2023 62nd IEEE Conference on Decision and Control (CDC)
December 13-15, 2023. Marina Bay Sands, Singapore

Set-Based Anomaly Detector and Stealthy Attack Impact using Constrained Zonotopes

Jonas Wagner, Tanner Kogel, and Justin Ruths

Abstract— **In this paper, a set-based detector is developed and analyzed based on the propagation of nominal (no attacks), actual (possibly with attacks), and perceived (corrupted by attacks) residuals due to bounded system and measurement noise uncertainties. The set of stealthy attacks (attacks that raise no alarms) is characterized and the impact of these stealthy attacks on the system state is quantified. When implemented through the set tools of constrained zonotopes, this approach provides accurate, time-varying attack-reachable sets that can be used to evaluate the safety and performance of systems in adversarial conditions. These tools are demonstrated through two case studies.**

I. Introduction

Model-based anomaly detection leverages knowledge about the system dynamics to identify when measurements indicate a departure from normal behavior. Detection in a deterministic setting with exact models is trivial since system behavior can be predicted perfectly. Detection becomes challenging in uncertain environments. Modeling this uncertainty has naturally taken two avenues: distribution-based and set-based. While distribution-based modeling has many advantages, the tools developed thus far are complicated to scale to nonlinear dynamics, more complex uncertainties, and transient behavior. Work along these lines has represented more complex distributions using moment families [1] and Gaussian Mixture Models [2], or proposed more general types of detectors based on Wasserstein or Kullback–Leibler Divergence discrepancies [3]. Alternative set-based tools offer a compelling opportunity to model uncertainty propagation in transient and nonlinear systems.

The maturity and expressiveness of set representations play a large role in the success of reachability analysis for dynamic systems. Ellipsoids are relatively simple and concise set structures highly suitable for certain applications, but lacks fine-grain resolution [4], [5]. Zonotopes are a versatile representation of convex polytopes, but the strict requirement for symmetry greatly limits their usefulness in many applications. In 2016, constrained zonotopes were introduced, allowing the symmetry to be broken and also enabling constrained zonotopes to be closed under set intersections [6]. Since then use of constrained zonotopes has grown in popularity in a variety of applications, including reachability of dynamic systems [7].

Recently, several approaches have been proposed to build estimators using zonotopes and constrained zonotopes, e.g., [8]–[11], including the original constrained zonotope paper which develops an estimator and demonstrates how it can be used as a fault detection methodology [6]. In much of this line of work, measurement sets (introduced by measurement noise) are used to corroborate forward reachable sets (introduced by current state and system noise uncertainty). This means that state set propagation forward in time is computed online and estimator values are updated through set intersections as new innovations arrive. Therefore, although the concept of developing a set-based anomaly detector with constrained zonotopes is not new, the approach outlined in this paper is novel and is leveraged to address a different line of questioning.

This paper proposes using forward reachability analysis under nominal operation (no faults/no attacks) to be used as a comparative baseline and forming the core of the proposed set-based detector. Crucially, set propagation is done in the error space using measurement residuals which removes the effect of exact state values, i.e., the reachable set of estimation error is tracked as opposed the state or estimate themselves. Because the nominal uncertainty reachability can be done offline, the sets of attack sequences, i.e. attack sequences that guaranteed no alarms under the set-based detector, can be studied. This offline analysis also provides fundamental information about the system vulnerability rather than only being informed by scenario or sample realizations. This paper outlines a complete framework to design and optimize stealthy attacks while also quantifying the impact upon the system state, estimate, and error compared to nominal operation.

II. System Definition

This paper focuses on discrete-time (DT) linear time invariant (LTI) systems described by

$$x_{k+1} = Ax_k + Bu_k + \nu_k. \tag{1}$$

For each time-step, $k \in \mathbb{N}$, the system state $x_k \in \mathbb{R}^n$ is evolved by the previous state according to state-update matrix $A \in \mathbb{R}^{n \times n}$, the input signal $u_k \in \mathbb{R}^m$ according to input matrix $B \in \mathbb{R}^{n \times m}$, and the bounded system noise $\nu_k \in V_k \subset \mathbb{R}^n$. The pair (A, B) is assumed to be stabilizable.

The noisy, and potentially corrupted, measurement is given by output equation

$$y_k = Cx_k + \eta_k + a_k, \tag{2}$$

This material is in part based upon work supported by the National Science Foundation under Grant No. CMMI-2143485.

The authors are with the Mechanical Engineering Department at the University of Texas at Dallas, Richardson, TX, USA jonas.wagner@, tanner.kogel@, jruths@utdallas.edu

979-8-3503-0123-6/23/$31.00 ©2023 IEEE

where the measurement, $y_k \in \mathbb{R}^p$, is dependent on the system state according to output matrix $C \in \mathbb{R}^{p\times n}$, the bounded measurement noise $\eta_k \in H_k \subset \mathbb{R}^p$, and the attack signal $a_k \in \mathbb{R}^p$. The pair (A, C) is assumed to be detectable.

An estimated state is provided by a static gain Luenberger observer,

$$\hat{x}_{k+1} = A\hat{x}_k + Bu_k + L(y_k - C\hat{x}_k), \tag{3}$$

where the estimated state $\hat{x}_k \in \mathbb{R}^n$ is updated each time-step by modifying the predicted evolution, $A\hat{x}_k + Bu_k$, by the residual, $r_k = y_k - C\hat{x}_k$, according to observer gain $L \in \mathbb{R}^{n\times p}$. The gain L is selected to drive the estimation error, $e_k = x_k - \hat{x}_k$, to zero; thus $(A - LC)$ is stable.

An estimate feedback controller is used to stabilize the system according to

$$u_k = K\hat{x}_k, \tag{4}$$

where the feedback gain $K \in \mathbb{R}^{m\times n}$ is selected to stabilize the closed loop system, thus $(A + BK)$ is stable.

The DT-LTI plant (1), with measurement (2), state estimator (3), and feedback controller (4) has the closed-loop state and error dynamics given by

$$\begin{cases} x_{k+1} = (A + BK)x_k - BKe_k + \nu_k, \\ e_{k+1} = (A - LC)e_k + \nu_k - L\eta_k + La_k, \\ \quad y_k = Cx_k + \eta_k + a_k. \end{cases} \tag{5}$$

III. Set-based Detector

This section proposes a novel set-based detector that checks to ensure the residuals are feasible based on nominal operating conditions, $a_k = 0$. The following is a review several basic set operations and their definitions:

Definition 1 (General Set Operations): Let sets $X, Y \subset \mathbb{R}^n$, $W \subset \mathbb{R}^m$, vector $v \in \mathbb{R}^n$, and matrix $R \in \mathbb{R}^{m\times n}$. The set operations are defined as follows: (i) the linear transformation of X under R is defined as $RX = \{Rx \mid x \in X\}$; (ii) the vector sum of v to X is $v + X = \{v + x \mid x \in X\}$; (iii) the Minkowski sum of X and Y is $X \oplus Y = \{x + y \mid x \in X, y \in Y\}$; (iv) the generalized intersection of X and W under R is $X \cap_R W = \{x \in X \mid Rx \in W\}$; and (v) the standard intersection of X and Y, corresponding to $X \cap_{\mathbf{I}_n} Y$, is denoted as $X \cap Y$.

Set-based Detector: For the closed-loop system (5), the residual, $r_k = y_k - C\hat{x}_k$ is tested to ensure it could be produced under nominal operation, i.e., while the projected nominal error $C\bar{e}_k$ is within the reachable nominal error set projected onto the measurement space $C\bar{\mathcal{E}}_k$, and raise an alarm otherwise:

$$\begin{cases} r_k \in C\bar{\mathcal{E}}_k \oplus H_k & \text{no alarm} \\ \text{otherwise} & \text{alarm} \end{cases} \tag{6}$$

The nominal error set, $\bar{\mathcal{E}}_k$ is updated according to

$$\bar{\mathcal{E}}_{k+1} = (A - LC)\bar{\mathcal{E}}_k \oplus V_k \oplus -LH_k. \tag{7}$$

The following result justifies the performance of this proposed detector.

Theorem 1: The detector (6) generates no alarms under nominal operation, $a_k = 0$.

Proof: With $a_k = 0$, the measurement (2) becomes

$$y_k = Cx_k + \eta_k = C(\hat{x}_k + \bar{e}_k) + \eta_k, \tag{8}$$

where $\bar{e}_k = x_k - \hat{x}_k$ is the estimation error when $a_k = 0$. Similarly, the residual can be expressed as

$$r_k = y_k - C\hat{x}_k = C(\hat{x}_k + \bar{e}_k) + \eta_k - C\hat{x}_k = C\bar{e}_k + \eta_k. \tag{9}$$

Since the detector aims to account for all possible residual values, set notation is used to represent all possible vectors of the nominal error, $\bar{e} \in \bar{\mathcal{E}}_k$. The evolution of this reachability set (7) is derived from (5) where $a_k = 0$.

Moreover, since the measurement noise is bounded by $\eta_k \in H_k$, the residual is similarly bounded as

$$r_k \in C\bar{\mathcal{E}}_k \oplus H_k, \tag{10}$$

and therefore no alarms would be triggered. ■

IV. Stealthy Attacks

To consider worst-case attack impact, this paper assumes that attacker is able to measure and manipulate the sensor values while also having access to the state estimate and control signal. Such capability and access would be possible if the attack has access to the local network or is installed as malware on the system controller. If attackers execute large, obvious attacks, they can have large and immediate impact on systems, but risk revealing their presence. The defending strategy in this scenario instead becomes one of mitigation. This paper considers the alternative motivation, that an attacker wishes to remain undetected, but still aims to disrupt system behavior. This motivates the notion of stealthiness.

Definition 2 (Stealthiness): A sequence of attach signals, $\{a_k\}_{k=0}^{N-1}$, is considered ***stealthy*** to a particular detector for N time steps if no alarms are raised for $k = 0, \ldots, N$.

A. Stealthy Attack Design

As discussed in previous work, one method to ensure stealthiness is to exploit the reliance of the detector on the residual and ensure that the attack will never trigger the alarm [5], [12]. We follow this same strategy for this new set-based framework.

Stealthy Attack Design: For residual-based detectors, the attack signal, a_k, can be defined to cancel out the measurement residual, $r_k = y_k - C\hat{x}_k$, and replace it with a designed attack signal, $\delta_k \in \Delta_k \subset \mathbb{R}^p$, using

$$a_k = -r_k + \delta_k = -(y_k - C\hat{x}_k) + \delta_k. \tag{11}$$

Note that implementation of this attack requires the attacker to have access to the noisy measurement $Cx_k + \eta_k$ and the estimate $\hat{x}_k$ but not the noise realization, η_k, on its own.

Remark 1: The primary advantage of the attack given in (11) is that the measurement noise is absorbed into a_k. This separates the set-based uncertainty propagation from real-time operation. A consequence of this is that it allows δ_k

to be designed offline since stealthiness is not dependent on operating conditions.

For a sequence of designed attack signals, $\{\delta_k\}_{k=0}^{N-1}$, the stealthy attack set, Δ_k, is defined for a particular detector as all the possible δ_k that guarantee stealthiness $\forall_{k=1,\dots,N}$.

B. Ensuring Attack Stealthiness

For a given detector framework, an attack sequence is guaranteed to be stealthy whenever no alarm is triggered regardless of system or sensor noise.

Theorem 2: The attack sequence (11) injected into the system (5) is guaranteed to be stealthy to the set-based detector (6) as long as

$$\delta_k \in \Delta_k = C\bar{\mathcal{E}}_k \oplus H_k, \tag{12}$$

where

$$\bar{\mathcal{E}}_k = (A-LC)^k\bar{\mathcal{E}}_0 \oplus \bigoplus_{i=0}^{k-1}(A-LC)^{k-1-i}(V_i \oplus -LH_i). \tag{13}$$

Proof: The attacked measurement, (**??**), results in an attacked residual of

$$r_k = y_k - C\hat{x}_k = (C\hat{x}_k + \delta_k) - C\hat{x}_k = \delta_k. \tag{14}$$

Since no alarm is triggered by the detector, (6) when $r_k \in C\bar{\mathcal{E}}_k \oplus H_k$, $\delta_k \in C\bar{\mathcal{E}}_k \oplus H_k$ implies the attack is stealthy for each individual time-step, thus (12). Further, the explicit definition of $\bar{\mathcal{E}}_k$ (13) can be derived from the recursive definition in (7), where $\bar{e}_0 \in \bar{\mathcal{E}}_0$. ■

C. Stealthy Attack Impact

By quantifying the stealthy attack sets Δ_k, it then enables the forward propagation of these attacks into the estimation error and then into the system state. This provides a quantification of the attack impact that can be used to assess and guarantee safety.

Substituting the stealthy attack signal, (11), into the closed loop system dynamics, (5), results in the attacked estimation error update equation,

$$e_{k+1} = Ae_k + \nu_k - L\delta_k. \tag{15}$$

Thus the attacked error set, $\mathcal{E}_k$, evolves according to

$$\mathcal{E}_{k+1} = A\mathcal{E}_k \oplus V_k \oplus -L\Delta_k. \tag{16}$$

Corollary 1 (Error Stealthy Reachable Set): For the system (5), the detector (6) ensures that the error reachable set under any stealthy attack is given by

$$\mathcal{E}_k = A^k\mathcal{E}_0 \oplus \bigoplus_{i=0}^{k-1} A^{k-1-i}\big(V_i \oplus -L(C\bar{\mathcal{E}}_i \oplus H_i)\big), \tag{17}$$

where $\bar{\mathcal{E}}_i$ is defined in (13).

Proof: An explicit definition for $\mathcal{E}_k$ can be derived from (16) as $\mathcal{E}_k = A^k\mathcal{E}_0 \oplus \bigoplus_{i=0}^{k-1} A^{k-1-i}(V_i \oplus -L\Delta_i)$ and then (17) can be derived by plugging in (12). ■

Corollary 2 (Error Stealthy Steady State Reachable Set): When A is stable and the noise sets are constant, $V_k = V$ and $H_k = H$), the set $\mathcal{E}_k$ will converge to

$$\mathcal{E}_\infty = (I-A)^{-1}\Big(V \oplus -L(I - A + LC)^{-1}(V \oplus -LH)\Big) \tag{18}$$

Proof: For static inputs, the steady-state response for the system plant, (1), is calculated as $x_\infty = \lim_{k\to\infty} x_k$. This limit can be found by calculating when $x_\infty = Ax_\infty + Bu$. This is equivalent to $(I-A)x_\infty = Bu$ and since A is stable, $(I-A)$ is invertible and thus $x_\infty = (I-A)^{-1}Bu$. Extending this to the set update equations in (7) and (17) results in

$$\bar{\mathcal{E}}_\infty = (I - (A - LC))^{-1}(V \oplus -LH) \tag{19}$$

$$\Delta_\infty = C\bar{\mathcal{E}}_\infty \oplus V \tag{20}$$

$$\mathcal{E}_\infty = (I-A)^{-1}(V \oplus -L\Delta_\infty) \tag{21}$$

which can then be used to derive (18). ■

Remark 2: Note that the strict stability of A, $|\text{eig}(A)| < 1$, is required for invertibility of $I - A$ and to ensure that the stealthy reachable set converges.

Corollary 3 (State Stealthy Reachable Set): For the system (5) with detector (6), the state reachable set under any stealthy attack is given by

$$\mathcal{X}_k = (A+BK)^k\mathcal{X}_0 \bigoplus_{i=0}^{k-1}(A+BK)^{k-1-i}(V_i \oplus -BK\mathcal{E}_i) \tag{22}$$

where $\mathcal{E}_i$ is defined by (17).

Proof: From (5), the set update for all reachable states under stealthy attacks, $x_k \in \mathcal{X}_k$, evolves according to

$$\mathcal{X}_{k+1} = (A+BK)\mathcal{X}_k \oplus -BK\mathcal{E}_k \oplus V_k. \tag{23}$$

(22) is then derived as an explicit definition of (23). ■

Note that the error and state reachable sets are driven by the combination of attack sets Δ_k and system noise sets V_k.

V. Implementation with Constrained Zonotopes

The proposed approach is agnostic to the set representation; however it is the versatility of the constrained zonotope set representation that enables the further usefulness of this approach. Specifically, complicated noise characteristics and nonlinearities motivates potentially needing asymmetric uncertainty sets. Moreover, the reachable sets can also be used for a variety of subsequent analyses - many of which would require set intersections and asymmetric operations, such as intersections with dangerous/critical states. Thus, constrained zonotopes are used to implement the proposed set-based detector effectively.

Definition 3 (Constrained Zonotopes [6]): A constrained zonotope (CG-rep) in $\mathbb{R}^n$ is defined by

$$Z_c = \{G, c, A, b\} = \{G\xi + c \mid \|\xi\|_\infty \leq 1, A\xi = b\} \tag{24}$$

with center $c \in \mathbb{R}^n$, generator matrix $G \in \mathbb{R}^{n \times n_g}$ consisting of n_g generators g_i, and $A \in \mathbb{R}^{n_c \times n_g}$ and $b \in \mathbb{R}^{n_c}$ describing n_c equality constraints.

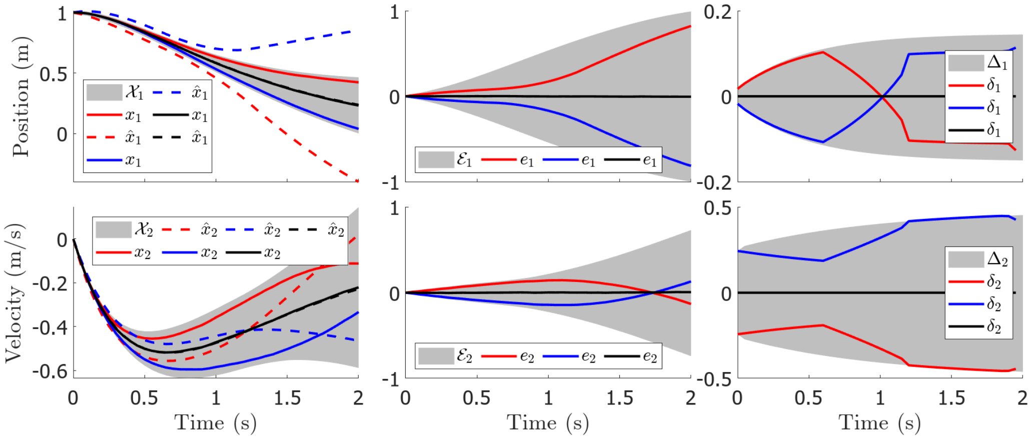

Fig. 1. Comparison of trajectories of the spring-mass-damper system under stealthy attack sequences over $2\,\mathrm{s}$ from rest at an initial position of $x_0 = \hat{x}_0 = 1\,\mathrm{m}$. Each column plots the reachable sets (in gray) and trajectories for state x (position x_1, velocity x_2) and state estimate $\hat{x}$ (left), attacked state error e (center), and attack signal δ (right). The black trajectories are the nominal case when no attack is present, while the red and blue trajectories correspond to attack sequences designed to maximize and minimize the position overtime, respectively.

The set operations of linear transformation, Minkowski sum, and generalized intersection for constrained zonotopes (CG-rep) are defined by (25), (26), and (27) respectively.

$$RX = \{RG_x, Rc_x, A_x, b_x\} \tag{25}$$

$$X \oplus Y = \left\{ \begin{bmatrix} G_x & G_y \end{bmatrix}, c_x + c_y, \begin{bmatrix} A_x & 0 \\ 0 & A_y \end{bmatrix}, \begin{bmatrix} b_x \\ b_y \end{bmatrix} \right\} \tag{26}$$

$$X \cap_R W = \left\{ \begin{bmatrix} G_x & 0 \end{bmatrix}, c_x, \begin{bmatrix} A_x & 0 \\ 0 & A_w \\ RG_x & -G_w \end{bmatrix}, \begin{bmatrix} b_x \\ b_w \\ c_w - Rc_x \end{bmatrix} \right\} \tag{27}$$

The set-based detector, (6), is implemented using CG-rep and the set operations (25)-(27). The nominal error is represented in CG-rep as $\bar{\mathcal{E}}_k = \{G_{\bar{\mathcal{E}}_k}, c_{\bar{\mathcal{E}}_k}, A_{\bar{\mathcal{E}}_k}, b_{\bar{\mathcal{E}}_k}\}$ and updated according to (7). For the alarm (6), an inclusion test can be performed using a simple linear program as described in Proposition 2 of [6].

VI. Numerical Case Studies

Two case studies are used to demonstrate the concepts of the set-based detector and its utility. These numerical examples were generated in MATLAB using the ConZono MATLAB toolbox [13], where optimization problems were solved using YALMIP [14] and Gurobi [15].

A. Spring-Mass-Damper Illustrative Example

A linear spring-mass-damper model is used to demonstrate the implementation of the set-based detector (6) and the effect of the stealthy attack strategy (11) on a two-state open-loop stable system. The state reachability under stealthy attacks are compared for different measurement methods and the ability for a stealthy attack to modify operation is demonstrated by attacks aiming to minimize and maximize the position.

An underdamped spring-mass-damper system is defined with physical constants of mass $m = 1\,\mathrm{kg}$, spring constant $k = 1\,\mathrm{N/m}$, and linear damping $b = 0.25\,\mathrm{N/m^2}$, modeled as a sampled-data-system with a zero-order hold and discretization of $0.05\,\mathrm{s}$. The resultant DT-LTI plant is defined by (1) with matrices

$$A = \begin{bmatrix} 0.9988 & 0.04967 \\ -0.04967 & 0.9863 \end{bmatrix}, \quad B = \begin{bmatrix} -0.001245 \\ -0.04967 \end{bmatrix}.$$

The system noise is bounded by a constrained zonotope, $\nu_k \in V = \{10^{-4}G_V, 10^{-5}c_V, 10^{-3}A_V, 10^{-4}b_V\}$, with

$$G_V = \begin{bmatrix} 0.5 & 1 & 0 & 0 \\ 2 & -0.5 & 0 & 0 \end{bmatrix}, \quad c_V = \begin{bmatrix} -2.5 \\ -2.5 \end{bmatrix}$$

$$A_V = \begin{bmatrix} -0.5 & -1 & 1.25 & 0 \\ -2 & 0.5 & 0 & 1.75 \end{bmatrix}, \quad b_V = \begin{bmatrix} -2.5 \\ -7.5 \end{bmatrix}.$$

The system is fully observed, (2) with $C = \mathbf{I}_2$, and the measurement noise is bounded as $\eta_k \in H = \{G_H, 10^{-4}c_H\}$ with

$$G_H = \begin{bmatrix} 0.005 & 0 \\ 0 & 0.25 \end{bmatrix}, \quad c_H = \begin{bmatrix} -0.5 \\ 5 \end{bmatrix}.$$

The observer, (3), is designed to ensure $\mathrm{eig}(A - LC) = \{0.9, 0.95\}$ and the estimate feedback controller, (4), is designed with pole placement to ensure $\mathrm{eig}(A + BK) = \{0.9, 0.95\}$ resulting in

$$L = \begin{bmatrix} 0.0988 & 0.0497 \\ -0.0497 & 0.0363 \end{bmatrix}, \quad K = \begin{bmatrix} 1.0129 & 2.6946 \end{bmatrix}.$$

The sets Δ_k, $\mathcal{E}_k$, and $\mathcal{X}_k$ are computed offline (Remark 1) which allows the sequence δ_k to also be calculated offline as an open-loop optimization problem. The results are visualized in Fig. 1 by projecting the sets onto each state individually (shown in gray). The stealthy attack sequences δ_k (Fig. 1, right column) designed to minimize (blue) or maximize (red) the position (x_1) are compared against the nominal no-attack case (black). For the first 0.5 s, the attack sequences hug the lower and upper boundary of the set in order to disrupt the position measurement as much as possible. The attack sequence then transitions to predominantly disrupt the velocity for the final 0.75 s. This results in the estimation error e_k (Fig. 1, center column) to approach the boundary of the reachable set $\mathcal{E}_k$, first in velocity and then position.

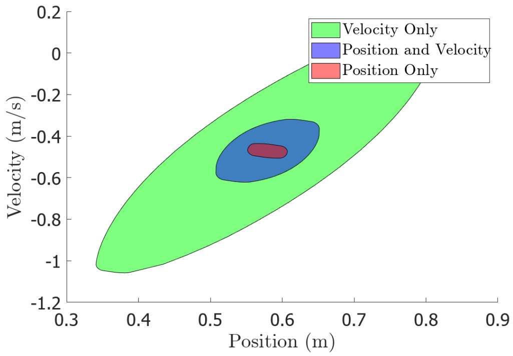

Fig. 2. Attack reachable sets for the spring-mass-damper system under stealthy attacks when measuring both states, C, only position, $C_1 = [1\ 0]$, or only velocity, $C_2 = [0\ 1]$, compared at time $t = 1$ s and initial conditions of $x_0 = \hat{x}_0 = [1\ 0]^T$.

Note that the presence of system noise makes it unlikely for the trajectory to reach the boundary exactly since that would require worst-case realization of system noise. As a result, the deviated estimation error (Fig. 1, left column) will corrupt the estimated state ($\hat{x}_k$) and the controller will act to respond to this perceived deviation; however, this instead causes the actual state (x_k) trajectory to be pushed even further from the nominal trajectory.

1) Stealthy Attack Reachability: Fig. 1 depicts projections of the reachable sets into the position and velocity states, however, the reachable set itself contains significantly more information, most notably the possible correlation/dependence between the two states. A visualization of the state reachability sets under stealthy attacks, $\mathcal{X}_k$ (22), is visualized in Fig. 2. Specifically, this depicts a comparison for the cases when the estimator is potentially limited to observing a single state. When measuring only position, $C_1 = [1\ 0]$, or only velocity, $C_2 = [0\ 1]$, the measurement noise is instead bounded by interval zonotopes $H_1 = C_1 H$ and $H_2 = C_2 H$, respectively. In each case, the observer and controller gains are selected to ensure $\text{eig}(A - LC) = \text{eig}(A + BK) = \{0.9, 0.95\}$. These three cases are visualized at $t = 1$ s and initial conditions of $x_0 = \hat{x}_k = [1\ 0]^T$.

When only measuring the velocity (green) $\mathcal{X}_k$ is largest due to the larger measurement noise in the velocity sensor. Conversely, only measuring the position directly (red) has less impact on the estimate uncertainty and results in being the most robust to stealthy attacks. When both sensors are used (blue), the size $\mathcal{X}_k$ is in between those dependent on individual sensors. This demonstrates that the addition of a more sensors may actually result in a less secure system as the additional measurement signals, and associated noise, can be exploited by an attacker. The set-based detection and attack impact framework allows us to quantify the difference and would be the starting point towards system design to make the system less sensitive to attacks.

B. Vehicle Platoon

The following example considers a one-dimensional model of n identical vehicles in platoon formation. This platoon model is marginally stable and is used to demonstrate how the tools presented in this paper can provide finite time guarantees on the security of systems. In particular, the ability for stealthy attacks to cause a crash the vehicles within a fixed time window is tested.

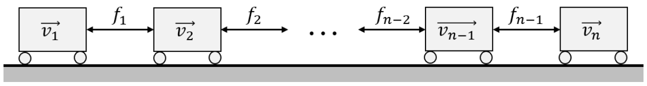

Fig. 3. Vehicle platoon model of n cars where the direction of travel is to the right.

Fig. 3 shows the schematic of a platoon of n identical vehicles traveling in the same direction with absolute positions x_i, absolute velocities v_i, and length l_c, for $i = 1, \ldots, n$. Each vehicle (i) is able to observe a noisy measurement of the positions and velocities of the vehicle immediately in front $(i+1)$ and behind $(i-1)$ it. Vehicle n leads the platoon with a constant velocity, v_n, while each subsequent car aims to maintain a velocity dependent distance between vehicles, $d_i = r + hv_{i+1}$, where d_i is the distance between vehicle i and $i + 1$, r is the desired distance at rest, and h is the time headway. Each vehicle implements an identical proportional-derivative controller (with gains k_p and k_d) based on the estimated state to track the desired distance with respect to the vehicle in front of it. Previous implementations of similar platoon models have used direct output feedback rather than the estimate feedback as consider here [16], [17].

To better study the stability and security, the platoon dynamics are often modeled in terms of the desired distance error $z_i = x_{i+1} - x_i - l_c - d_i$ and relative velocities $w_i = v_{i+1} - v_i$, resulting in the closed-loop system dynamics

$$\begin{aligned} \dot{z}_i &= hk_p(\hat{z}_i - \hat{z}_{i+1}) + hk_d(\hat{w}_i - \hat{w}_{i+1}) + w_i \\ \dot{w}_i &= k_p(\hat{z}_{i-1} - 2\hat{z}_i + \hat{z}_{i+1}) + k_d(\hat{w}_{i-1} - 2\hat{w}_i + \hat{w}_{i+1}) \\ \dot{z}_{n-1} &= w_{n-1} \\ \dot{w}_{n-1} &= k_p(\hat{z}_{n-1} - \hat{z}_{n-1}) + k_d(\hat{w}_{n-2} - \hat{w}_{n-1}) \end{aligned} \tag{28}$$

which can then be formulated in a state-space format with state $\mathbf{x} = [z_1\ \ldots\ z_{n-1}\ w_1\ \ldots w_{n-1}]^T$. These relative states are related to the actual following distances and absolute velocities by, respectively,

$$\begin{aligned} f_i &= z_i + r + hv_{i+1}, \\ v_i &= v_n - (w_i + \cdots + w_{n-1}). \end{aligned} \tag{29}$$

This example considers a platoon of four cars ($n = 4$) with modeling parameters $l_c = 4.5$ m, $v_4 = 30$ m/s, $r = 1$ m, $h = 0.33$ s, $k_p = 1$, and $k_d = 5$, leading to the following continuous time LTI dynamics matrices,

$$\begin{aligned} A_c &= \begin{bmatrix} \mathbf{0} & \mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad C = \mathbf{I_6}, \quad K_c = \begin{bmatrix} k_p\mathbf{I_3} & \mathbf{0} \\ \mathbf{0} & k_d\mathbf{I_3} \end{bmatrix}, \\ B_c &= \begin{bmatrix} h & -h & 0 & h & -h & 0 \\ 0 & h & -h & 0 & h & -h \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -2 & 1 & 0 & -2 & 1 & 0 \\ 1 & -2 & 1 & 1 & -2 & 1 \\ 0 & 1 & -1 & 0 & 1 & -1 \end{bmatrix}. \end{aligned} \tag{30}$$

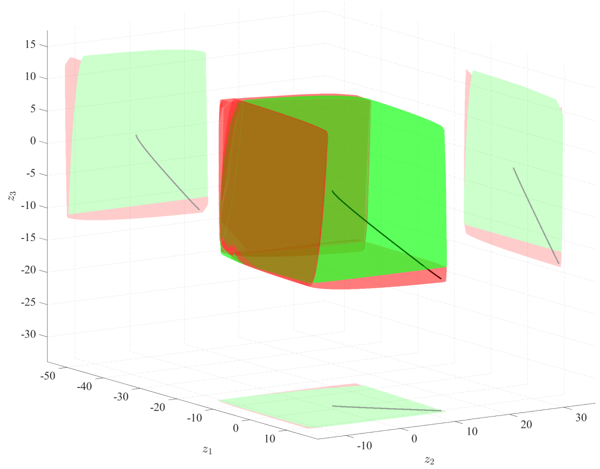

Fig. 4. The reachable set of relative following distances within the state $\mathcal{X}_k$ (green), when stealthily attacked, observed at time $t = 5.95$ s projected onto the respective two dimensional planes. The regions where collisions will occur ($f_i \leq 0$) are represented in red on these projections. A single trajectory corresponding to a single choice of stealthy attack on the system is denoted by the solid black curve that drives the system to cross $f_3 = 0$ plane, causing vehicles 3 and 4 to collide.

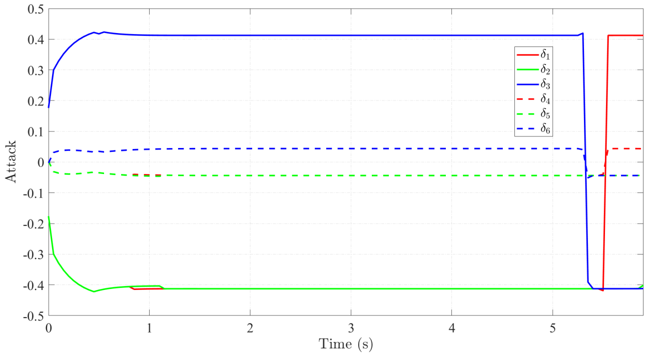

Fig. 5. An example of a sequence of stealthy attacks that drive the platoon to crash two of the vehicles. The corresponding trajectory is plotted in Fig. 4.

The continuous time system is discretized with a zero order hold at $\Delta t = 0.05\,\mathrm{s}$ to obtain the DT-LTI dynamics considered in this paper. The gain matrix L of the discrete time observer is determined by pole placement such that all the eigenvalues $\mathrm{eig}_i(A - LC) = 0.8$. The vehicles are operating at highway speeds in steady-state with $\mathbf{x}_0 = \mathbf{0}_6$.

The measurement noise for each vehicle adds the interval of uncertainty $[-0.22, 0.22]$ m and $[-0.01, 0.01]$ m/s to the relative position z_i and relative velocity w_i measurements, respectively. This is modeled as an unconstrained zonotope $H = \left\{\left[\begin{smallmatrix}\sigma_z \mathbf{I}_3 & \mathbf{0}_3\\ \mathbf{0}_3 & \sigma_2 \mathbf{I}_3\end{smallmatrix}\right], \mathbf{0}\right\}$, where $\sigma_z = 0.22$ and $\sigma_w = 0.01$. No system noise is included in this example.

The framework outlined in this paper is then used to propagate the inherent uncertainty due to measurement noise under normal operation using (7) and use these sets to define the stealthy attack sets as in (12). The reachable set of states due to the stealthy attacks is then computed using (17) and (22).

A crash occurs between vehicles if the following distance reaches zero, i.e., a crash between vehicle i and $i+1$ occurs if/when $f_i \leq 0$. Thus, from (29), at least one pair of vehicles collide if any of the following inequalities hold,

$$\begin{aligned}\begin{bmatrix}1 & 0 & 0 & 0 & -h & -h\end{bmatrix}\mathbf{x} &\leq -r - hv_4,\\ \begin{bmatrix}0 & 1 & 0 & 0 & 0 & -h\end{bmatrix}\mathbf{x} &\leq -r - hv_4,\\ \begin{bmatrix}0 & 0 & 1 & 0 & 0 & 0\end{bmatrix}\mathbf{x} &\leq -r - hv_4.\end{aligned} \tag{31}$$

Since A is marginally stable, by Remark 2, the effect of stealthy attacks can be unbounded. In this context, the set-based framework can be used to find the smallest time horizon where the reachable set intersects one of the critical crash planes, $f_i \leq 0$, which occurs at $k = 102$ (5.1 seconds). The intersection of $\mathcal{X}_k$ and the half-space $f_i \leq 0$ is computed with set operations [18] and checking the emptiness of this intersection is formulated as a linear program [6]. Note that in the absence of set-based methods, an upper bound on this horizon could be computed by, e.g., dynamic programming; however, it would be necessary to guess the terminal state that would lead to a crash. Doing this exhaustively to prove safety would be intractable.

In Fig. 4, the reachable set $\mathcal{X}_k$ (projection onto the three distance errors) is shown for a horizon of $k = 120$ (5.95 seconds). The reachable set intersects with all of the critical crash regions, $f_1 \leq 0$, $f_2 \leq 0$, $f_3 \leq 0$, indicating that it is possible to crash all pairs of vehicles. The trajectory of the platoon corresponding to the stealthy attack in Fig. 5 is show in Fig. 4 as a black curve, demonstrating one of the many stealthy attacks that lead to crashes between vehicles.

VII. Conclusion

This paper develops a set-based framework to quantify and propagate nominal uncertainty through the system and error dynamics to define and calibrate a set-based detector. The detector definition permits the characterization of all possible attacks that are stealthy and evade detection. These stealthy attack sets are then used to propagate their potential impact on the system to find attack-induced state reachable sets. A distinctive aspect of this approach is that this analysis can be accomplished offline and thus represents fundamental security and safety assessments of the system dynamics and parameters, which is not based on run-time or sample-based information. Although the framework is agnostic to the set representation used, most other set representations would require over-approximations to accomplish this analysis, therefore this approach is largely enabled by the capabilities of constrained zonotopes. Access to the attack-induced state reachable sets is the starting point to design/re-design systems for increased security and for evaluating safety by observing the intersection of these sets with dangerous system states; which remains a focus for our ongoing work. We also aim to apply this set-based detector method to both nonlinear and time-varying systems.

References

[1] V. Renganathan, N. Hashemi, J. Ruths, and T. H. Summers, "Higher-order moment-based anomaly detection," *IEEE Control Systems Letters*, vol. 6, pp. 211–216, 2022.

[2] N. Hashemi and J. Ruths, “Generalized chi-squared detector for lti systems with non-gaussian noise,” in *2019 American Control Conference (ACC)*. IEEE, 2019, pp. 404–410.

[3] D. Li and S. Martínez, “High-confidence attack detection via wasserstein-metric computations,” *IEEE Control Systems Letters*, vol. 5, no. 2, pp. 379–384, 2020.

[4] A. A. Kurzhanskiy and Varaiya, *Ellipsoidal calculus for estimation and control*. Nelson Thornes, 1997.

[5] N. Hashemi, C. Murguia, and J. Ruths, “A comparison of stealthy sensor attacks on control systems,” in *2018 Annual American Control Conference (ACC)*, 2018, pp. 973–979.

[6] 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*, vol. 69, pp. 126–136, 2016.

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

[8] Y. Wang, Z. Wang, V. Puig, and G. Cembrano, “Zonotopic set-membership state estimation for discrete-time descriptor lpv systems,” *IEEE Transactions on Automatic Control*, vol. 64, no. 5, pp. 2092–2099, 2018.

[9] B. S. Rego, G. V. Raffo, J. K. Scott, and D. M. Raimondo, “Guaranteed methods based on constrained zonotopes for set-valued state estimation of nonlinear discrete-time systems,” *Automatica*, vol. 111, p. 108614, 2020.

[10] 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)*. IEEE, 2022, pp. 875–881.

[11] J. Li, Z. Wang, Y. Shen, and L. Xie, “Attack detection for cyber-physical systems: A zonotopic approach,” *IEEE Transactions on Automatic Control*, pp. 1–8, 2023.

[12] C. Murguia and J. Ruths, “On model-based detectors for linear time-invariant stochastic systems under sensor attacks,” *IET Control Theory & Applications*, vol. 13, no. 8, pp. 1051–1061, 2019.

[13] “Conzono,” https://github.com/ESCL-at-UTD/ConZono, 2023.

[14] J. Löfberg, “Yalmip : A toolbox for modeling and optimization in matlab,” in *In Proceedings of the CACSD Conference*, Taipei, Taiwan.

[15] Gurobi Optimization, LLC, “Gurobi Optimizer Reference Manual,” 2023. [Online]. Available: https://www.gurobi.com

[16] S. Dadras, R. M. Gerdes, and R. Sharma, “Vehicular platooning in an adversarial environment,” in *Proceedings of the 10th ACM Symposium on Information, Computer and Communications Security*, 2015, pp. 167–178.

[17] J. Giraldo, S. H. Kafash, J. Ruths, and A. A. Cardenas, “Daria: Designing actuators to resist arbitrary attacks against cyber-physical systems,” in *2020 IEEE European Symposium on Security and Privacy (EuroS&P)*, 2020, pp. 339–353.

[18] V. Raghuraman and J. P. Koeln, “Set operations and order reductions for constrained zonotopes,” *Automatica*, vol. 139, p. 110204, 2022.