lambda) + G(\lambda) \begin{bmatrix} z \\ b_1 \\ .. \\ b_{m_k} \end{bmatrix}
\end{aligned}$$

□

Choosing an appropriate λ: In order to find an appropriate over-approximation for the intersection of a zonotope with a family of strips, λ is typically chosen to minimize an approximation criterion. The authors in [25] proposed two approaches for intersecting a zonotope with a strip. The first approach is a segment-minimization approach which has a low computational complexity by minimizing the Frobenius norm of $G(\lambda)$. The second approach provides a better approximation; it is a volume-minimizing approach and requires solving a convex optimization problem.

However, if we take a careful look at (15), which is the measurement update equation that propagates the measurement effect into the estimates in the diffusion Kalman filter [21], we can find that the structure is very similar to our formula (9) which finds the new center of the intersection of a zonotope with a family of strips.

$$\psi_i^k = \hat{x}_{i|i-1}^k + P_{i|i}^k \sum_{j \in \mathscr{N}_k} H_i^{*^{k,j}} R_i^{j^{-1}} [y_i^{k,j} - H_i^{k,j} \hat{x}_{i|i-1}^k] \tag{15}$$

Thus, we choose to use the λ that is aligned with diffusion Kalman filter theory at each node k as shown in step 1 of Algorithm 1:

$$\lambda_i^{k,j} = P_{i|i}^k {H_i^*}^{k,j} {R_i^j}^{-1} \tag{16}$$

We choose the center of the reachable set on every node k as the estimate ψ_i^k. Also, as the size of the generators is increasing in each step by doing the previous measurement update, we reduce the order of the corrected zonotope $\mathscr{Z}_{\psi_{i|i}}^k = \langle \psi_i^k, G_{\psi_{i|i}}^k \rangle$ order by the method from [26, p.7].

IV. Secure Diffusion

In the diffusion step [21], every node shares its own local estimate ψ_i^j with its neighbors. Then every node averages the shared estimates ψ_i^j from the neighbors to achieve a better estimate of the system state. The averaging is based on some weights $w_i^{k,j}$ for each neighbor j [21]. However, these shares may be under attack. Thus, we make use of reachability analysis to protect against attacks during the diffusion step.

We propose to let every node compute the next corrected state-set $\mathscr{Z}_{\psi_{i|i}}^k$. Then, the combination in the diffusion step [21] is executed over shares ψ_i^j inside the corrected reachable set $\mathscr{Z}_{\psi_{i|i}}^k$ of the node. If the share ψ_i^j is outside its corrected set, it would be marked as "attacked share" and thus excluded. Thus, we can limit the effect of the attack on the diffusion shares ψ_i^j. More specifically, we assign the weight to zero if the share is outside the reachable set $\mathscr{Z}_{\psi_{i|i}}^k$. Shares inside $\mathscr{Z}_{\psi_{i|i}}^k$ take new weights $\hat{w}_i^{k,j}$ where $\sum_{j \in \mathscr{N}_k} \hat{w}_i^{k,j} = 1$.

$$w_i^{k,j} = \begin{cases} 0 & \text{if } \psi_i^j \notin \mathscr{Z}_{\psi_{i|i}}^k \\ \hat{w}_i^{k,j} & \text{else} \end{cases} \tag{17}$$

This illustrates Step 2 of Algorithm 1.

V. Evaluation

Our proposed algorithm is implemented in Matlab 2017 on a similar example to the one presented in [21], where a network of eight nodes attempts to track the position of a rotating object. All computations run on a single thread of an Intel(R) Core(TM) i7-8750 with 16 GB RAM. We made use of Cora [27]–[29] for zonotope operations. Our example is quite representative for secure state estimation, since it includes modeling noise and measurements noise. The state of each node consists of the unknown 2-dimensional position of the object combined with the attack on the measurements. The state matrix in (5) is

$$F = \begin{bmatrix} 0.992 & -0.1247 & 0 \\ 0.1247 & 0.992 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{12}$$

and the measurement matrix $H_i^{k,j}$ is [0 1 1] or [1 0 1] in the sequence of the taken measurements. This means that the nodes take measurements of the unknown position of the object either in the x or y direction and the measurements are under attack a_i^k. We generate the attacks a_i^k as following:

$$a_i^k = \begin{cases} 2\ \texttt{rand} + 4 & \text{if } t < 1/3T_{\text{sim}}, \\ \texttt{randp}(3,2) + 8 & \text{if } 1/3T_{\text{sim}} < t < 2/3T_{\text{sim}}, \\ 2\ \texttt{randn} + 50 & \text{if } t > 2/3T_{\text{sim}}, \end{cases} \tag{13}$$

where rand and randn return pseudo-random values drawn from the standard uniform distribution on the open interval (0,1) and the standard normal distribution, respectively. On the other hand, $\texttt{randp}(3,2)$ generates values form the Pareto distribution, where the shape equals 3, and the scale equals 2. T_{sim} is the total simulation time.

The implemented random attacks have time-varying statistics, as shown in (13). The simulations without protection and with protection are shown in Figures 2a and 2b, respectively, by attacking the measurement step only. Then, we attack the diffusion shares ψ_i^k using the three presented random generators. The insecure and secure versions run on the same values of random numbers for a fair comparison and the

Algorithm 1 Secure Diffusion Kalman Filter

Start with $\hat{x}_{0|-1}^k = x_0$, $P_{0|-1}^k = \Pi_0$ and zonotope $\mathscr{Z}_{0|-1}^k = \langle \hat{x}_{0|-1}^k, G_{0|-1}^k \rangle$. For all k, and at every time instant i, compute at every node k:

Step 1: Measurement update (Sec. III):

$$\begin{aligned} {P_{i|i}^{-1}}^k &= {P_{i|i-1}^{-1}}^k + \sum_{j \in \mathscr{N}_k} {H_i^*}^{k,j} {R_i^j}^{-1} H_i^{k,j} \\ \lambda_i^{k,j} &= P_{i|i}^k {H_i^*}^{k,j} {R_i^j}^{-1} \\ \psi_i^k &= \hat{x}_{i|i-1}^k + \sum_{j \in \mathscr{N}_k} \lambda_i^{k,j} (y_i^{k,j} - H_i^{k,j} \hat{x}_{i|i-1}^k) \\ G_{\psi_{i|i}}^k &= [(I - \sum_{j \in \mathscr{N}_k} \lambda_i^{k,j} H_i^{k,j}) G_{i|i-1}^k, \quad \lambda_i^{k,1} R_i^1, \\ &\quad ..., \lambda_i^{k,m_k} R_i^{m_k}] \end{aligned}$$

Reduce the order of the corrected zonotope $\mathscr{Z}_{\psi_{i|i}}^k = \langle \psi_i^k, G_{\psi_{i|i}}^k \rangle$ order by Girard method [26, p.7].

Step 2: Diffusion update (Sec. IV):
Filter ψ_i^j based on the reachability analysis, i.e. average ψ_i^j from neighbors if they are within the expected reachable set and assign the weights $w_i^{k,j}$ accordingly. ($w_i^{k,j} = 0$ if $\psi_i^j \notin \mathscr{Z}_{\psi_{i|i}}^k$).

$$\begin{aligned} \hat{x}_{i|i}^k &= \sum_{j \in \mathscr{N}_k} w_i^{k,j} \psi_i^j \\ G_{i|i}^k &= G_{\psi_{i|i}}^k \end{aligned}$$

Step 3: Time update:

$$\begin{aligned} \hat{x}_{i+1|i}^k &= F_i \hat{x}_{i|i}^k \\ P_{i+1|i}^k &= F_i P_{i|i}^k F_i^* + Q_i \\ G_{i+1|i}^k &= [F_i G_{i|i}^k, Q_i] \end{aligned}$$

diffusion step only is under attack. The outputs are shown in Figures 3a and 3b. Finally, we attack both the measurement and diffusion steps and report the output in Figures 4a and 4b. We obtain the reported means and standard deviations in Table I with and without the proposed protection on the same set of attacks. The mean is around 3m for the secure version with a standard deviation around 1.5 regardless of the attack type.

VI. Conclusion

We combine reachability analysis with secure state estimation to obtain a secure and fully distributed estimator. Our approach works for discrete-time, linear systems affected by disturbances, and measurement noises. Our proposed solution is fully distributed and does not require a fusion center. We consider attacks on the sensor levels as well as the communication links. At each time step, estimation output is supervised by reachability analysis to provide secure estimation shares. Reachability analysis allows us to have secure diffusion in distributed secure state estimation. We demonstrate the applicability of our approach with a simulation of a rotating target where the measurements and the diffusion shares are under attack.

Acknowledgements

We gratefully acknowledge partial financial support by the project justITSELF funded by the European Research Council (ERC) under grant agreement No 817629 and the project interACT under grant agreement No 723395; both projects are funded within the EU Horizon 2020 program.

References

[1] M. Uma and G. Padmavathi, "A survey on various cyber attacks and their classification." *I. J. Network Security*, vol. 15, no. 5, pp. 390–396, 2013.

[2] J. Slay and M. Miller, "Lessons learned from the Maroochy water breach," in *International Conference on Critical Infrastructure Protection*, 2007, pp. 73–82.

[3] D. Kushner, "The real story of Stuxnet," *IEEE Spectrum*, vol. 50, no. 3, pp. 48–53, 2013.

[4] R. Langner, "Stuxnet: Dissecting a cyberwarfare weapon," *IEEE Security & Privacy*, vol. 9, no. 3, pp. 49–51, 2011.

[5] V. M. Igure, S. A. Laughter, and R. D. Williams, "Security issues in SCADA networks," *Computers & Security*, vol. 25, no. 7, pp. 498–506, 2006.

[6] D. F. Kune, J. Backes, S. S. Clark, D. Kramer, M. Reynolds, K. Fu, Y. Kim, and W. Xu, "Ghost talk: Mitigating EMI signal injection attacks against analog sensors," in *Symposium on Security and Privacy*. IEEE, 2013, pp. 145–159.

[7] H. Fawzi, P. Tabuada, and S. Diggavi, "Secure estimation and control for cyber-physical systems under adversarial attacks," *IEEE Transactions on Automatic Control*, vol. 59, no. 6, pp. 1454–1467, 2014.

[8] M. S. Chong, M. Wakaiki, and J. P. Hespanha, "Observability of linear systems under adversarial attacks," in *American Control Conference*. IEEE, 2015, pp. 2439–2444.

[9] M. Pajic, J. Weimer, N. Bezzo, P. Tabuada, O. Sokolsky, I. Lee, and G. J. Pappas, "Robustness of attack-resilient state estimators," in *ACM/IEEE 5th International Conference on Cyber-Physical Systems*, 2014, pp. 163–174.

[10] F. Pasqualetti, F. Dorfler, and F. Bullo, "Control-theoretic methods for cyberphysical security: Geometric principles for optimal cross-layer resilient control systems," *IEEE Control Systems*, vol. 35, no. 1, pp. 110–127, 2015.

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

[12] Y. Mo and B. Sinopoli, "Secure control against replay attacks," in *47th annual Allerton conference on communication, control, and computing*. IEEE, 2009, pp. 911–918.

[13] F. Miao, M. Pajic, and G. J. Pappas, "Stochastic game approach for replay attack detection," in *52nd Conference on Decision and Control*. IEEE, 2013, pp. 1854–1859.

[14] S. Amin, A. A. Cárdenas, and S. Sastry, "Safe and secure networked control systems under denial-of-service attacks." in *ACM International Conference on Hybrid Systems: Computation and Control*, 2009, pp. 31–45.

[15] Y. Mo, E. Garone, A. Casavola, and B. Sinopoli, "False data injection attacks against state estimation in wireless sensor networks," in *49th Conference on Decision and Control*. IEEE, 2010, pp. 5967–5972.

[16] F. Pasqualetti, F. Dörfler, and F. Bullo, "A divide-and-conquer approach to distributed attack identification," in *54th Conference on Decision and Control*. IEEE, 2015, pp. 5801–5807.

[17] W. Ao, Y. Song, and C. Wen, "Distributed secure state estimation and control for CPSs under sensor attacks," *IEEE Transactions on Cybernetics*, pp. 1–11, 2018, [available online].

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

[19] F. S. Cattivelli and A. H. Sayed, "Distributed nonlinear Kalman filtering with applications to wireless localization," in *International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2010, pp. 3522–3525.

[20] M. Althoff, "Reachability analysis and its application to the safety assessment of autonomous cars," Ph.D. dissertation, Technische Universität München, 2010.

[21] F. S. Cattivelli, C. G. Lopes, and A. H. Sayed, "Diffusion strategies for distributed Kalman filtering: formulation and performance analysis," *Proceedings Cognitive Information Processing*, pp. 36–41, 2008.

[22] A. Alanwar, B. Etzlinger, H. Ferraz, J. Hespanha, and M. Srivastava, "SecSens: Secure state estimation with application to localization and time synchronization," *arXiv preprint arXiv:1801.07132*, 2018.

[23] A. Alanwar, H. Ferraz, K. Hsieh, R. Thazhath, P. Martin, J. Hespanha, and M. Srivastava, "D-slats: Distributed simultaneous localization and time synchronization," in *Proceedings of the 18th ACM International Symposium on Mobile Ad Hoc Networking and Computing*. ACM, 2017, p. 14.

[24] Y. Wang, V. Puig, and G. Cembrano, "Set-membership approach and Kalman observer based on zonotopes for discrete-time descriptor systems," *Automatica*, vol. 93, pp. 435–443, 2018.

[25] T. Alamo, J. M. Bravo, and E. F. Camacho, "Guaranteed state estimation by zonotopes," *Automatica*, vol. 41, no. 6, pp. 1035–1043, 2005.

[26] A. Girard, "Reachability of uncertain linear systems using zonotopes," in *International Workshop on Hybrid Systems: Computation and Control*, 2005, pp. 291–305.

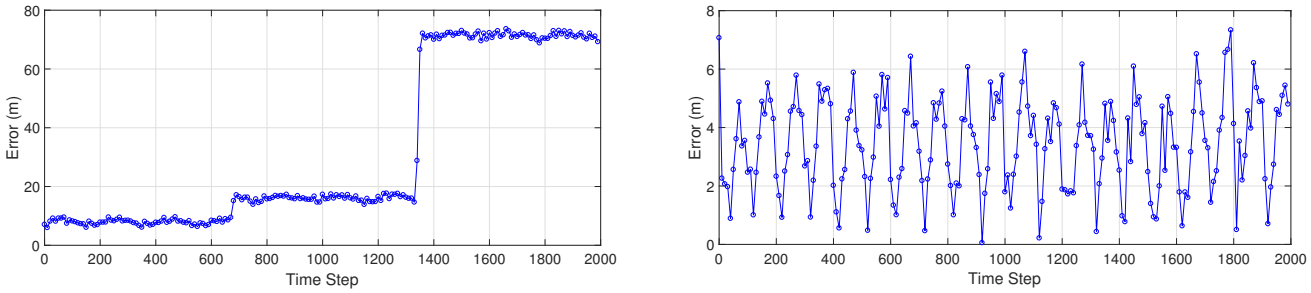
[27] M. Althoff, "An introduction to CORA 2015," in *Proc. of the Workshop on Applied Verification for Continuous and Hybrid Systems*, 2015, pp. 120–151.

[28] M. Althoff and D. Grebenyuk, "Implementation of interval arithmetic in CORA 2016," in *Proc. of the 3rd International Workshop on Applied Verification for Continuous and Hybrid Systems*, 2016, pp. 145–173.

[29] M. Althoff, D. Grebenyuk, and N. Kochdumper, "Implementation of taylor models in CORA 2018," in *Proc. of the 5th International Workshop on Applied Verification for Continuous and Hybrid Systems*, 2018, pp. 91–105.

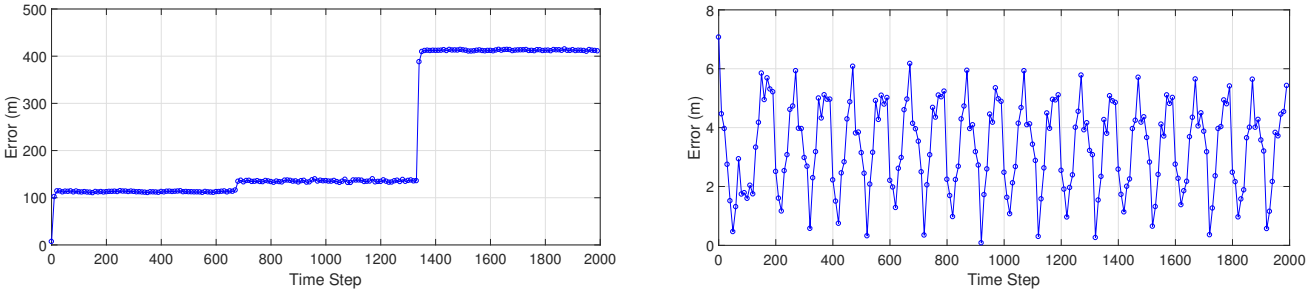
Steps under attack	Insecure		Secure	
	mean	std	mean	std
Measurement step only	31.558	28.119	3.306	1.523
Diffusion step only	219.325	136.567	3.194	1.474
Measurement and diffusion steps	250.161	164.489	3.220	1.518

TABLE I: The mean and standard deviation of the localization error (m) of the rotating target at one node with and without the proposed protection algorithm.



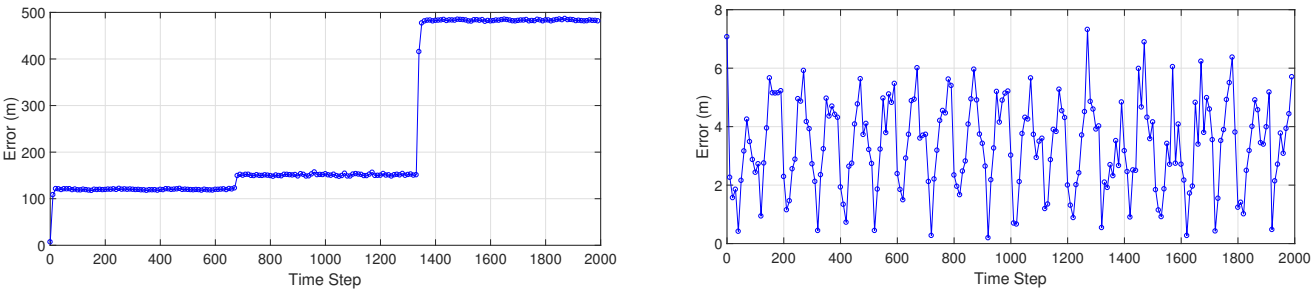
(a) Insecure (b) Secure

Fig. 2: Localization error at one node of the rotating target where all the measurements are under attack. The measurements only are under attack while the diffusion step is not under attack. Attacks are generated from uniform, normal and Pareto pseudo-random distributions, as shown in (13). Y-scales are different in Figures (a) and (b).



(a) Insecure (b) Secure

Fig. 3: Localization error at one node of the rotating target where all the diffusion shares are only under attack. Attacks are generated from uniform, normal and Pareto pseudo-random distributions. Y-scales are different in Figures (a) and (b).



(a) Insecure (b) Secure

Fig. 4: Localization error at one node of the rotating target where all the measurements and the diffusion shares are under attack with time varying values. Attacks are generated from uniform, normal and Pareto pseudo-random distributions as shown in (13). Y-scales are different in Figures (a) and (b).