Secure Set-Based State Estimation for Linear Systems under Adversarial Attacks on Sensors

Secure Set-Based State Estimation for Linear Systems under Adversarial Attacks on Sensors

M. Umar B. Niazi [a,b], Michelle S. Chong [c], Amr Alanwar [d,e], Karl H. Johansson [b]

[a] *Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge MA 02139, USA*

[b] *Division of Decision and Control Systems, Digital Futures, KTH Royal Institute of Technology, SE-100 44 Stockholm, Sweden*

[c] *Department of Mechanical Engineering, Eindhoven University of Technology, the Netherlands*

[d] *School of Computation, Information and Technology, Technical University of Munich, Germany*

[e] *School of Computer Science and Engineering, Constructor University, Germany*

Abstract

When a strategic adversary can attack multiple sensors of a system and freely choose a different set of sensors at different times, how can we ensure that the state estimate remains uncorrupted by the attacker? The existing literature addressing this problem mandates that the adversary can only corrupt less than half of the total number of sensors. This limitation is fundamental to all point-based secure state estimators because of their dependence on algorithms that rely on majority voting among sensors. However, in reality, an adversary with ample resources may not be limited to attacking less than half of the total number of sensors. This paper avoids the above-mentioned fundamental limitation by proposing a set-based approach that allows attacks on all but one sensor at any given time. We guarantee that the true state is always contained in the estimated set, which is represented by a collection of constrained zonotopes, provided that the system is bounded-input-bounded-state stable and redundantly observable via every combination of sensor subsets with size equal to the number of uncompromised sensors. Additionally, we show that the estimated set is secure and stable irrespective of the attack signals if the process and measurement noises are bounded. To detect the set of attacked sensors at each time, we propose a simple attack detection technique. However, we acknowledge that intelligently designed stealthy attacks may not be detected and, in the worst-case scenario, could even result in exponential growth in the algorithm's complexity. We alleviate this shortcoming by presenting a range of strategies that offer different levels of trade-offs between estimation performance and complexity. To illustrate the efficacy of our approach, we apply it to a vertically interconnected mechanical system that models a three-story building structure. Our results demonstrate that the proposed set-based method provides a robust and secure state estimation method that can handle a greater number of attacked sensors than existing point-based estimators.

Key words: Secure state estimation, set-based methods, zonotopic filtering, attack detection.

1 Introduction

The interconnectedness of control systems makes them vulnerable to malicious attacks. Sensors are particularly susceptible to attacks compared to other control system components because they can be tampered with to provide false measurements. These measurements, when employed for monitoring and control, can deceive the system operator and cause undesirable disruptions.

To mitigate this risk and ensure accurate state estimation, several methods have been proposed recently. Known as *resilient* or *secure* state estimation, these methods leverage the redundancy of sensors to obtain estimates that converge to a neighborhood of the true state, even in the presence of additive sensor attacks. These methods (see, e.g., [5, 12, 15, 16, 18, 28]) are designed to prevent corrupted sensor data from degrading estimation accuracy and maintain reasonable control

Email addresses: niazi@mit.edu (M. Umar B. Niazi), m.s.t.chong@tue.nl (Michelle S. Chong), alanwar@tum.de (Amr Alanwar), kallej@kth.se (Karl H. Johansson).

[1] This work is supported by the Swedish Research Council and the Knut and Alice Wallenberg Foundation, Sweden. It has also received funding from the European Union's Horizon Research and Innovation Programme under grant agreement No. 830927 and Marie Skłodowska-Curie grant agreement No. 101062523.

performance. However, most methods proposed in the literature are point-based, i.e., the estimated state at any given time is a point in the state space. The estimation error bounds of point-based methods are not precise because of their dependence on comparison functions, which are known to be conservative. Moreover, the reliance of point-based methods on majority voting-type algorithms presents a fundamental limitation that only less than half the total number of sensors are allowed to be attacked at any given time. However, an adversary with ample resources would be capable of attacking more than half the number of sensors.

1.1 Related literature

To overcome the limitations of point-based methods and obtain tighter robust guarantees, the set-based zonotopic filtering paradigm [1, 4, 7, 8, 14, 19, 31] is promising. It has found many real-world applications, including fault diagnosis in industrial systems [9], underwater robotics [17], vehicle localization [10], and leakage detection in water distribution networks [30]. Moreover, in safety-critical applications, guaranteed state inclusion in a bounded set is crucial to provably avoid unsafe regions in the state space. All these issues further motivate the need for set-based state estimation techniques that provide a set of all possible states under unknown disturbances and measurement errors belonging to known bounded sets.

Different from zonotopic estimators, other set-based estimators include interval observers [29, 37], which estimate a box at every time instant, guaranteeing the true state's inclusion. However, a box fails to capture the inter-dependencies between the state variables and is less accurate. In contrast, zonotopic filters are shown to not only provide an accurate set-based estimate but are also computationally efficient.

Despite the growing importance of secure estimation in the presence of adversarial attacks, the existing set-based methods have several shortcomings, including restrictive assumptions, limited robustness to stealthy attacks, and reliance on specific attack strategies. The reachability-based approach of [33] requires the full state vector to be measured by any subset of sensors with cardinality equal to the number of safe sensors. This is a highly restrictive assumption, and without it, the guarantees of estimation accuracy provided in [33] become excessively conservative. Recent works [11, 21, 25] employ standard interval-based or zonotopic filters by identifying and discarding the sensors that have been corrupted. However, this approach fails to exclude stealthy attacks on the sensors, which can easily evade the proposed attack detection procedure. As a result, stealthy attacks can significantly affect estimation accuracy by constantly injecting small signals in the sensor measurements, which can accumulate over time without being detected. Although this issue is addressed by [22, 23] and [24, 35, 39, 40], the proposed methods are either restricted to a particular attack strategy or rely on the assumption that the attack signal is bounded.

1.2 Our Contribution

In this paper, we propose a new set-based state estimation method that neither requires a full state vector to be measured by any subset of sensors nor allows stealthy attacks to corrupt the estimation performance significantly. We also do not assume the boundedness of the attack signals or that the attacker resorts to a particular attack strategy. Moreover, to address the fundamental limitation of point-based secure estimators that strictly less than half the number of sensors can be attacked at any given time, we allow the attacker to corrupt all but one sensor under the assumption that the system remains observable from the remaining attack-free sensors. Subject to these assumptions, we present a zonotope-based state estimation algorithm for linear systems under time-varying sensor attacks and show that the estimated set is guaranteed to contain the true state.

We would like to point out that although the algorithm presented in this paper can also handle sensor faults, it was developed with *adversarial attacks* in mind. Attacks differ from faults in that an attack is an intelligently designed strategy to inflict a maximal negative impact on the system and/or evade detection, or in other words, is stealthy. For example, injecting small signals to corrupt the sensor measurements, which cannot be detected at any instant in time, but its effect can accumulate over a long time horizon. Another example is the so-called zero dynamics attack [32], where a specially designed sensor attack signal based on the unstable zero dynamics of the plant can grow unbounded while the plant's state is driven away from the attack-free trajectory. Such an attack is directed close to the output null space such that the sensor measurement is close to zero (modulo noise). These intelligently designed attacks can cause conventional state estimation algorithms to provide inaccurate state estimates, where the estimation accuracy is dependent on the attack signals. We eliminate the undesirable effect of sensor attacks with our proposed secure set-based estimation algorithm.

Our algorithm operates in a series of steps at every time instant. Firstly, it calculates the time update by utilizing the model and the bounds on the process noise. Subsequently, subsets of state-space consistent with the sensor measurements are computed. These subsets, known as *measurement consistent sets*, are determined based on the bounds of the measurement noise. Essentially, the measurement consistent sets correspond to the possible states in the state space that could have generated the obtained sensor measurements, given that the exact realization of measurement noise can be anywhere within

the specified bounds described by a zonotope. In practical scenarios, the system might not be observable from every sensor, and an attacker could compromise a subset of sensors. To account for this, we create multiple agreement sets by intersecting various combinations of measurement consistent sets. By doing so, we can eliminate measurement consistent sets that correspond to attacked sensors. We establish that at least one agreement set contains the true system state. Lastly, the measurement update is computed by intersecting the time update with the agreement set, which further eliminates the agreement sets affected by non-stealthy attack signals. We also propose a simple procedure to identify the set of compromised sensors, or its subset if the attacks are stealthy, at any given time.

The main strength of our algorithm is that it can handle attacks on different subsets of sensors at any given time. In addition, our approach can accommodate any cardinality of attacked sensors as long as they are less than the total number of sensors. The reason for this is that we do not rely on a simple majority vote among the sensors. Instead, we use a combination of intersections between their measurement consistent sets, which lead to the agreement sets. These combinations are compared with the model-based time update set in such a way that we can verify the validity of multiple agreement sets if only one sensor is uncompromised.

A shortcoming of our algorithm is that its complexity may exponentially increase in the worst-case scenario when intelligent and stealthy attacks are executed. Despite this, we argue that it is challenging for attackers to achieve this worst-case scenario because it requires a complete understanding of both the system and the algorithm, as well as ample computational resources to calculate the optimal attack within a single time sample. Nonetheless, to tackle this complexity issue, we suggest various strategies to reduce complexity and facilitate the implementation of our zonotope-based secure state estimation algorithm.

Furthermore, when less than half of the sensors have been attacked, we integrate a point-based resilient state observer into our algorithm to prune the candidate sets. However, this strategy's practicality depends on an accurate approximation of the guaranteed estimation error provided by the resilient point-based state observer. The modified algorithm provides asymptotic convergence guarantees with an explicit bound that depends on known process and measurement noise, independent of the attack signals.

In summary, our contributions in this paper include a novel secure set-based state estimation algorithm [2] that guarantees the following:

(1) state inclusion, i.e., the true state is always guaranteed to be inside the estimated set;
(2) secure estimation, i.e., large attack signals are automatically discarded, while the impact of stealthy attacks on the estimation accuracy is negligible;
(3) stability guarantees, i.e., the estimated set remains bounded irrespective of the unbounded attack if the process and measurement noise signals are bounded;
(4) attack detection scheme that identifies the set of attacked sensors; and
(5) methods to address the algorithm's complexity with minimal compromise on the estimation accuracy.

To this end, we assume that the system is redundantly observable, i.e., observable from every possible combination of sensors with cardinality less than or equal to the number of safe sensors. This assumption is more realistic than the one in our previous work [27], which required observability from every sensor. Also, the analysis under this assumption turns out to be significantly different and non-trivial compared to [27]. Furthermore, we provide stability guarantees and an attack detection algorithm in this paper, which is missing in [27]. Finally, we demonstrate the effectiveness of our proposed method through an illustrative example and a more practical example of set-based state estimation of a three-story building structure [36] during an earthquake and when an adversary has compromised some of the sensors.

1.3 Outline

The rest of the paper is organized as follows. Section 2 defines the notations and summarizes the required preliminaries on set representations. The main assumptions and the problem are stated in Section 3. Section 4 presents the secure zonotopic state estimation algorithm, provides the inclusion guarantees, and discusses methods to reduce the algorithm's complexity. Section 5 provides the stability analysis of the estimation algorithm. Section 6 presents the attack detection scheme and discusses the cases where it works. We also discuss the worst-case complexity of the proposed algorithm under stealthy attacks and proposes methods to handle it effectively. Finally, Section 7 demonstrates the effectiveness of the proposed algorithm through simulation examples, and Section 8 concludes the paper.

2 Notations and Preliminaries

2.1 Notations

The set of real numbers and integers are denoted by $\mathbb{R}$ and $\mathbb{Z}$, respectively. We let $\mathbb{Z}_{\geqslant i} \doteq \{i, i+1, i+2, \dots\}$ and $\mathbb{Z}_{[i,k]} \doteq \{i, i+1, i+2, \dots, k\}$ for $k \geqslant i$. The maximum norm of a vector $x \in \mathbb{R}^n$ is denoted as $\|x\| \doteq$

[2] The code is available online in our GitHub repository https://github.com/aalanwar/Secure-Set-Based-Estimation.

$\max_{i\in\{1,\dots,n\}} |x_i|$. Given a signal $v : \mathbb{Z}_{\geqslant 0} \to \mathbb{R}^n$, we denote its restriction to the interval $[0, k]$ by $v_{[0,k]}$, for some $k \in \mathbb{Z}_{\geqslant 0}$. For a set $\mathcal{S}$, $|\mathcal{S}|$ denotes its cardinality. Given multiple sets $\mathcal{S}_1, \dots, \mathcal{S}_n$, we denote their collection as $\mathcal{S} = \{\mathcal{S}_i\}_{i\in\mathbb{Z}_{[1,n]}}$. The notation $\binom{n}{k} = \frac{n!}{n!(n-k)!}$ denotes the binomial coefficient or the number of possible combinations when k out of n elements are chosen. The Cartesian product is denoted by $\times$.

2.2 Set Representations

Given a center $c_z \in \mathbb{R}^n$ and a generator matrix $G_z \in \mathbb{R}^{n\times\xi_z}$, a *zonotope* $\mathcal{Z} \subset \mathbb{R}^n$ is a set

$$\mathcal{Z} \doteq \{c_z + G_z\beta_z : \beta_z \in [-1, 1]^{\xi_z}\}$$

where ξ_z is the number of generators of $\mathcal{Z}$. Since a zonotope can be completely characterized by its center and generator matrix, the notation $\mathcal{Z} = \langle c_z, G_z\rangle$ is used throughout the paper for brevity.

A matrix $L \in \mathbb{R}^{n'\times n}$ multiplied with a zonotope $\mathcal{Z}$ yields a linearly transformed zonotope $L\mathcal{Z} = \langle Lc_z, LG_z\rangle$. Given two zonotopes $\mathcal{Z}_1 = \langle c_{z_1}, G_{z_1}\rangle$ and $\mathcal{Z}_2 = \langle c_{z_2}, G_{z_2}\rangle$, each being a subset of $\mathbb{R}^n$, their Minkowski sum is given by

$$\mathcal{Z}_1 \oplus \mathcal{Z}_2 = \langle c_{z_1} + c_{z_2}, [\, G_{z_1} \;\; G_{z_2} \,]\rangle.$$

Similarly, the Cartesian product of two zonotopes is defined and computed as

$$\begin{aligned}\mathcal{Z}_1 \times \mathcal{Z}_2 &\doteq \left\{ \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} : z_1 \in \mathcal{Z}_1, z_2 \in \mathcal{Z}_2 \right\} \\ &= \left\langle \begin{bmatrix} c_{z_1} \\ c_{z_2} \end{bmatrix}, \begin{bmatrix} G_{z_1} & 0 \\ 0 & G_{z_2} \end{bmatrix} \right\rangle.\end{aligned}$$

A *constrained zonotope* is a set

$$\mathcal{Z} \doteq \{c_z + G_z\beta_z : \beta_z \in [-1, 1]^{\xi_z}, A\beta_z = b\}$$

where $A \in \mathbb{R}^{n\times\xi_z}$ and $b \in \mathbb{R}^n$ with $n \in \mathbb{Z}_{>0}$. In other words, a zonotope is an affine transformation of the hypercube and a constrained zonotope is an affine transformation of the corresponding linearly constrained hypercube.

The radius of a zonotope, or a constrained zonotope, is computed as follows

$$\mathrm{rad}(\mathcal{Z}) = \min \Delta \text{ subject to } \mathcal{Z} \subseteq \mathcal{H}(c_z, \Delta),$$

which is the radius Δ of a minimal hypercube $\mathcal{H}(c_z, \Delta)$ centered at c_z that inscribes $\mathcal{Z}$.

3 Problem Definition

Consider an LTI system with $p \in \mathbb{Z}_{>0}$ sensors, in discrete-time, for $k \in \mathbb{Z}_{\geqslant 0}$

$$x(k+1) = Ax(k) + Bu(k) + w(k) \tag{1a}$$
$$y_i(k) = C_i x(k) + v_i(k) + a_i(k), \quad i \in \mathbb{Z}_{[1,p]} \tag{1b}$$

where $x(k) \in \mathcal{X} \subset \mathbb{R}^{n_x}$ is the state, $u(k) \in \mathcal{U} \subset \mathbb{R}^{n_u}$ is a known bounded input, and $y_i(k) \in \mathbb{R}^{m_i}$ is the measured output of the i-th sensor with $i \in \mathbb{Z}_{[1,p]}$. The vector $w(k) \in \mathcal{W}$ represents the process noise, which is bounded and assumed to be contained in the zonotope $\mathcal{W} = \langle c_w, G_w\rangle$, and the vector $v_i(k) \in \mathcal{V}_i$ represents the measurement noise of the i-th sensor, which is also bounded and assumed to be contained in the zonotope $\mathcal{V}_i = \langle c_{v_i}, G_{v_i}\rangle$, for every $i \in \mathbb{Z}_{[1,p]}$. Finally, $a_i(k) \in \mathbb{R}^{m_i}$ represents the attack signal injected by the attacker to corrupt the measurement of the i-th sensor, and it can be arbitrary and unbounded.

Assumption 1 *We assume the following:*

(i) ***Upper bound on the number of attacked sensors****: The attacker can attack up to q number of sensors, where $q \leqslant p - 1$ is known a priori. However, the exact number and the particular set of attacked sensors need not be known.*

(ii) ***Redundant observability****: For every sensor subset $\mathsf{J} \subset \mathbb{Z}_{[1,p]}$ with cardinality $|\mathsf{J}| = c_\mathsf{J}$, for some $c_\mathsf{J} \leqslant p - q$, the pair (A, C_J) is observable, where C_J is obtained by stacking C_j, for all $j \in \mathsf{J}$, in row blocks.*

(iii) ***Bounded state space****: For every time instant $k \in \mathbb{Z}_{\geqslant 0}$ and input $u_{[0,k]} \in \mathcal{U}$, each element $i \in \mathbb{Z}_{[1,n_x]}$ of the state vector satisfies $|x_i(k)| \leqslant \chi_i$, where $\chi_i > 0$ is known. That is, the state always remains bounded inside an n_x-dimensional box $\mathcal{X} = \langle 0, D_x\rangle$, where $D_x = \mathrm{diag}(\chi_1, \dots, \chi_{n_x}) \in \mathbb{R}^{n_x\times n_x}$ describes the dimensions of the box in any orthant.*

Assumption 1(i) is fundamental in this paper because it ensures that, at every time $k \in \mathbb{Z}_{\geqslant 0}$, there exists a set of *uncompromised* (or *safe*) sensors

$$\mathsf{S}_k \subset \mathbb{Z}_{[1,p]} \text{ with } |\mathsf{S}_k| \geqslant p - q$$

such that $a_i(k) = 0_{m_i}$ for every $i \in \mathsf{S}_k$. This, along with the redundant observability (Assumption 1(ii)), allows us to ensure that the true state can be theoretically reconstructed from the set of uncompromised sensors under the absence of noise. In addition, the assumption entails that the attacker, even though omniscient about the system dynamics and noise bounds, has limited resources at hand and cannot attack all the sensors. We remark that this assumption is not restrictive because it neither restricts the set of attacked sensors to be static

with respect to time nor requires that q is less than half the number of sensors p — an assumption that is fundamental in the secure state estimation literature. On the contrary, at any time instant, our problem setup allows the attacker to inject arbitrary signals to any subset of sensors with cardinality less than or equal to q, where q is only required to be strictly less than p — meaning that at least one sensor needs to be safe at any time instant.

Assumption 1(ii) is required to enable decentralized set-based operations for secure state estimation without violating the robustness guarantees. Moreover, because Assumption 1(i) allows the attacker to attack up to $q \leqslant p-1$ sensors, it is necessary that the observability is guaranteed from the remaining safe sensors.

Assumption 1(iii) demarcates the class of systems considered in this paper and assumes bounded input bounded state (BIBS) stability. While this may appear to be restrictive in comparison to other secure state estimation schemes, we argue that the class of BIBS stable systems is not restrictive as it is a property all control systems strive to achieve via state or output feedback.

Under the standing assumptions stated above, we study the following problem:

Problem Statement Given the uncertain system (1) subject to Assumption 1, we aim to estimate a set $\hat{\mathcal{X}}_k \subset \mathcal{X}$ guaranteeing the inclusion $x(k) \in \hat{\mathcal{X}}_k$ for every $k \in \mathbb{Z}_{\geqslant 1}$, where $x(k)$ is the true state of (1). Moreover, irrespective of the attack signals $a_i(k)$, $\hat{\mathcal{X}}_k$ must satisfy the following stability condition

$$\text{rad}(\hat{\mathcal{X}}_k) \leqslant \beta(\text{rad}(\mathcal{X}), k) + \gamma(\max\{\text{rad}(\mathcal{W}), \text{rad}(\mathcal{V})\}) \tag{2}$$

where β is a class-$\mathcal{KL}$ function and γ is a class-$\mathcal{K}$ function [3].

This paper achieves the aforementioned secure set-based state estimation problem via modifications to the conventional zonotopic filtering [14], which will be developed in the forthcoming sections.

4 Secure Set-based State Estimation

This section presents our main algorithm for secure set-based state estimation. The algorithm comprises four steps that are summarized in the main loop of Algorithm 1. In the following subsections, we describe each step of the algorithm in detail, provide an error bound, and discuss the complexity issue.

4.1 Time update set

The time update in our proposed algorithm is a one-step computation of the reachability set and is given by

$$\hat{\mathcal{X}}_{k|k-1} = A\hat{\mathcal{X}}_{k-1} \oplus Bu(k-1) \oplus \mathcal{W}. \tag{3}$$

Here, $\hat{\mathcal{X}}_{k-1}$ is obtained in the previous time step through a measurement update, also known as the correction step, which is described in Section 4.3. Given $x(k-1) \in \hat{\mathcal{X}}_{k-1}$, the time update set $\hat{\mathcal{X}}_{k|k-1}$ at time $k \in \mathbb{Z}_{\geqslant 0}$ is the set of states to which the system can evolve subject to the model (A, B), the input $u(k-1)$, and the noise zonotope $\mathcal{W}$. Although the attacker cannot directly influence the time update (3), it can do so indirectly through the previous measurement update $\hat{\mathcal{X}}_{k-1}$. Thus, the resilience against sensor attacks is achieved by carefully devising the measurement update.

4.2 Measurement-consistent set

Before presenting the measurement update, it is crucial to estimate a region in the state space which is consistent with the sensor measurements. In other words, given the measurements $y_1(k), \ldots, y_p(k)$, we estimate a set of points in $\mathbb{R}^{n_x}$, called measurement-consistent set, that could have produced $y_1(k), \ldots, y_p(k)$ up to the measurement noise bounds $\mathcal{V}_1, \ldots, \mathcal{V}_p$.

Given that the system (1) is observable via each subset of sensors with cardinality not more than $p-q$ according to Assumption 1(ii), we find the region of the state space consistent with the measurement $y_{\mathsf{J}}(k)$ provided by the sensor subset $\mathsf{J} \subset \mathbb{Z}_{[1,p]}$, where $y_{\mathsf{J}}(k)$ is obtained by stacking $y_i(k)$, for all $i \in \mathsf{J}$. For instance, if $\mathsf{J} = \{i_1, \ldots, i_{n_{\mathsf{J}}}\}$, then

$$\underbrace{\begin{bmatrix} y^{i_1} \\ \vdots \\ y^{i_{n_{\mathsf{J}}}} \end{bmatrix}}_{y_{\mathsf{J}}(k)} = \underbrace{\begin{bmatrix} C_{i_1} \\ \vdots \\ C_{i_{n_{\mathsf{J}}}} \end{bmatrix}}_{C_{\mathsf{J}}} x + \underbrace{\begin{bmatrix} v_{i_1} \\ \vdots \\ v_{i_{n_{\mathsf{J}}}} \end{bmatrix}}_{v_{\mathsf{J}}} + \begin{bmatrix} a_{i_1} \\ \vdots \\ a_{i_{n_{\mathsf{J}}}} \end{bmatrix}.$$

Here, the stacked vector of measurement noise is

$$v^{\mathsf{J}} \in \mathcal{V}_{\mathsf{J}} = \bigtimes_{i \in \mathsf{J}} \mathcal{V}_i \doteq \langle c_{v^{\mathsf{J}}}, G_{v^{\mathsf{J}}} \rangle$$

where $\times$ denotes the Cartesian product of zonotopes.

Given the output equation (1b), output matrix C_i, and the measurement noise zonotope $\mathcal{V}_i$, a method to find a

[3] A continuous function $\gamma : \mathbb{R}_{\geqslant 0} \to \mathbb{R}_{\geqslant 0}$ is a class $\mathcal{K}$ function, if it is strictly increasing and $\gamma(0) = 0$. A continuous function $\beta : \mathbb{R}_{\geqslant 0} \times \mathbb{R}_{\geqslant 0} \to \mathbb{R}_{\geqslant 0}$ is a class $\mathcal{KL}$ function, if: (i) $\beta(., s)$ is a class $\mathcal{K}$ function for any $s \geqslant 0$; (ii) $\beta(r, .)$ is non-increasing and (iii) $\beta(r, s) \to 0$ as $s \to \infty$ for any $r \geqslant 0$.

subset of the state space $\mathcal{X}$ consistent with each sensor i's measurement $y_i(k)$ is described in [2,3]. In the following, we present its straightforward extension.

Lemma 2 *Let Assumption 1(ii) and (iii) hold. Then, for every $\mathsf{J} \subset \mathbb{Z}_{[1,p]}$ with $|\mathsf{J}| = c_\mathsf{J}$, for some $c_\mathsf{J} \leqslant p-q$, the region $\mathcal{Y}_k^\mathsf{J} \subset \mathbb{R}^{n_x}$ consistent with the measurement $y_\mathsf{J}(k) = C_\mathsf{J} x(k) + v_\mathsf{J}(k) + a_\mathsf{J}(k)$, is given by the measurement-consistent set*

$$\mathcal{Y}_k^\mathsf{J} = \langle c_{y_\mathsf{J}}(k), G_{y_\mathsf{J}}(k) \rangle, \text{ where } \begin{cases} c_{y_\mathsf{J}}(k) = C_\mathsf{J}^\dagger \big(y_\mathsf{J}(k) - c_{v_\mathsf{J}}\big) \\ G_{y_\mathsf{J}}(k) = [\, C_\mathsf{J}^\dagger G_{v_\mathsf{J}} \;\; (I_{n_x} - C_\mathsf{J}^\dagger C_\mathsf{J}) D_x \,] \end{cases} \tag{4}$$

where D_x is the generator matrix of the state space $\mathcal{X}$ given by Assumption 1(iii).

Moreover, if $\mathsf{J} \subseteq \mathsf{S}_k$, where S_k is the index set of uncompromised sensors at time k, then $x(k) \in \mathcal{Y}_k^\mathsf{J}$ for every $k \in \mathbb{Z}_{\geqslant 0}$.

PROOF. Consider a subset $\mathsf{J} \subseteq \mathsf{S}_k$ with $|\mathsf{J}| \leqslant p-q$. Then, $a_i(k) = 0$ for every $i \in \mathsf{J}$. Since the pair (A, C_J) is observable (Assumption 1(ii)), we have that

$$y_\mathsf{J}(k) = C_\mathsf{J} x(k) + v_\mathsf{J}(k)$$

has a non-trivial solution

$$x(k) = C_\mathsf{J}^\dagger (y_\mathsf{J}(k) - v_\mathsf{J}(k)) + (I_{n_x} - C_\mathsf{J}^\dagger C_\mathsf{J}) x(k).$$

Secondly, note that $\mathrm{im}(I_{n_x} - C_\mathsf{J}^\dagger C_\mathsf{J}) = \ker(C_\mathsf{J})$. Finally, $\mathcal{X} = \langle 0, D_x \rangle$ by Assumption 1(iii). Thus, by considering sets on the right hand side and using Minkowski sum operation, we obtain

$$x(k) \in C_\mathsf{J}^\dagger (y_\mathsf{J}(k) - \mathcal{V}_\mathsf{J}) \oplus (I_{n_x} - C_\mathsf{J}^\dagger C_\mathsf{J}) \langle 0, D_x \rangle = \mathcal{Y}_k^\mathsf{J}$$

which completes the proof. □

By Assumption 1(iii), the forward invariant reachable set of system (1) (i.e., state space in which the trajectories of system (1) reside in for all $k \in \mathbb{Z}_{\geqslant 0}$) is given by the zonotope $\mathcal{X} = \langle 0, D_x \rangle$. Subject to this assumption, (4) in the above lemma computes a subset of $\mathcal{X}$ that is consistent with the measurements provided by the sensor subset J. Thus, if the subset J of sensors is attack-free at time k, Lemma 2 guarantees that the true state $x(k)$ is inside the set $\mathcal{Y}_k^\mathsf{J}$. However, the guarantee doesn't hold when the sensor subset J is under attack at time k. By taking the intersection of the consistent sets $\mathcal{Y}_k^\mathsf{J}$, we discard all the sensors whose measurements are corrupted by large valued attack signals. However, sensors that are corrupted by stealthy attack signals, i.e., signals within the noise bounds, remain undetected. Nonetheless, we can ensure that there is at least one subset of sensors with cardinality $p-q$ containing the true state $x(k)$.

Theorem 3 *Let Assumption 1 hold. Then, there exist distinct sets $\mathsf{J}_1, \mathsf{J}_2, \ldots, \mathsf{J}_\eta \subset \mathbb{Z}_{[1,p]}$, each with cardinality $|\mathsf{J}_1| = \cdots = |\mathsf{J}_\eta| = c_\mathsf{J}$, for some $c_\mathsf{J} \leqslant p-q$, such that the agreement set*

$$\mathcal{I}_k = \mathcal{Y}_k^{\mathsf{J}_1} \cap \cdots \cap \mathcal{Y}_k^{\mathsf{J}_\eta} \tag{5}$$

is non-empty and contains the true state $x(k)$, where zonotopes $\mathcal{Y}_k^{\mathsf{J}_\alpha}$ are given in (4) *and*

$$\eta = \binom{p-q}{c_\mathsf{J}} \geqslant 1. \tag{6}$$

PROOF. Let $k \in \mathbb{Z}_{\geqslant 0}$ and $c_\mathsf{J} \leqslant p-q$. Notice that the measurement-consistent sets $\mathcal{Y}_k^{\mathsf{J}_\alpha}$ for any $\mathsf{J}_\alpha \subset \mathbb{Z}_{[1,p]}$ with cardinality $|\mathsf{J}_\alpha| = c_\mathsf{J} > 0$ and $\alpha \in \mathbb{Z}_{[1,\eta]}$ are non-empty according to Lemma 2, because the generator matrices $G_{y_{\mathsf{J}_\alpha}}(k)$ are non-zero. Moreover, the quantity η determines the total number of sensor subsets when c_J number of sensors out of possibly safe $p-q$ sensors are chosen. When $c_\mathsf{J} = p-q$, we have $\eta = 1$, and there exists at least one subset $\mathsf{J}_1 \subset \mathbb{Z}_{[1,p]}$ with cardinality $p-q$ that contains only uncompromised sensors. That is, $\mathsf{J}_1 \subseteq \mathsf{S}_k$ because only q number of sensors can be attacked (Assumption 1(i)) and, by considering $\binom{p}{p-q}$ combinations of sensor subsets, there is at least one subset that doesn't contain any attacked sensor at time k. Therefore, in this case, $\mathcal{I}_k = \mathcal{Y}_k^{\mathsf{J}_1} \neq \varnothing$ and $x(k) \in \mathcal{I}_k$ by Lemma 2. On the other hand, when $c_\mathsf{J} < p-q$, we have $\eta > 1$, and there exist distinct $\mathsf{J}_1, \ldots, \mathsf{J}_\eta \subset \mathbb{Z}_{[1,p]}$, each with cardinality c_J and containing only uncompromised sensors. That is, $\mathsf{J}_1, \ldots, \mathsf{J}_\eta \subset \mathsf{S}_k$ because we can choose at least η subsets of cardinality c_J from the safe, uncompromised sensors S_k at time k. By Assumption 1(ii) and Lemma 2, we have that $x(k) \in \mathcal{Y}_k^{\mathsf{J}_1}$, $x(k) \in \mathcal{Y}_k^{\mathsf{J}_2}$ …, and $x(k) \in \mathcal{Y}_k^{\mathsf{J}_\eta}$. Therefore, the intersection (5) yields a non-empty set containing $x(k)$. □

Remark 4 *If the system* (1) *is redundantly observable for $c_\mathsf{J} = p-q$, then $\eta = 1$. In this case, the agreement set $\mathcal{I}_k$ in* (5) *will equal to the measurement-consistent set $\mathcal{Y}_k^\mathsf{J}$ for some J with $|\mathsf{J}| = p = q$. To have an agreement protocol* (5) *comparing multiple measurement-consistent sets, i.e., $\eta \geqslant 2$, which refines the agreement set $\mathcal{I}_k$ by a set-based voting mechanism, it is necessary that the redundant observability holds for $c_\mathsf{J} < p-q$ and $q < p-1$. For instance, if $c_\mathsf{J} = p-q-1$ and $q = p-2$, then $\eta = p-q = 2$. In this case, the agreement protocol* (5) *consists of a pairwise intersection between two measurement-consistent sets.*

We have shown that the measurement-consistent sets $\mathcal{Y}_k^\mathsf{J}$ formed out of attack-free sensors contain the true state

and, hence, their intersection contains the true state. In the proof of Theorem 3, we saw that by removing the number of attacked sensors q, we guarantee the existence of at least one index set J with cardinality $|\mathsf{J}| = c_\mathsf{J}$ which is attack-free. This allows us to ensure that the agreement set $\mathcal{I}_k$ is non-empty and, moreover, contains the true state. However, in the presence of stealthy attacks, we cannot exclude the measurement-consistent sets formed by the stealthily attacked sensors, as these sets may yield non-empty intersections, where only some of them may contain the true state. Therefore, in the next step, we leverage model-based information to estimate the state space region that contains the true state.

4.3 Measurement update step

We exploit model-based information via computing the time update $\hat{\mathcal{X}}_{k|k-1}$ for trajectories starting from the estimated set $\hat{\mathcal{X}}_{k-1}$ from the previous time step, using (3). Measurement update $\hat{\mathcal{X}}_k$ corrects the conservative estimate of the model-based time update by incorporating new information from the sensor measurements (4).

In the presence of sensor attacks, we saw in the previous section that the formation of the agreement set $\mathcal{I}_k$ is paramount. In particular, to show the state inclusion $x(k) \in \mathcal{I}_k$, it is required that a certain number η of measurement-consistent sets are formed out of attack-free measurements. However, as stated in Assumption 1(i), we only know the maximal number of attacked sensors, but which sensors have been attacked is unknown. Hence, we need to search over all possible intersections of measurement-consistent sets.

Let

$$n_c = \binom{n_\mathsf{J}}{\eta}, \quad n_\mathsf{J} = \binom{p}{c_\mathsf{J}} \tag{7}$$

where we recall that c_J is the cardinality of the index sets J_i. Obtain all the subsets of sensors $\mathsf{J}_1, \dots, \mathsf{J}_{n_\mathsf{J}} \subset \mathbb{Z}_{[1,p]}$ with cardinality c_J. Also, obtain all possible n_c combinations of indices $1, \dots, n_\mathsf{J}$ with size η, where η is given by (6). Let P_h represent each of these combinations for $h = 1, \dots, n_c$, respectively, where $|\mathsf{P}_h| = \eta$. Then, for $h \in \mathbb{Z}_{[1,n_c]}$ and $k \in \mathbb{Z}_{\geqslant 0}$, the agreement sets are obtained as

$$\mathcal{I}_k^h = \bigcap_{\alpha \in \mathsf{P}_h} \mathcal{Y}_k^{\mathsf{J}_\alpha} \tag{8}$$

where $\mathsf{J}_\alpha \subset \mathbb{Z}_{[1,p]}$ with cardinality $c_\mathsf{J} \leqslant p - q$.

Here, notice that n_J is the total number of sensor subsets $\mathsf{J} \subset \mathbb{Z}_{[1,p]}$ with $|\mathsf{J}| = c_\mathsf{J}$ that can be obtained. Further, by Theorem 3, there exists at least one $h \in \mathbb{Z}_{[1,n_c]}$ such that the true state $x(k)$ is included in the agreement set $\mathcal{I}_k^h$; however, we do not know which h due to the unknown set of compromised sensors. Hence, as proposed in (8), it is necessary to check intersections between all the possible combinations of η measurement-consistent sets, which totals to n_c given in (7), i.e., the number of (unordered) ways to choose η sets out of n_J sets.

Then, the measurement update $\hat{\mathcal{X}}_k$ is obtained as

$$\hat{\mathcal{X}}_k = \hat{\mathcal{X}}_{k|k-1} \cap \{\mathcal{I}_k^h\}_{h \in \mathbb{Z}_{[1,n_c]}} \tag{9}$$

where $\mathcal{I}_k^h$ is defined in (8) and the time update set $\hat{\mathcal{X}}_{k|k-1}$ is given in (3) with the following intialization

$$\hat{\mathcal{X}}_{1|0} = A\mathcal{X} \oplus Bu(0) \oplus \mathcal{W}, \quad \mathcal{X} = \langle 0, D_x \rangle.$$

Note that the measurement update set $\hat{\mathcal{X}}_k$ is a collection of multiple constrained zonotopes.

Theorem 5 *Let Assumption 1 hold. Then, for every $k \in \mathbb{Z}_{\geqslant 1}$ and $x(0) \in \mathcal{X}$, the inclusion $x(k) \in \hat{\mathcal{X}}_k$ is guaranteed, where the measurement update $\hat{\mathcal{X}}_k$ is computed in* (9)*.*

PROOF. We prove this result by induction. By Assumption 1(iii), we have $x(0) \in \mathcal{X} = \langle 0, D_{n_x} \rangle$. Therefore,

$$x(1) \in \hat{\mathcal{X}}_{1|0} = A\mathcal{X} \oplus Bu(0) \oplus \mathcal{W}$$

because $w(0) \in \mathcal{W}$. Moreover, by Theorem 3, there exists $h \in \mathbb{Z}_{[1,n_c]}$ such that $x(1) \in \mathcal{I}_1^h$. Therefore, from (9), we have

$$x(1) \in \hat{\mathcal{X}}_1 = \hat{\mathcal{X}}_{1|0} \cap \{\mathcal{I}_1^h\}_{h \in \mathbb{Z}_{[1,n_c]}}.$$

This, in turn, implies that

$$x(2) \in \hat{\mathcal{X}}_{2|1} = A\hat{\mathcal{X}}_1 \oplus Bu(1) \oplus \mathcal{W}$$

which is the one-step reachability or time update set (3). Now, assume $x(k') \in \hat{\mathcal{X}}_{k'|k'-1}$ for some $k' \in \mathbb{Z}_{\geqslant 2}$. Then, by Theorem 3, there exists $h \in \mathbb{Z}_{[1,n_c]}$ such that $x(k') \in \mathcal{I}_{k'}^h$, implying

$$x(k') \in \hat{\mathcal{X}}_{k'} = \hat{\mathcal{X}}_{k'|k'-1} \cap \{\mathcal{I}_{k'}^h\}_{h \in \mathbb{Z}_{[1,n_c]}}.$$

Therefore, we have the inclusion

$$x(k'+1) \in \hat{\mathcal{X}}_{k'+1|k'} = A\hat{\mathcal{X}}_{k'} \oplus Bu(k') \oplus \mathcal{W}.$$

Thus, the proof is completed because we showed that, for every $k \in \mathbb{Z}_{\geqslant 1}$, $x(k-1) \in \hat{\mathcal{X}}_{k-1|k-2}$ implies $x(k-1) \in \hat{\mathcal{X}}_{k-1}$, which, in turn, implies $x(k) \in \hat{\mathcal{X}}_{k|k-1}$. Hence, $x(k) \in \hat{\mathcal{X}}_k$. □

Although the above theorem guarantees the inclusion of the true state, it is important to remark that the number of sets in the measurement update (9) may increase with respect to time under stealthy attacks. We address this

issue in Section 4.5 by proposing several techniques that facilitate the computational efficiency of the algorithm.

It is worth mentioning that the proposed algorithm is resilient because the attacker cannot deteriorate the estimation accuracy over time. If an index set J is such that the reading of sensor $i \in \mathsf{J}$ is injected with a large-valued attack signal, it will be automatically discarded because the corresponding measurement-consistent set $\mathcal{Y}_k^{\mathsf{J}}$ will not intersect with either its counterparts or the time update set. Therefore, in order to yield a non-empty agreement set $\mathcal{I}_k^h$ containing the attacked sensors, the attacker can only inject small attack signals whose magnitude is within the measurement noise bounds $\mathcal{V}_i$, which, therefore, maintains the estimation accuracy.

The proposed algorithm is summarized in Algorithm 1.

Algorithm 1 Secure set-based state estimation

Require: System matrices A, B, and C_i, and noise zonotopes $\mathcal{W}$ and $\mathcal{V}_i$, for every $i \in \mathbb{Z}_{[1,p]}$; time sequence of sensor measurements $\{y^1(k), y^2(k), \dots, y^p(k)\}_{k \in \mathbb{Z}_{\geqslant 0}}$.

1: Initialize: $\hat{\mathcal{X}}_0 \subseteq \mathcal{X}$
2: **for** $k = 1, 2, 3, \dots$ **do**
3: Obtain the time update $\hat{\mathcal{X}}_{k|k-1}$ using (3).
4: Obtain the measurement-consistent sets $\mathcal{Y}_k^{\mathsf{J}_\alpha}$, for $\alpha = 1, \dots, n_{\mathsf{J}}$, where $\mathsf{J}_\alpha \subset \mathbb{Z}_{[1,p]}$ with $|\mathsf{J}_\alpha| = c_{\mathsf{J}}$ and $\mathsf{J}_\alpha \neq \mathsf{J}_{\alpha'}$ for $\alpha \neq \alpha'$, using (4).
5: Obtain the agreement sets $\mathcal{I}_k^h$, for $h = 1, \dots, n_c$, using (8).
6: Obtain the measurement update $\hat{\mathcal{X}}_k$ using (9).
7: **end for**

We illustrate Algorithm 1 on a simple example of $p = 3$ sensors, where $q = 1$ sensor has been corrupted. For the purposes of this illustration in Figure 1, we chose the first sensor y_1 to be corrupted. Suppose that the system is observable via every sensor. Therefore, the measurement consistent sets $\mathcal{Y}_k^{\mathsf{J}_1}$,$\mathcal{Y}_k^{\mathsf{J}_2}$,and $\mathcal{Y}_k^{\mathsf{J}_3}$ can be obtained from sensors $y_{\mathsf{J}_1}, y_{\mathsf{J}_2}, y_{\mathsf{J}_3}$, respectively, with $\mathsf{J}_1 = \{1\}$, $\mathsf{J}_2 = \{2\}$ and $\mathsf{J}_3 = \{3\}$. Based on the illustration in Figure 1, the agreement sets are $\mathcal{I}_k^1 = \mathcal{Y}_k^{\mathsf{J}_1} \cap \mathcal{Y}_k^{\mathsf{J}_2} = \varnothing$; $\mathcal{I}_k^2 = \mathcal{Y}_k^{\mathsf{J}_1} \cap \mathcal{Y}_k^{\mathsf{J}_3}$ and $\mathcal{I}_k^3 = \mathcal{Y}_k^{\mathsf{J}_2} \cap \mathcal{Y}_k^{\mathsf{J}_3}$ are depicted as the yellow untextured region in Figure 1. Finally, the estimated set $\hat{\mathcal{X}}_k$ is the union of the textured green regions in Figure 1.

4.4 Bound on the estimation error

Since Theorem 5 guarantees that the true state $x(k)$ of system (1) lies in at least one of the zonotopes in the measurement update $\hat{\mathcal{X}}_k$ at each $k \in \mathbb{Z}_{\geqslant 0}$, it must also lie in a zonotope that overbounds $\hat{\mathcal{X}}_k$. That is, let the collection of the constrained zonotopes in $\hat{\mathcal{X}}_k$ be overbounded by a zonotope

$$\hat{\mathcal{Z}}_k = \langle \hat{c}_z(k), \hat{G}_z(k) \rangle \tag{10}$$

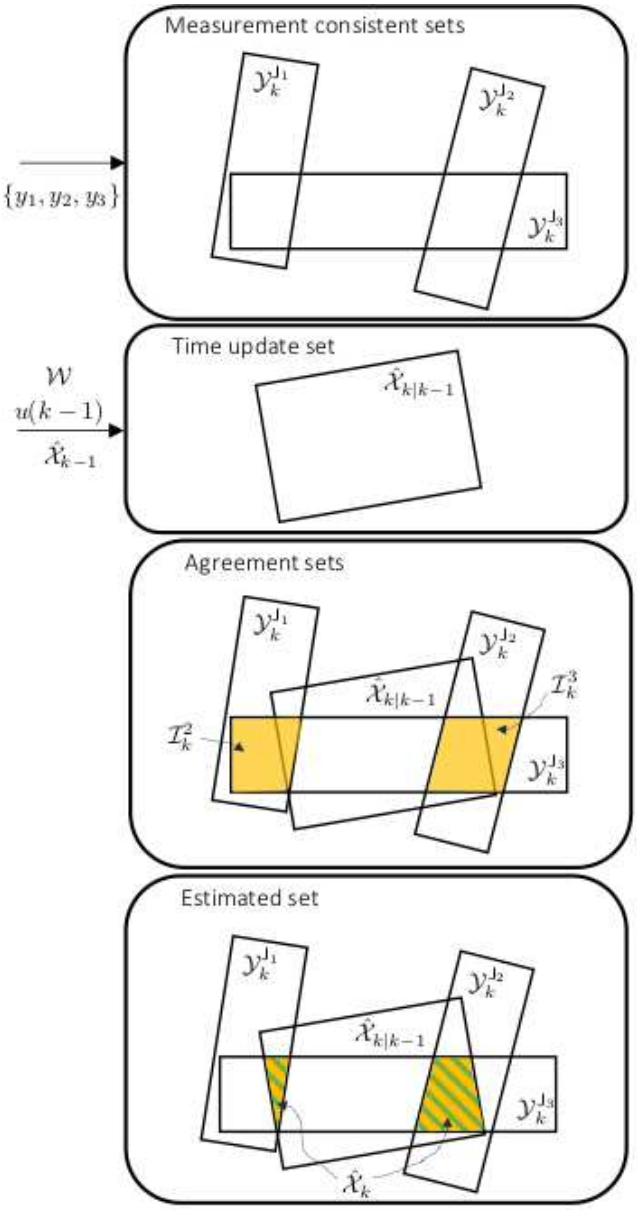


Fig. 1. Illustration of Algorithm 1 with $p = 3$ and $q = 1$, where sensor 1 has been attacked.

which is obtained by solving

$$\min \operatorname{rad}(\hat{\mathcal{Z}}_k) \text{ subject to } \hat{\mathcal{X}}_k \subseteq \hat{\mathcal{Z}}_k. \tag{11}$$

Then, the estimation error can be bounded by

$$\|\hat{c}_z(k) - x(k)\| \leqslant \operatorname{rad}(\hat{\mathcal{Z}}_k).$$

The stability analysis of this error bound is provided in Section 5 by using a point-based secure estimator. Nonetheless, we remark that this error bound is significantly smaller in practice than the error bounds obtained by point-based secure estimators [16, 28].

4.5 Methods to reduce the complexity

The major computational challenge of Algorithm 1 that the attacker can exploit lies in the measurement update step (9) for computing $\hat{\mathcal{X}}_k$, which is a collection of zonotopes whose cardinality (i.e., the number of zonotopes) could grow over time. To reduce computational complexity resulting from the increasing cardinality of the measurement update, we propose several pruning methods below.

The first obvious step is to remove any empty sets or subsets of other sets in the measurement update intersection (9). It is also possible to obtain a single overbounding zonotope of $\hat{\mathcal{X}}_k$ as in (11), and use it in the next time update step [6]. However, a better trade-off between accuracy and complexity is to not overbound the whole collection but only the intersecting zonotopes in the collection $\hat{\mathcal{X}}_k$. This may not make the cardinality of $\hat{\mathcal{X}}_k$ equal to one, but it reduces it significantly by allowing minimal loss of accuracy.

Another method is to employ zonotope reduction methods [38] to reduce the number of generators in the zonotopes, which are often increased when performing the Minkowski sum operations. However, this technique may result in a larger radius of $\hat{\mathcal{Z}}_k$ in (11) and can also yield conservative estimates.

Finally, one could employ a point-based resilient estimator, if it exists, in parallel with the set-based secure estimator. In this case, we may consider only those candidates in the measurement update collection that lie within the intersection of $\hat{\mathcal{X}}_k$ and an error margin generated by a point-based resilient state estimator. However, the existing point-based resilient state estimators [12, 16, 18, 28, 34] require that the total number of sensors be strictly greater than twice the number of compromised sensors $q < p/2$ and the members of the attacked sensors also remain unchanged over time, which are tighter requirements than our standing Assumption 1(i). Moreover, the error margins obtained by point-based estimators are usually very conservative. Nonetheless, such a technique is useful in the stability analysis of the proposed secure set-based state estimator, as discussed in the next section.

5 Stability Analysis

To assist in the stability analysis of the secure zonotopic state estimation algorithm, in the sense that the estimated set $\hat{\mathcal{X}}_k$ containing the true state $x(k)$ is bounded with respect to time $k \in \mathbb{Z}_{\geqslant 0}$ given that the initial set $\mathcal{X}$ is bounded, we employ a secure *point-based* state estimator in parallel. Such an estimator provides a point-based estimate of the state and ensures that the bound on the estimation error is unaffected by the attack signal. In the following, we define a point-based secure estimator and use it to modify our proposed algorithm, which ensures that the set-based estimate is stable in the sense of (2).

5.1 Point-based secure state estimator

In the measurement update (9), we observe that the number of zonotopes in the collection given by the estimated set $\hat{\mathcal{X}}_k$ could be very large. To limit the number of candidates in our collection of zonotopes in the measurement update set $\hat{\mathcal{X}}_k$ and consequently, guarantee stability of our algorithm, we make use of a point-based estimator which is *secure* with respect to the attack signals a_i and robust with respect to disturbances w and noise v as follows.

Assumption 6 *There exists a secure point-based state estimator $E : \mathbb{R}^m \times \mathcal{U} \to \mathbb{R}^{n_x}$ which provides a point-based state estimate $\hat{x}(k) = E(y^1(k), \ldots, y^p(k), u(k))$ to system (1) satisfying the following for $k \in \mathbb{Z}_{\geqslant 0}$,*

$$\|\hat{x}(k) - x(k)\| \leqslant \beta(\|\hat{x}(0) - x(0)\|, k) + \alpha \max(\bar{w}, \bar{v}) \quad (12)$$

for some class $\mathcal{KL}$ function β and $\alpha \in \mathbb{R}_{\geqslant 0}$, where $\|w(k)\| \leqslant \bar{w}$ and $\|v(k)\| \leqslant \bar{v}$ for all $k \in \mathbb{Z}_{\geqslant 0}$.

Note that in Assumption 6, the upper bound on the state estimation error is not affected by the attack signals a_i. We call estimators that possess this property *secure* and results exist in the literature for linear (e.g., [13, 16, 20, 26]). So far, these results require at most $q \leqslant \lceil p/2 \rceil - 1$ number of attacked sensors, i.e., strictly less than half the number of sensors can be attacked. Thereby, this imposes a stricter requirement than Assumption 1(i), which we state below.

Assumption 7 *The number of attacked sensors $q \in \mathbb{Z}_{\geqslant 0}$ satisfies $2q < p$, where $p \in \mathbb{Z}_{\geqslant 0}$ is the total number of sensors. The integers p and q are known, but the exact attacked sensors is unknown.*

In the following section, we describe how we use a point-based *secure* state estimator to guarantee the stability of the algorithm by pruning the collection of estimated states $\hat{\mathcal{X}}_k$ at each iteration $k \in \mathbb{Z}_{\geqslant 0}$.

5.2 Modified measurement update

Modification of Algorithm 1, Step 6: We modify the measurement update by the following

$$\hat{\mathcal{X}}_k^{\text{mod}} = \hat{\mathcal{X}}_{k|k-1} \cap \{\mathcal{I}_k^h\}_{h \in \mathbb{Z}_{[1,n_c]}} \cap \mathcal{H}(\hat{x}(k), \Delta_k) \quad (13)$$

where $\hat{\mathcal{X}}_{k|k-1}, \mathcal{I}_k^h$ come from (3) and (8), respectively; the point-based state estimate $\hat{x}(k)$ is provided by a *secure*

state estimator satisfying Assumption 6; and the radius is

$$\Delta_k \doteq \beta(\|x(0) - \hat{x}(0)\|, k) + \alpha \max(\bar{w}, \bar{v}), \tag{14}$$

where $\beta \in \mathcal{KL}$, $\alpha \in \mathbb{R}_{\geqslant 0}$, $\bar{w} \in \mathbb{R}_{\geqslant 0}$ and $\bar{v} \in \mathbb{R}_{\geqslant 0}$ come from Assumption 6. Figure 2 illustrates the modified measurement update (13).

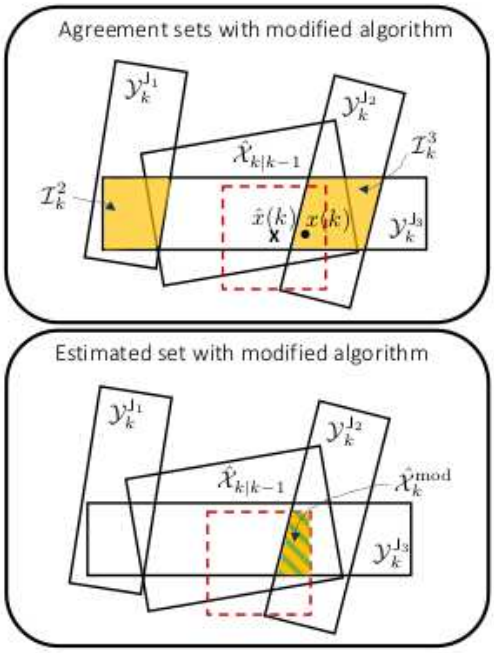


Fig. 2. Modified measurement update

It is important to note that the practicality of this pruning algorithm relies on reasonable estimates of $\beta \in \mathcal{KL}$ and $\alpha > 0$, which is usually difficult to obtain. For linear systems, they can be obtained easily; however, they are usually quite conservative because of their dependence on the condition number of a certain matrix. Nonetheless, this technique is useful in bounding the estimation error.

To this end, we define the estimation error in terms of the center of $\hat{\mathcal{Z}}_k^{\text{mod}} := \langle \hat{c}_z^{\text{mod}}(k), \hat{G}_z^{\text{mod}}(k) \rangle$, the minimal radius zonotope that overbounds $\hat{\mathcal{X}}_k^{\text{mod}}$ in (13) with maximal radius Δ_k, i.e.,

$$\begin{aligned} &\min \ \text{rad}(\hat{\mathcal{Z}}_k^{\text{mod}}) \leqslant \Delta_k \\ &\text{subject to } \hat{\mathcal{X}}_k^{\text{mod}} \subseteq \hat{\mathcal{Z}}_k^{\text{mod}}. \end{aligned} \tag{15}$$

Since $\lim_{k\to\infty} \Delta_k = \alpha \max(\bar{w}, \bar{v})$,

$$\lim_{k\to\infty} \text{rad}(\hat{\mathcal{Z}}_k^{\text{mod}}) \leqslant \alpha \max(\bar{w}, \bar{v}),$$

which only depends on the bounds on the process and measurement noise, and not on the attack signals.

5.3 Bound on the estimation error

We can bound the estimation error $\hat{c}_z^{\text{mod}}(k) - x(k)$ as follows.

Proposition 8 *Let Assumptions 1(ii)-(iii), 6, and 7 hold. Consider Algorithm 1 with a modified measurement update* (13)*. Then, for every $k \in \mathbb{Z}_{\geqslant 0}$, the overall estimation error bound is given by the following*

$$\|\hat{c}_z^{\text{mod}}(k) - x(k)\| \leqslant \Delta_k \tag{16}$$

where Δ_k is defined in (14) *and $\hat{c}_z^{\text{mod}}(k)$ is the center of $\hat{\mathcal{Z}}_k^{\text{mod}}$ given by* (15)*.*

PROOF. Note that $x(k) \in \hat{\mathcal{X}}_{k|k-1} \cap \{\mathcal{I}_k^h\}_{h\in\mathbb{Z}_{[1,n_c]}}$ according to Theorem 5 and $x(k) \in \mathcal{H}(\hat{x}(k), \Delta_k)$ according to Assumption 6. Therefore, we have that $x(k) \in \hat{\mathcal{X}}_k^{\text{mod}}$. Since $\hat{\mathcal{X}}_k^{\text{mod}} \subseteq \hat{\mathcal{Z}}_k^{\text{mod}}$ because of (15), we obtain (16). □

Therefore, Proposition 8 enables us to conclude that by implementing a secure point-based estimator with properties given by Assumption 6, we can guarantee that the estimation error satisfies (16) with a bound Δ_k that asymptotically converges to a bound that does not depend on the attack signals a_i, which is unknown, but only depends on the bounds on the process and measurement noise, which are assumed to be known. This is a guarantee that is absent in the purely secure set-based estimation algorithm (Algorithm 1) developed in the previous section.

However, we now have the more restrictive Assumption 7 (in comparison to Assumption i) that guarantees the existence of a resilient point-based estimator possessing the property in Assumption 6. We summarise the assumptions and guarantees provided by Algorithm 1 and its modification in the following section.

5.4 Comparison and guarantees

We have chosen to develop the secure set-based state estimation scheme in two stages, first in Section 4, then providing a modified algorithm by incorporating a point-based state estimation algorithm in Section 5, to show that while the estimated set $\hat{\mathcal{X}}_k$, $k \in \mathbb{Z}_{\geqslant 0}$ always contains the true state $x(k)$, we can ensure that its radius asymptotically decreases up to a margin of error that depends only the process and measurement noise, which is assumed to be known. This is achieved with the more restrictive assumption on the allowed number of attacked sensors, which is needed to implement the modified algorithm that incorporates a *secure* point-based state estimator. We summarise this in Table 1.

Table 1
Comparison between the measurement update in Algorithm 1 and the modified measurement update in Section 4.3.

	Secure set-based state estimator (Algorithm 1)	*Modified measurement update* (Section 5.2)
Constraint on the number of attacks	$q < p$ (Assumption 1(i))	$2q < p$ (Assumption 7)
Guarantees on the estimation error	$\hat{\mathcal{X}}_k$ is bounded at every time k.	$\hat{\mathcal{X}}_k^{\text{mod}}$ is bounded by a shrinking ball Δ_k in (14), which converges to a smaller ball of radius $\alpha \max(\bar{w}, \bar{v})$ irrespective of the attack signals.

6 Discussion on Attack Detection and Computational Complexity

6.1 Algorithm to detect attacked sensors

A notable contribution of the set-based state estimation scheme in this paper over other secure schemes is Assumption 1(i), which allows the attacker to compromise not only up to $p-1$ sensors at each time instant but also different subsets of sensors at different times. To this end, we remark that we can detect only those compromised sensors that are injected with non-stealthy attack signals. The detection algorithm is fairly simple and can be summarized in Algorithm 2.

Algorithm 2 Detection of a subset of attacked sensors at time $k \in \mathbb{Z}_{\geqslant 1}$

Require: Time update $\hat{\mathcal{X}}_{k|k-1}$, measurement-consistent sets $\mathcal{Y}_k^{\mathsf{J}_1}, \ldots, \mathcal{Y}_k^{\mathsf{J}_{n_\mathsf{J}}}$, and a collection $\{\mathsf{P}_h\}_{h \in \mathbb{Z}_{[1,n_c]}}$ containing $n_c = \binom{n_\mathsf{J}}{\eta}$ combinations of indices.

1: Initialize $\hat{\mathsf{S}}_k = \varnothing$.
2: **for** $h = 1, \ldots, n_c$ **do**
3: Obtain $\mathcal{I}_k^h$ using (5).
4: **if** $\hat{\mathcal{X}}_{k|k-1} \cap \mathcal{I}_k^h \neq \varnothing$ **then**
5: Estimated safe subset $\hat{\mathsf{S}}_k \leftarrow \hat{\mathsf{S}}_k \cup \big(\bigcup_{\alpha \in \mathsf{P}_h} \mathsf{J}_\alpha\big)$.
6: **end if**
7: **end for**
8: Detected attacked sensors $\hat{\mathsf{A}}_k = \mathbb{Z}_{[1,p]} \backslash \hat{\mathsf{S}}_k$.

At time k and for $h \in \mathbb{Z}_{[1,n_c]}$, Algorithm 2 checks if both the agreement set $\mathcal{I}_k^h$ and its intersection with the time update $\hat{\mathcal{X}}_{k|k-1}$ are non-empty. If that is the case, then the sensors with indices in $\bigcup_{\alpha \in \mathsf{P}_h} \mathsf{J}_\alpha$ are either safe or compromised with a stealthy attack signal. Otherwise, the set $\bigcup_{\alpha \in \mathsf{P}_h} \mathsf{J}_\alpha$ contains at least one attacked sensor. By checking all the combinations P_h and storing a 'potentially' safe subset of sensors in $\hat{\mathsf{S}}_k$ at every iteration, a subset of attacked sensors $\hat{\mathsf{A}}_k$ are estimated as those sensors that are not in $\hat{\mathsf{S}}_k$.

Remark 9 *The estimated safe subset $\hat{\mathsf{S}}_k$ contains the true safe subset S_k at time k, i.e., $\mathsf{S}_k \subseteq \hat{\mathsf{S}}_k$. If there are sensors that are injected with small-valued stealthy attack signals, our detection algorithm fails to recognize those attacks and considers those sensors to be uncompromised. Therefore, the detected attacked sensors $\hat{\mathsf{A}}_k$ is only a subset of the true subset of attacked sensors.*

Remark 10 *In Algorithm 2, we detect a subset of attacked sensors at every time k. However, in the time-invariant attack setting where the attacker does not change the subset of compromised sensors, Algorithm 2 can be adapted to cumulatively detect and remove the attacked sensors over time. This can have application in sensor fault detection and isolation as faults can be modeled as naive attacks.*

6.2 Attack detection under naive attacks/sensor faults

Algorithm 1 automatically discards attacked sensors when a naive attacker injects large or random attack signals. Consequently, Algorithm 2 can detect such attacks. These attack signals have no effect on the estimated set $\hat{\mathcal{X}}_k$ since the measurement-consistent sets corresponding to the attacked sensors are automatically excluded in (5) and (9), resulting in empty intersections. Even random attack signals that fall within the noise bounds can be detected eventually if the attacker is not intelligent enough to account for the changing orientation and position of the time update set and the measurement-consistent sets with respect to time.

Our proposed framework is well-suited for sensor faults like a complete failure, sensor deterioration, intermittent transmissions, and random bias, all of which can be treated as naive attacks and detected by Algorithm 2. Random attack signals, on the other hand, can be viewed as a consequence of the attacker's limited knowledge of the system and/or noise bounds. Such an attack could also arise because the attacker has limited resources and cannot generate an optimal attack signal at every time instance to guarantee the worst-case complexity, as discussed in the subsequent subsection.

Under the assumption of naive attacks and/or sensor faults, the attacked sensors can be easily detected and discarded, which results in a significant reduction of complexity of the measurement update step (9). The proposed framework is, therefore, well-suited for robust system monitoring, providing a reliable defense against attacks and faults that could compromise system performance and integrity.

6.3 Worst-case complexity under stealthy attacks

In the worst-case scenario, stealthy attacks on sensors may result in non-empty intersections in (8), resulting in an empty detected set $\hat{\mathsf{A}}_k$. Moreover, the collection of sets obtained from the intersections (5) and (9) can also increase the complexity of Algorithm 1 exponentially, which can overwhelm the available computation resources.

Precisely, if at every time instant $k \in \mathbb{Z}_{\geqslant 0}$, the agreement sets $\mathcal{I}_k^h, \mathcal{I}_k^{h'}$ are non-empty and distinct for $h \neq h'$, and they all intersect with the time update $\hat{\mathcal{X}}_{k|k-1}$, then the number of sets in the measurement update $\hat{\mathcal{X}}_k$ turn out to be $|\hat{\mathcal{X}}_{k|k-1}| \times n_c$, where $|\hat{\mathcal{X}}_{k|k-1}|$ denotes the number of constrained zonotopes in the time update and n_c is given in (7). Note that, since initially $|\hat{\mathcal{X}}_{1|0}| = 1$, we have $|\hat{\mathcal{X}}_1| = n_c$ and, by using (3), $|\hat{\mathcal{X}}_{2|1}| = n_c$, implying $|\hat{\mathcal{X}}_2| = n_c^2$. Thus, it is now straightforward to see that the number of constrained zonotopes in the measurement update $\hat{\mathcal{X}}_k$ is n_c^k in the worst-case scenario.

It is important to mention that generating such an intelligent stealthy attack is quite difficult for an attacker. First, it requires that the attacker has complete knowledge of not only the system but also the estimation algorithm. Second, it assumes that the attacker has ample computational resources to compute a feasible attack signal within one time sample guaranteeing that (i) all agreement sets are non-empty, (ii) they are distinct, and (iii) they all intersect with the time update. From an attacker's perspective, this turns out to be a very difficult task as it not only imposes multiple hard constraints but also requires the computation within significantly small time. Further investigation into the computational complexity of generating such an attack is out of the scope of the current paper and will be addressed in the future.

Apart from the worst-case, there is no doubt that the number of constrained zonotopes in the measurement update could increase with time. Therefore, at every time instant, it is crucial to employ methods discussed in Section 4.5 to reduce the complexity and ensure the computational feasibility of the proposed algorithm. Since each complexity reduction method offers a trade-off between estimation accuracy and complexity, the best method is the one that offers maximum accuracy under the available computational resources.

7 Numerical Simulation

We evaluated our proposed algorithms on two examples. The first illustrative one lacks observability from each sensor, which is provided in the three-story building structure example. The CORA toolbox [6] is used to generate our simulations. We start by analyzing the illustrative example.

7.1 Illustrative example

In this example, we consider a simple two-dimensional linear system without a control input $u(t) \equiv 0$, where the system matrices are given by

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \qquad \begin{aligned} C_1 &= [\,1\ 0\,],\ C_2 = [\,1\ 1\,], \\ C_3 &= [\,0\ 1\,],\ C_4 = [\,1\ 2\,]. \end{aligned}$$

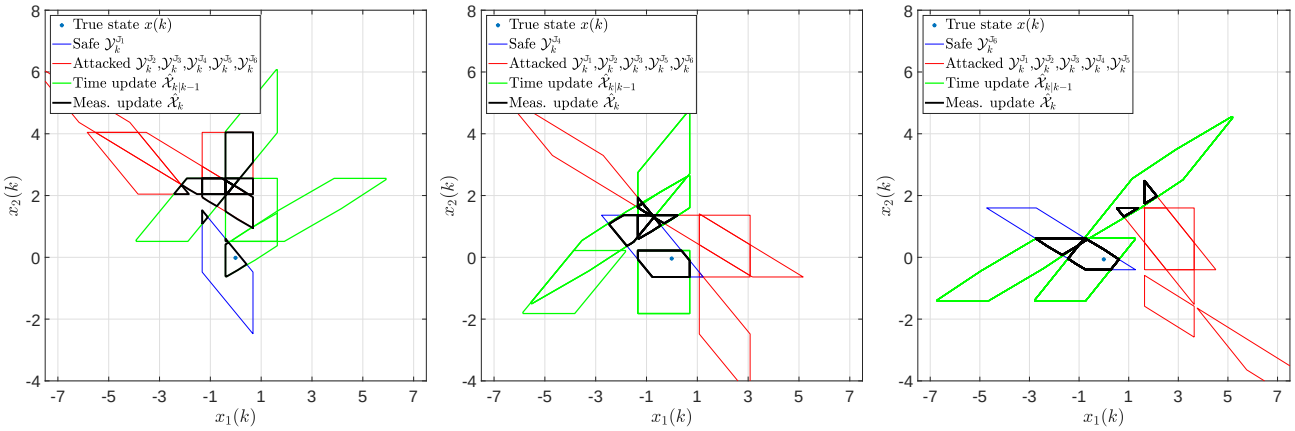
The number of sensors are $p = 4$, and we suppose that the attacker can target $q = 2$ number of sensors at any time instant. In this case, since $q = p/2$, the point-based state estimators that require $q < p/2$ cannot be employed. The process noise bound is $\mathcal{W} = \langle 0, \mathrm{diag}([0.02\,0.02]) \rangle$. We have the measurement noise bounds of the four sensors as follows.

$$\mathcal{V}_1 = \mathcal{V}_2 = \mathcal{V}_3 = \mathcal{V}_4 = \langle 0, 1 \rangle,$$

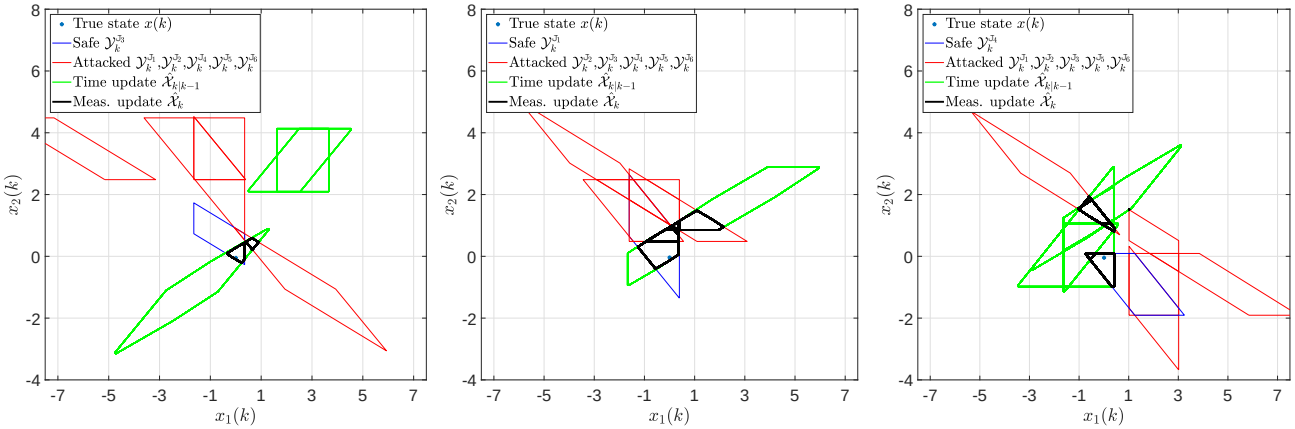
Notice that Assumption 1(ii) holds for $c_\mathsf{J} = p - q = 2$ because the pair $\left(A, \begin{bmatrix} C_i^\top & C_j^\top \end{bmatrix}^\top\right)$ is observable for any $i, j = 1, 2, 3, 4$ and $i \neq j$. Hence, we consider the following subsets J_α of sensors with cardinality $|\mathsf{J}_\alpha| = c_\mathsf{J}$, where $\alpha = 1, \ldots, n_\mathsf{J}$ with $n_\mathsf{J} = 6$,

$$\begin{aligned} &\mathsf{J}_1 = \{1, 2\}, \quad &&\mathsf{J}_2 = \{1, 3\}, \quad &&\mathsf{J}_3 = \{1, 4\} \\ &\mathsf{J}_4 = \{2, 3\}, \quad &&\mathsf{J}_5 = \{2, 4\}, \quad &&\mathsf{J}_6 = \{3, 4\}. \end{aligned}$$

For these sets, we compute the measurement-consistent sets $\mathcal{Y}_k^{\mathsf{J}_\alpha}$ using (4). Notice that since $\eta = 1$, we do not need to perform intersections in (8) to compute the agreement sets $\mathcal{I}_k^h$. In this example, it turns out that $\mathcal{I}_k^h = \mathcal{Y}_k^{\mathsf{J}_\alpha}$, where $h = \alpha$ and $h, \alpha \in \{1, \ldots, 6\}$ since $n_c = n_\mathsf{J} = 6$. Then, the time and measurement updates can be computed using (3) and (9). Fig. 3 illustrates the measurement-consistent sets $\mathcal{Y}_k^{\mathsf{J}_\alpha}$ (safe in blue and attacked in red), time update $\hat{\mathcal{X}}_{k|k-1}$ (green), measurement update $\hat{\mathcal{X}}_k$ (blue), and the true state $x(k)$ for different sets of sensors attacked at different times in which the attacker chooses two sensors at each time step in a rotating manner. At time step $k = 2$, we have in Figure 3a sensors 3 and 4 under attack, i.e., we have $\mathcal{Y}_k^{\mathsf{J}_\alpha}$, for all $\alpha \in \{2, \ldots, 6\}$ under attack. Then, sensors 1 and 4 are attacked in Fig. 3b in which the true state is still included in the measurement update set $\hat{\mathcal{X}}_k$. We should note that some attacked measurement-consistent sets are detected and isolated in Fig. 3c as they do not intersect with any of the time update sets. Interestingly, the measurement update sets have small volumes in Fig. 3d, which still contain the true state $x(k)$. Similarly, we present the computed sets in Fig. 3e and 3f.



(a) Time $k = 2$. Sensors $\{3, 4\}$ are attacked. **(b)** Time $k = 3$. Sensors $\{1, 4\}$ are attacked. **(c)** Time $k = 4$. Sensors $\{1, 2\}$ are attacked.



(d) Time $k = 5$. Sensors $\{2, 3\}$ are attacked. **(e)** Time $k = 6$. Sensors $\{3, 4\}$ are attacked. **(f)** Time $k = 7$. Sensors $\{1, 4\}$ are attacked.

Fig. 3. Snapshots of estimated sets using Algorithm 1 under time-varying attack, where different sensors are attacked at different time steps.

7.2 Three-story building structure

We now consider a three-story building structure of [36] described by a mechanical system

$$M\ddot{q}(t) + D\dot{q}(t) + Sq(t) = Ge(t), \tag{17}$$

where $q(t) \in \mathbb{R}^3$ is the vector of relative horizontal displacements of the floors and $e(t) \in \mathbb{R}$ is the ground acceleration due to earthquake. Also, $M \in \mathbb{R}^{3\times3}$ is the mass matrix, $D \in \mathbb{R}^{3\times3}$ is the damping matrix, $S \in \mathbb{R}^{3\times3}$ is the stiffness matrix, and $G \in \mathbb{R}^3$ is the loading vector. The parameter values of the system (17) are provided by [36, Appendix A] as:

$$M = \mathrm{diag}([\,478350\ 478350\ 517790\,]) \quad (\mathrm{kg})$$

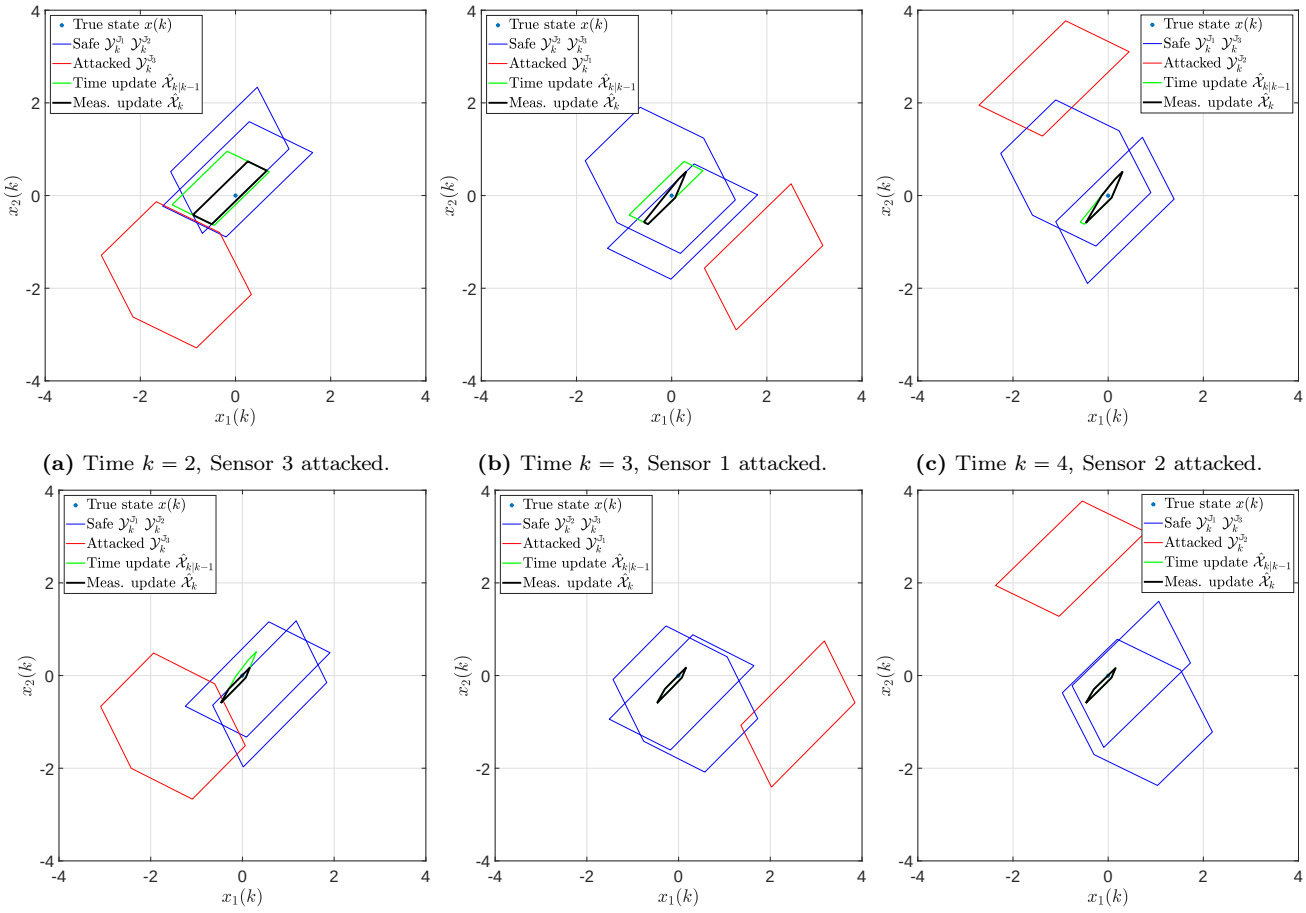
$$D = 10^5 \times \begin{bmatrix} 7.7626 & -3.7304 & 0.6514 \\ -3.7304 & 5.8284 & -2.0266 \\ 0.6514 & -2.0266 & 2.4458 \end{bmatrix} \quad (\mathrm{Ns/m})$$

$$S = 10^8 \times \begin{bmatrix} 4.3651 & -2.3730 & 0.4144 \\ -2.3730 & 3.1347 & -1.2892 \\ 0.4144 & -1.2892 & 0.9358 \end{bmatrix} \quad (\mathrm{N/m})$$

$$G = [\,478350\ 478350\ 517790\,]^T \quad (\mathrm{kg}).$$

By considering the state $x(t) = [\,q(t)^T\ \ \dot{q}(t)^T\,]^T$, we can obtain the state-space representation in continuous time

$$\dot{x} = A_c x + E_c e$$



(a) Time $k = 2$, Sensor 3 attacked. **(b)** Time $k = 3$, Sensor 1 attacked. **(c)** Time $k = 4$, Sensor 2 attacked.

(d) Time $k = 5$, Sensor 3 attacked. **(e)** Time $k = 6$, Sensor 1 attacked. **(f)** Time $k = 7$, Sensor 2 attacked.

Fig. 4. Snapshots of estimated sets using Algorithm 1 under time-varying attack, where different sensors are attacked at different time steps.

where

$$A_c = \begin{bmatrix} 0_{3\times 3} & I_3 \\ -M^{-1}S & -M^{-1}D \end{bmatrix}, \quad E_c = \begin{bmatrix} 0_{3\times 1} \\ -M^{-1}G \end{bmatrix}.$$

After discretization with sample time δ, we obtain the system in the form (1a), where

$$A = \exp(A_c\delta)$$
$$w(k) = A_c^{-1}(A - I_6)E_c e(k).$$

Notice that we do not consider a control input in this example, i.e., $Bu(k) \equiv 0$. Here, our goal is to monitor the building dynamics under an earthquake, which is assumed to be the process noise or disturbance, using secure set-based state estimation.

We assume that each floor of the building is equipped with a sensor, i.e., $p = 3$, that measures the relative displacement and the velocity of that floor, which can be collected in the output vector $y_i(k) \in \mathbb{R}^3$ as given by (1b), for $i \in \mathbb{Z}_{[1,3]}$, where

$$C_1 = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

$$C_2 = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

$$C_3 = \begin{bmatrix} -1 & 0 & 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

We suppose that the attacker can compromise only one at each time, i.e., $q = 1$. Assumption 1(ii) holds for $c_\mathcal{J} = 1$ because all the pairs $(A, C_1), (A, C_2), (A, C_3)$ are observable. Then, our goal is to monitor the building displacements and velocities irrespective of the compro-

mised sensor. The noise bound $e(k) \in \langle 0, 500 \rangle$. We have the measurement noise bounds of the four sensors as follows.

$$\mathcal{V}_1 = \mathcal{V}_2 = \mathcal{V}_3 = \langle 0, \text{diag}([1\ 1\ 1]) \rangle.$$

To illustrate the efficacy of our algorithm, we apply a powerful, time-varying attack to our system in which the attacker randomly chooses a sensor i at every time step k and injects false data into its measurement $y_i(k)$. In Fig. 4a, Sensor 3 is under attack with a large attack value, and we have one estimated measurement update set (black). Then, Sensor 1 is attacked in Fig. 4b in which the number of estimated measurement update sets is increasing due to having a small attack value. Finally, other sensors are attacked in Fig. 4c–4f. Although the complexity increases in such attacks, it is worth noting that the true state $x(k)$ remains enclosed by the estimated measurement update sets at all time steps. Also, the estimation error remains bounded, and the attacker cannot destroy the accuracy of the set-based state estimate.

8 Conclusion and future work

We presented a novel set-based state estimation algorithm that can estimate the system state even when all but one sensor could be compromised by an adversary. Our proposed algorithm overcomes the limitation of point-based secure estimators that restrict the number of attacked sensors to strictly less than half the total number of sensors. We achieved this by constructing agreement sets from the intersection of various combinations of measurement consistent sets. We showed that our algorithm guarantees the inclusion of the true state in the estimated set, provided that the system remains observable from every combination of the number of uncompromised sensors. Moreover, we proposed a simple algorithm to identify the set of compromised sensors, which can aid in addressing intelligent and stealthy attacks.

While our algorithm's worst-case complexity may increase exponentially under intelligent and stealthy attacks, we argued that it is challenging for attackers to execute such attacks due to the requirement of a complete understanding of the system and algorithm and substantial computational resources. We suggested various strategies to reduce the complexity of our algorithm to facilitate its implementation.

We also incorporated a point-based resilient state observer into our algorithm to prune the candidate sets when less than half of the sensors have been attacked. This strategy's effectiveness depends on an accurate approximation of the guaranteed estimation error provided by the resilient point-based state observer. Nonetheless, the modified algorithm provides asymptotic convergence guarantees with an explicit bound that depends on the known process and measurement noise, independent of the attack signals. Our future work will focus on the set-based secure state estimation of nonlinear systems and developing a data-driven approach for secure estimation when the system model is unknown.

References

[1] T. Alamo, J. M. Bravo, and E. F. Camacho. Guaranteed state estimation by zonotopes. *Automatica*, 41(6):1035–1043, 2005.

[2] A. Alanwar, A. Berndt, K. H. Johansson, and H. Sandberg. Data-driven set-based estimation using matrix zonotopes with set containment guarantees. In *2022 European Control Conference (ECC)*, pages 875–881, 2022.

[3] A. Alanwar, M. U. B. Niazi, and K. H. Johansson. Data-driven set-based estimation of polynomial systems with application to SIR epidemics. In *2022 European Control Conference (ECC)*, pages 888–893, 2022.

[4] A. Alanwar, J. J. Rath, H. Said, and M. Althoff. Distributed set-based observers using diffusion strategy. *arXiv:2003.10347*, 2020.

[5] A. Alanwar, H. Said, and M. Althoff. Distributed secure state estimation using diffusion kalman filters and reachability analysis. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pages 4133–4139. IEEE, 2019.

[6] M. Althoff. An introduction to CORA 2015. In *Proceedings of the Workshop on Applied Verification for Continuous and Hybrid Systems*, 2015.

[7] M. Althoff, G. Frehse, and A. Girard. Set propagation techniques for reachability analysis. *Annual Review of Control, Robotics, and Autonomous Systems*, 4:369–395, 2021.

[8] M. Althoff and J. J. Rath. Comparison of guaranteed state estimators for linear time-invariant systems. *Automatica*, 130:109662, 2021.

[9] J. Blesa, V. Puig, and J. Saludes. Robust fault detection using polytope-based set-membership consistency test. *IET Control Theory & Applications*, 6(12):1767–1777, 2012.

[10] P. Bouron, D. Meizel, and P. Bonnifait. Set-membership nonlinear observers with application to vehicle localisation. In *2001 European Control Conference (ECC)*, pages 1255–1260, 2001.

[11] A. Chen, T. Ngoc Dinh, T. Raissi, and Y. Shen. Outlier-robust set-membership estimation for discrete-time linear systems. *International Journal of Robust and Nonlinear Control*, 32(4):2313–2329, 2022.

[12] M. S. Chong, H. Sandberg, and J. P. Hespanha. A secure state estimation algorithm for nonlinear systems under sensor attacks. In *2020 59th IEEE Conference on Decision and Control (CDC)*, pages 5743–5748, 2020.

[13] M. S. Chong, M. Wakaiki, and J. P. Hespanha. Observability of linear systems under adversarial attacks. In *2015 American Control Conference (ACC)*, pages 2439–2444, 2015.

[14] A. A. de Paula, G. V. Raffo, and B. O. Teixeira. Zonotopic filtering for uncertain nonlinear systems: Fundamentals, implementation aspects, and extensions [applications of control]. *IEEE Control Systems Magazine*, 42(1):19–51, 2022.

[15] H. Fawzi, P. Tabuada, and S. Diggavi. Secure estimation and control for cyber-physical systems under adversarial attacks. *IEEE Transactions on Automatic Control*, 59(6):1454–1467, 2014.

[16] X. He, X. Ren, H. Sandberg, and K. H. Johansson. How to secure distributed filters under sensor attacks. *IEEE Transactions on Automatic Control*, 67(6):2843–2856, 2021.

[17] L. Jaulin. Robust set-membership state estimation: Application to underwater robotics. In *Automatica*, volume 45, pages 202–206, 2009.

[18] J. Kim, C. Lee, H. Shim, Y. Eun, and J. H. Seo. Detection of sensor attack and resilient state estimation for uniformly observable nonlinear systems having redundant sensors. *IEEE Transactions on Automatic Control*, 64(3):1162–1169, 2018.

[19] V. T. H. Le, C. Stoica, T. Alamo, E. F. Camacho, and D. Dumur. Zonotopic guaranteed state estimation for uncertain systems. *Automatica*, 49(11):3418–3424, 2013.

[20] J. G. Lee, J. Kim, and H. Shim. Fully distributed resilient state estimation based on distributed median solver. *IEEE Transactions on Automatic Control*, 65(9):3935–3942, 2020.

[21] J. Li, Y. Wang, X. Liu, F. Yao, and M. Zhang. Zonotopic resilient state estimation for unmanned surface vehicles subject to integrity attacks. *Ocean Engineering*, 273:113934, 2023.

[22] X. Li, G. Wei, and L. Wang. Distributed set-membership filtering for discrete-time systems subject to denial-of-service attacks and fading measurements: A zonotopic approach. *Information Sciences*, 547:49–67, 2021.

[23] L. Liu, L. Ma, Y. Wang, J. Zhang, and Y. Bo. Distributed set-membership filtering for time-varying systems under constrained measurements and replay attacks. *Journal of the Franklin Institute*, 357(8):4983–5003, 2020.

[24] L. Liu, L. Ma, J. Zhang, and Y. Bo. Distributed non-fragile set-membership filtering for nonlinear systems under fading channels and bias injection attacks. *International Journal of Systems Science*, 52(6):1192–1205, 2021.

[25] N. Meslem and A. Hably. Robust set-membership state estimator against outliers in data. *IET Control Theory & Applications*, 14(13):1752–1761, 2020.

[26] A. Mitra and S. Sundaram. Byzantine-resilient distributed observers for LTI systems. *Automatica*, 108:108487, 2019.

[27] M. U. B. Niazi, A. Alanwar, M. S. Chong, and K. H. Johansson. Resilient set-based state estimation for linear time-invariant systems using zonotopes. *European Journal of Control*, page 100837, 2023.

[28] M. Pajic, I. Lee, and G. J. Pappas. Attack-resilient state estimation for noisy dynamical systems. *IEEE Transactions on Control of Network Systems*, 4(1):82–92, 2016.

[29] T. Raïssi, D. Efimov, and A. Zolghadri. Interval state estimation for a class of nonlinear systems. *IEEE Transactions on Automatic Control*, 57(1):260–265, 2011.

[30] B. S. Rego, S. G. Vrachimis, M. M. Polycarpou, G. V. Raffo, and D. M. Raimondo. State estimation and leakage detection in water distribution networks using constrained zonotopes. *IEEE Transactions on Control Systems Technology*, 2021.

[31] J. K. Scott, D. M. Raimondo, G. R. Marseglia, and R. D. Braatz. Constrained zonotopes: A new tool for set-based estimation and fault detection. *Automatica*, 69:126–136, 2016.

[32] H. Shim, J. Back, Y. Eun, G. Park, and J. Kim. Zero-dynamics attack, variations, and countermeasures. In *Security and Resilience of Control Systems: Theory and Applications*, pages 31–61. Springer, 2022.

[33] T. Shinohara and T. Namerikawa. Reach set-based secure state estimation against sensor attacks with interval hull approximation. *SICE Journal of Control, Measurement, and System Integration*, 11(5):399–408, 2018.

[34] Y. Shoukry, P. Nuzzo, A. Puggelli, A. L. Sangiovanni-Vincentelli, S. A. Seshia, and P. Tabuada. Secure state estimation for cyber-physical systems under sensor attacks: A satisfiability modulo theory approach. *IEEE Transactions on Automatic Control*, 62(10):4917–4932, 2017.

[35] H. Song, P. Shi, C.-C. Lim, W.-A. Zhang, and L. Yu. Set-membership estimation for complex networks subject to linear and nonlinear bounded attacks. *IEEE Transactions on Neural Networks and Learning Systems*, 31(1):163–173, 2019.

[36] T. H. Truong, P. Seiler, and L. E. Linderman. Analysis of networked structural control with packet loss. *IEEE Transactions on Control Systems Technology*, 30(1):344–351, 2021.

[37] Z. Wang, C.-C. Lim, and Y. Shen. Interval observer design for uncertain discrete-time linear systems. *Systems & Control Letters*, 116:41–46, 2018.

[38] X. Yang and J. K. Scott. A comparison of zonotope order reduction techniques. *Automatica*, 95:378–384, 2018.

[39] Y. Zhang, Y. Zhu, and Q. Fan. A novel set-membership estimation approach for preserving security in networked control systems under deception attacks. *Neurocomputing*, 400:440–449, 2020.

[40] Y. Zhu, H. Liu, C. Li, and J. Yu. Consensus and security control of multi-agent systems based on set-membership estimation with time-varying topology under deception attacks. *International Journal of Control, Automation and Systems*, 20(11):3624–3636, 2022.