

Constraining Attacker Capabilities Through Actuator Saturation

Sahand Hadizadeh Kafash¹, Jairo Giraldo², Carlos Murguia³, Alvaro A. Cardenas², and Justin Ruths¹

Abstract—For LTI control systems, we provide mathematical tools – in terms of Linear Matrix Inequalities – for computing outer ellipsoidal bounds on the reachable sets that attacks can induce in the system when they are subject to the physical limits of the actuators. Next, for a given set of dangerous states, states that (if reached) compromise the integrity or safe operation of the system, we provide tools for designing new artificial limits on the actuators (smaller than their physical bounds) such that the new ellipsoidal bounds (and thus the new reachable sets) are as large as possible (in terms of volume) while guaranteeing that the dangerous states are not reachable. This guarantees that the new bounds cut as little as possible from the original reachable set to minimize the loss of system performance. Computer simulations using a platoon of vehicles are presented to illustrate the performance of our tools.

I. INTRODUCTION

Security and privacy in cyber-physical systems (CPS) have become a major concern in the control community due to tight interaction between communication networks and physical processes [1][2]. Several high-profile incidents such as StuxNet [3], the 2016 breach of Ukrainian power grid [4], as well as attacks on modern cars [5] have exposed a wide range of vulnerabilities in CPS. As a consequence, the use of control techniques to analyze the impact of cyber-attacks and to design anomaly detection and mitigation tools have drawn significant attention in recent years [6]–[15].

Most of the work on security analysis does not take into account the physical constraints of actuators; however, it is well known that constrained control actions have significant implications in stability and reachability of control systems [16]. Since actuators cannot inject arbitrarily large amounts of energy into the system, there are always physical limitations restricting the trajectories that actuators can induce. In most physical dynamical systems actuator saturation arises from these physical limits (e.g., the power that can be injected to an electrical system; the acceleration possible by an engine due to limited torque; the maximum flow rate of an inlet pipe). From a control designer perspective, this translates into a reachability problem: whether it is possible

to drive the system from an initial state to a final state given the actuator bounds.

On the other hand, from an attacker's viewpoint and since attacks on sensors or in control commands lead to anomalous actuators actions, actuator bounds will tend to reduce the adversary capabilities in terms of the states that can be reached by the attacker's action. So the question arises: given the actuator bounds, is it possible for the attacker to drive the system to an undesired or dangerous state? That is, given a set of unsafe states (i.e., the states where the integrity or safe operation of the system is compromised) \mathcal{D} , is there a sequence of attacker actions that is capable of driving the system state to \mathcal{D} given the physical restrictions on the actuators? This question reduces to identifying the intersection between the attack-induced reachable set of states and the dangerous state set \mathcal{D} .

The contributions of this work are twofold. First, we approximate the reachable set induced by individually-bounded control inputs. Because quantifying this exact set is mathematically intractable, we turn to construct outer ellipsoidal bounds of the reachable sets. We provide tools in terms of Linear Matrix Inequalities (LMIs) to obtain tight ellipsoidal bounds for the inherent actuator bounds, using an approach similar to that in [17]. Second, we then formulate a design strategy that determines artificial limits on actuators smaller than their physical bounds to reduce the potential impact of attackers on the behavior of the system. To avoid the trivial (and useless) solution of setting the artificial bounds to zero and to minimize the loss of system performance, we find new input bounds that make the new reachable set as large as possible without overlapping a given set of dangerous states \mathcal{D} . Effectively our goal is to maximize, through the choice of new actuator bounds, the size of the reachable set without intersecting with unsafe states. In lieu of maximizing the actual reachable set, which is generally intractable, we maximize the volume of the ellipsoidal bounds, phrasing this as a synthesis LMI. Intuitively, the ellipsoidal bound is maximized until it touches but does not cross the boundary of the dangerous states. Finally, we show the viability of our approach in a vehicle platooning example subject to false-data injection attacks.

II. BACKGROUND

We study Linear Time-Invariant (LTI) discrete-time systems with individually-bounded control inputs:

$$x_{k+1} = Fx_k + Gu_k, \quad (1)$$

with $k \in \mathbb{N}$; state $x_k \in \mathbb{R}^n$; state matrix $F \in \mathbb{R}^{n \times n}$; input matrix $G \in \mathbb{R}^{n \times m}$; and symmetrically bounded control input

*This work was partially supported by the National Research Foundation (NRF), Prime Minister's Office, Singapore, under its National Cybersecurity R&D Programme (Award No. NRF2014NCR-NCR001-40) and administered by the National Cybersecurity R&D Directorate. Thiby the Air Force Office of Scientific Research under award number FA9550-17-1-0135 and by NSF under award number CNS 1553683.

¹These authors are with the Departments of Mechanical and Systems Engineering at the University of Texas at Dallas, Richardson, Texas, USA Sahand.HadizadehKafash, jruths@utdallas.edu

²These authors are with the Computer Science Department at the University of Texas at Dallas, Richardson, Texas, USA jag140730, Alvaro.Cardenas@utdallas.edu

³This author is with the iTrust Center at Singapore University of Technology and Design, Singapore murguia.rendon@sutd.edu.sg

$u_k \in \mathbb{R}^m$ such that

$$[u_k]_i^2 \leq \gamma_i, \quad i = 1, \dots, m, \quad (2)$$

where $\gamma_i > 0$ is a constant bound for the magnitude of each control input, i.e., $[\xi]_j$ is the j^{th} element of ξ .

Physical laws and energy constraints lead real control systems to have practical limits on the actuators used to steer the system dynamics. For example, a vehicle cannot accelerate or decelerate infinitely fast; the engine and brakes have limits. These limits imply a saturation in the mapping from input signal to actuation (e.g., once the engine limit is reached, increasing the throttle will not add more torque). These are the input signal limitations modeled by γ_i .

In this work, we study a generic model that captures any attack that, directly or indirectly, propagates itself to the input signal regardless of the specific mechanism (e.g., feedback law, etc). In particular, we characterize the set of possible reachable states induced by individually-bounded control inputs. We then consider imposing artificial bounds to constrain this reachable set to avoid states that are harmful to the system or unsafe. Because quantifying the exact reachable set is not tractable, we turn to construct outer ellipsoidal bounds on these reachable sets.

In a similar context, the authors in [18] developed tools to quantify outer ellipsoidal bounds on the reachable set of states for LTI systems with peak bounded input, where the norm of input vector is bounded in aggregate, i.e., $\|u_k\| < \gamma$ (in this paper $\|\cdot\|$ denotes Euclidian norm). We restate their result before we build on it for this work.

Lemma 1: [18]. Let V_k be a positive definite function, $V_1 = 0$, and $\zeta_k^T \zeta_k \leq \kappa \in \mathbb{R}_{>0}$. If there exists a constant $a \in (0, 1)$ such that the following holds, then $V_k \leq 1$:

$$V_{k+1} - aV_k - \frac{1-a}{\kappa} \zeta_k^T \zeta_k \leq 0. \quad (3)$$

III. RESULTS

From a theoretical viewpoint, if (F, G) is controllable, the reachable set of (1) is the complete state space \mathbb{R}^n – even if arbitrarily large inputs are required to reach some states. However, actuators have practical limitations and as a result the entire state-space is not reachable.

Our goal is to find an ellipsoids which encapsulates the entire reachable set. We define these ellipsoids as:

$$\mathcal{E}(P, \alpha) := \{x \in \mathbb{R}^n \mid x^T P x \leq \alpha\}, \quad (4)$$

where $P \in \mathbb{R}^{n \times n}$ is a positive-definite matrix and $\alpha \in \mathbb{R}_{>0}$ is a positive constant. When $\alpha = 1$, we omit writing it out explicitly, i.e., $\mathcal{E}(P, 1) = \mathcal{E}(P)$. We adapt the result in [18] to find these ellipsoids (proof is similar, omitted for brevity).

Lemma 2: Let V_k be a positive definite function, $V_1 = 0$, and $[u_k]_i^2 \leq \gamma_i$, $i = 1, \dots, m$. If there exists a constant $a \in (0, 1)$ such that the following holds, then $V_k \leq m$:

$$V_{k+1} - aV_k - (1-a) \sum_{i=1}^m \frac{[u_k]_i^2}{\gamma_i} \leq 0. \quad (5)$$

A. Analysis

We can now employ Lemma 2 to derive an outer ellipsoidal bound on the reachable set of (1) subject to known individual actuator bounds (2). The reachable set we seek to quantify is given by

$$\mathcal{R} := \left\{ x_k \in \mathbb{R}^n \mid \begin{array}{l} x_{k+1} = Fx_k + Gu_k, \quad x_1 = \mathbf{0}, \\ [u_k]_i^2 \leq \gamma_i, \quad i = 1, \dots, m, \quad \forall k \in \mathbb{N} \end{array} \right\}. \quad (6)$$

Notice that

$$\sum_{i=1}^m \frac{[u_k]_i^2}{\gamma_i} = u_k^T R u_k \leq m, \quad R := \begin{bmatrix} \frac{1}{\gamma_1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \frac{1}{\gamma_m} \end{bmatrix}, \quad (7)$$

where the actuator bounds are collected in the matrix R . Note that \mathcal{R} is unbounded if the system is open-loop unstable.

Theorem 1: Consider the LTI system (1) with matrices (F, G) , the actuator bounds $\gamma_i > 0$, $i = 1, \dots, m$, and matrix R in (7). For given $a \in (0, 1)$, if there exists a positive definite matrix $P \in \mathbb{R}^{n \times n}$ solution of the following convex optimization:

$$\begin{cases} \min_P & -\log \det P, \\ \text{s.t.} & P > 0, \text{ and} \\ & \begin{bmatrix} aP - F^T P F & -F^T P G \\ -G^T P F & (1-a)R - G^T P G \end{bmatrix} \geq 0, \end{cases} \quad (8)$$

then, $\mathcal{R} \subseteq \mathcal{E}(P, m)$ and the ellipsoid $\mathcal{E}(P, m)$ has minimum volume.

Proof: For some positive definite matrix $P \in \mathbb{R}^{n \times n}$, let $V_k = x_k^T P x_k$ in Lemma 2. Substituting (1) and this V_k in (5) yields

$$\nu^T \underbrace{\begin{bmatrix} aP - F^T P F & -F^T P G \\ -G^T P F & (1-a)R - G^T P G \end{bmatrix}}_Q \nu \geq 0 \quad (9)$$

where $\nu = [x_k^T, u_k^T]^T$. This inequality is satisfied if and only if Q is positive semi-definite.

To ensure that the ellipsoid bound is as tight as possible, we minimize $(\det P)^{-1/2}$ since this quantity is proportional to the volume of $x_k^T P x_k = m$. We instead minimize $\log \det P^{-1}$ as it shares the same minimizer and because for $P > 0$ this objective is convex [19]. ■

Lemma 2 indicates that the solution to the optimization problem (8) may exist for some values of the parameter $a \in (0, 1)$. Out of these values, we are interested in selecting the one that leads to the ellipsoid with minimum volume. We employ a straightforward grid search to find this value of a .

Remark 1: In the case in which all control inputs have identical bounds, i.e., $R = \frac{1}{\gamma} I_m$, with I_m the $m \times m$ identity matrix, the common scalar bound γ can be factored out of the LMI in (9) by defining $\hat{P} = \gamma P$, then:

$$Q = \frac{1}{\gamma} \underbrace{\begin{bmatrix} a\hat{P} - F^T \hat{P} F & -F^T \hat{P} G \\ -G^T \hat{P} F & (1-a)I_m - G^T \hat{P} G \end{bmatrix}}_{\hat{Q}} \geq 0. \quad (10)$$

Since $\gamma > 0$, $\hat{Q} \geq 0 \Leftrightarrow Q \geq 0$. This implies that in the case of common actuator bounds, the optimization can be solved independent of the actual bound. After a solution \hat{P} satisfying $\hat{Q} \geq 0$ in (10) is found, it is simply scaled by the bound γ to recover the desired ellipse, i.e., $\mathcal{R} \subseteq \mathcal{E}\left(\frac{1}{\gamma}\hat{P}, m\right)$.

B. Synthesis

In most physical dynamical systems, actuator saturation arises from physical limits. In the past section, Theorem 1 gives us the tools necessary to quantify the outer ellipsoidal bounds for the reachable states according to such inherent actuator bounds. In the context of security, it is intriguing to impose artificial limits on actuators smaller than their physical bounds to reduce the potential impact of attackers on the behavior of the system. Such a design problem would be informed by a region of state space \mathcal{D} which is considered unsafe. Such a region might represent *dangerous states* in which, for example, the pressure of a holding vessel will exceed its pressure rating or the level of a liquid in a tank exceeds its capacity.

Thus, we aim here to design new bounds $\hat{\gamma}_i$, $i = 1, \dots, m$ such that the new reachable set bounding ellipsoid does not overlap with the unsafe states. Corresponding to these new bounds, we define the new reachable set $\hat{\mathcal{R}}$ as:

$$\hat{\mathcal{R}} := \left\{ x_k \in \mathbb{R}^n \mid \begin{array}{l} x_{k+1} = Fx_k + Gu_k, \quad x_1 = \mathbf{0}, \\ [u_k]_i^2 \leq \hat{\gamma}_i, \quad i = 1, \dots, m, \quad \forall k \in \mathbb{N} \end{array} \right\}. \quad (11)$$

The dangerous state sets in many, if not most, practical applications can be captured through the union of half-spaces defined by their boundary hyperplanes:

$$\mathcal{D} := \left\{ x \in \mathbb{R}^n \mid \bigcup_{i=1}^{\kappa} c_i^T x \geq b_i \right\}, \quad (12)$$

where each pair (c_i, b_i) , $c_i \in \mathbb{R}^n$, $b_i \in \mathbb{R}$, $i = 1, \dots, \kappa$ quantifies a hyperplane that defines a single half-space.

Theorem 2: Consider the LTI system (1) with matrices (F, G) , the original actuator bounds $\gamma_i > 0$, $i = 1, \dots, m$, matrix R in (7), and a set \mathcal{D} of dangerous states bounded by the hyperplanes $c_i^T x = b_i$, $i = 1, \dots, \kappa$. For given $a \in (0, 1)$, if there exist a positive definite matrix $Y \in \mathbb{R}^{n \times n}$ and diagonal matrix $\hat{R} := \text{diag}(\hat{r}_1, \dots, \hat{r}_m) \in \mathbb{R}^{m \times m}$, $\hat{r}_i > 0$, solution of the following convex optimization:

$$\begin{cases} \min_{\hat{R}, Y} \text{tr}(\hat{R}), \\ \text{s.t. } \hat{R} \geq R, \quad Y > 0, \\ c_i^T Y c_i \leq \frac{b_i^2}{m}, \quad \text{for } i = 1, \dots, \kappa, \\ \begin{bmatrix} aY & 0 & YF^T \\ 0 & (1-a)\hat{R} & G^T \\ FY & G & Y \end{bmatrix} \geq 0, \end{cases} \quad (13)$$

then, the new actuator bounds $\hat{\gamma}_i := (1/\hat{r}_i)$, $i = 1, \dots, m$, enforce that \mathcal{D} does not intersect with the new reachable set $\hat{\mathcal{R}}$ in (11) and maximize the volume of the new minimum-volume ellipsoid $\mathcal{E}(Y^{-1}, m)$ bounding $\hat{\mathcal{R}}$.

Proof: The minimum distance d between an ellipsoid $\mathcal{E}(P, m)$ centered at zero and a hyperplane $c^T x = b$ is given by the formula [20]

$$d = \frac{|b| - \sqrt{mc^T P^{-1} c}}{\sqrt{c^T c}}. \quad (14)$$

We aim to obtain the largest ellipsoid (in terms of volume) that does not cross the hyperplane. This would maximize the size of the reachable set restricted to not crossing into the dangerous set. This is accomplished if we let the ellipsoid and hyperplane to touch at a single point, i.e., $d = 0$,

$$c^T P^{-1} c = \frac{b^2}{m}. \quad (15)$$

Given the hyperplane parameters c and b , the choice of P^{-1} that satisfies this relationship is then the largest ellipse $\mathcal{E}(P, m)$ that does not overlap with the dangerous states. Note that the zero-distance condition (15) is written in terms of P^{-1} which is not linear in P . Thus, to maintain a tractable convex semi-definite optimization problem, we write the original analysis LMI Q in (9) in terms of $Y := P^{-1}$ and the new matrix \hat{R} . The new Q can be written as the Schur complement of a higher dimensional matrix \tilde{Q} such that the positive semi-definiteness of \tilde{Q} implies the positive semi-definiteness of Q ,

$$Q \geq 0 \Leftrightarrow \tilde{Q} = \begin{bmatrix} aY^{-1} & 0 & F^T Y^{-1} \\ 0 & (1-a)\hat{R} & G^T Y^{-1} \\ Y^{-1}F & Y^{-1}G & Y^{-1} \end{bmatrix} \geq 0. \quad (16)$$

Finally, applying the congruence transformation $\mathcal{Y} = \text{diag}(Y, I_n, Y)$ to \tilde{Q} results in (13).

The constraint (15) can be added for each hyperplane that specifies the boundary of the dangerous set. Once multiple hyperplanes specify the boundary, it is possible that the ellipse cannot touch all boundaries simultaneously; therefore, we relax each distance constraint into the inequality in (13) ($c^T P^{-1} c = c^T Y c \leq \frac{b^2}{m}$). Taken together, these distance constraints ensure that the outer ellipsoidal bound of reachable states does not extend beyond any of the hyperplanes. In order to guarantee that we find the largest possible bounds, we maximize a function of the bounds, or in this case minimize a function of the diagonal elements of \hat{R} (the trace of \hat{R}). This bound maximization coupled with the hyperplane boundaries have the additional effect of ensuring the tightness ellipsoidal bound (minimum volume) to the reachable set. This allows us to omit the $\log \det P^{-1}$ objective we use in Theorem 1 (which is not a convex objective with the decision variable as $Y = P^{-1}$). Maximizing the bounds first expands the ellipse until it hits one or more hyperplanes. Then, continuing to maximize the bounds maintains the same ellipsoid, however, increases the tightness of the ellipsoid onto the reachable states, until this cannot be improved. The restriction $\hat{R} \geq R$ is added to ensure that the new bounds are not larger than the original physical bounds. ■

A particular case in which a (semi) closed-form expression for the new actuator bounds can be obtained is when all the bounds are designed to be equal, i.e., $\hat{\gamma}_i := \hat{\gamma} \in \mathbb{R}_{>0}$,

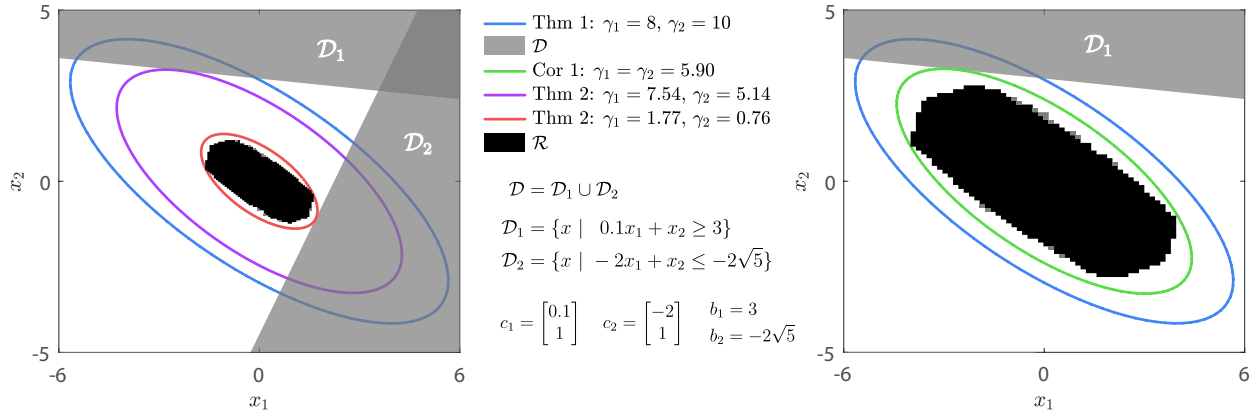


Fig. 1. The solution of Theorem 1 provides the (blue) ellipsoid which outer-bounds the reachable set corresponding to the system driven by the original physical input bounds $\gamma = [8, 10]$. A dangerous state set \mathcal{D} , is defined by the single half-space $c_1^T x \geq b_1$ and the solution of Theorem 2 provides the bounds $\gamma = [7.54, 5.14]$ and corresponding (purple) ellipsoid that outer-bounds the largest reachable set that avoids the dangerous state set. A second half-space is added to \mathcal{D} , $c_2^T x \leq b_2$ and Theorem 2 again provides the bounds $\gamma = [1.77, 0.76]$ and the (red) ellipsoid that outer-bounds the largest reachable set that avoids both half-spaces. Corollary 1 can be used to derive equal bounds $\gamma = [5.9, 5.9]$ and the corresponding (green) ellipsoid that bound the reachable set that avoids the single half-space $c_1^T x \geq b_1$. We show a Monte-Carlo simulation of many (10,000) trajectories to construct an empirical reachable set according to the corresponding bounds, which demonstrates the tightness of the ellipsoidal bounds.

$i = 1, \dots, m$. This is stated in the following corollary of Theorem 2.

Corollary 1: Let $\gamma_{\min} := \min(\gamma_i)$, $i = 1, \dots, m$. For given $a \in (0, 1)$, if there exists a positive definite matrix $\hat{P} \in \mathbb{R}^{n \times n}$ solution of the following convex optimization:

$$\begin{cases} \min_{\hat{P}} -\log \det \hat{P}, \\ \text{s.t. } \hat{P} > 0, \text{ and} \\ \begin{bmatrix} a\hat{P} - F^T \hat{P} F & -F^T \hat{P} G \\ -G^T \hat{P} F & (1-a)I_m - G^T \hat{P} G \end{bmatrix} \geq 0, \end{cases} \quad (17)$$

then, the new actuator bound:

$$\hat{\gamma} := \min \left(\frac{b_i^2}{m c_i^T \hat{P}^{-1} c_i}, \gamma_{\min} \right), \quad i = 1, \dots, m, \quad (18)$$

enforces that \mathcal{D} does not intersect with the new reachable set $\hat{\mathcal{R}}$ in (11) with $\hat{\gamma}_i = \hat{\gamma}$ and maximize the volume of the new minimum-volume ellipsoid $\mathcal{E}(\frac{1}{\hat{\gamma}} \hat{P}, m)$ bounding $\hat{\mathcal{R}}$.

Proof: Because $\hat{\gamma}_i = \hat{\gamma}$, we have $\hat{R} = \frac{1}{\hat{\gamma}} I_m$ in (13). Applying the same congruence transformation \mathcal{Y}^{-1} , the Schur complement, and the change of variables $P = Y^{-1}$ to the bottom LMI in (13), we get the equivalent inequality

$$\frac{1}{\hat{\gamma}} \begin{bmatrix} a\hat{P} - F^T \hat{P} F & -F^T \hat{P} G \\ -G^T \hat{P} F & (1-a)I_m - G^T \hat{P} G \end{bmatrix} \geq 0,$$

with $\hat{P} = \hat{\gamma} P$. Which is the bottom inequality in (17) scaled by $\frac{1}{\hat{\gamma}}$. This implies that for a solution \hat{P} of (17), we can simply scale \hat{P} by the inversed new bound $\hat{\gamma}$ to recover the desired ellipse, i.e., $\hat{\mathcal{R}} \subseteq \mathcal{E}(\frac{1}{\hat{\gamma}} \hat{P}, m)$. Note that $\mathcal{E}(\frac{1}{\hat{\gamma}} \hat{P}, m)$ is a minimum-volume ellipsoid for any $\hat{\gamma} > 0$ because $\mathcal{E}(\hat{P}, m)$ is minimum-volume and both ellipsoids have the same shape and orientation (i.e., the eigenvectors of P and \hat{P} are the same). Hence, to have a maximal minimum-volume ellipsoid, we have to select $\hat{\gamma}$ to be the largest $\hat{\gamma}$ such that the dangerous

states are avoided. For this, as in Theorem 2, we use the distance condition $c_i^T Y c_i = c_i^T P^{-1} c_i = \hat{\gamma} c_i^T \hat{P}^{-1} c_i \leq \frac{b_i^2}{m}$, $i = 1, \dots, m$. The largest $\hat{\gamma}$ satisfying the latter inequality and being smaller than the smallest original physical bound is given by $\hat{\gamma}$ in (18). ■

IV. ILLUSTRATIVE EXAMPLE

To illustrate these results, we consider a linear 2×2 system with two inputs

$$F = \begin{bmatrix} 0.84 & 0.23 \\ -0.47 & 0.12 \end{bmatrix}, \quad G = \begin{bmatrix} 0.07 & 0.3 \\ 0.23 & 0.1 \end{bmatrix}. \quad (19)$$

The actuators of this system have physical limitations which saturate the inputs at $[u_k]_1^2 \leq \gamma_1 = 8$ and $[u_k]_2^2 \leq \gamma_2 = 10$, $k \in \mathbb{N}$. These inherent bounds impose a reachable set of states which are outer-bounded by an ellipse that is the solution of Theorem 1. This ellipse is the blue ellipse in Fig. 1. This solution (and all others we present here) is found using YALMIP, a MATLAB toolbox for optimization, especially semidefinite programs (SDP), and the SDP solver SeDuMi [21],[22]. As mentioned before, the parameter a is incrementally varied over the range $(0, 1)$ and the optimization is solved at each value of a . The optimal solution must first be a successful solution (for some values of a there is no solution to the problem) and second must yield an ellipse of minimum volume.

We now define a set of dangerous states that represent states which jeopardize the safety and/or operation of the system. Let $\mathcal{D} = \mathcal{D}_1 = \{x \mid 0.1x_1 + x_2 \geq 3\}$. It is immediately apparent that the inherent bounds allow the system to reach some of the dangerous states. Using Theorem 2, we can find new artificial bounds on the inputs so that the system can no longer reach the danger states. We solve the optimization for the purple ellipse in Fig. 1 corresponding to bounds $\gamma_1 = 7.54$ and $\gamma_2 = 5.14$. This outer ellipsoidal bound touches, but does not cross the hyperplane that defines

\mathcal{D}_1 . Thus, following these new bounds, the system cannot reach any of the dangerous states.

It is also possible to avoid the dangerous states while enforcing equivalent artificial bounds on all inputs. Using Corollary 1, we identify the ellipsoid \hat{P} . Using (18), we find the bound $\gamma_1 = \gamma_2 = \gamma = 5.9$ and the corresponding green ellipsoid in Fig. 1. This ellipsoid (as well as the others above) tightly bound the reachable set. We show this by plotting an extensive Monte-Carlo simulation of 10,000 trajectories (each of length 1000 steps) of the system with the bounds $\gamma_1 = \gamma_2 = \gamma = 5.9$. This empirical reachable set is well approximated by the outer bounding ellipsoid.

We further demonstrate that the formulation in Theorem 2 can handle a dangerous state set composed of the union of multiple half-spaces. Let $\mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_2$, where $\mathcal{D}_2 = \{x \mid -2x_1 + x_2 \leq -2\sqrt{5}\}$. Theorem 2 with both half-space constraints produces the bounds $\gamma_1 = 1.77$ and $\gamma_2 = 0.76$ along with the corresponding red ellipse in Fig. 1. Again we demonstrate the tightness of this ellipsoidal bound by plotting the empirical reachable set.

V. CASE STUDY: PLATOONING

In order to illustrate the viability of our analysis, we consider the problem of vehicular platooning. In particular, platooning offers many benefits over solo driving such as better reaction times, decrease of CO₂ emissions, and lower fuel consumption [23]. The objective of the platoon is to maintain an adequate distance between vehicles, such that sudden changes in the leader's speed (e.g., braking) will not cause any crash in the preceding vehicles. This is known as the string stability of the platoon and has been widely studied in the literature [24], [25], [26]. Typically, the Adaptive Cruise Control (ACC) system controls the distance and/or relative velocity between adjoining vehicles by measuring (radar/lidar) and reacting to the relative distance and/or velocity between adjacent vehicles compared to a desired setpoint. More recently, work has leveraged vehicle-to-vehicle or infrastructure-to-vehicle communication to inject feed-forward commands. Such Cooperative Adaptive Cruise Control (CACC) systems improve the string stability of the platoon and allows vehicles to follow each other with a closer distance than with ACC, thereby improving traffic flow capacity. CACC gathers information of vehicles further in front according to a specific communication network topology.

Let us consider a simplified discrete-time cruise control model of a platoon with n vehicles as follows

$$\begin{aligned} d_{i-1,k+1} &= d_{i-1,k} + \Delta_t(v_{i,k} - v_{i-1,k}) \\ v_{j,k+1} &= v_{j,k} + \beta_j v_{j,k} + \Delta_t u_{j,k} \end{aligned}$$

where $i = 1, \dots, n-1$, $j = 1, \dots, n$, $k \in \mathbb{Z}_+$ is the sampling instant, $d_{i,k}$ is the distance between vehicle $i+1$ and i , $v_{j,k}$ is the speed of the j^{th} vehicle, and $u_{j,k}$ is the control input that changes the acceleration. Δ_t is the sampling period and $\beta_j < 0$ is the velocity loss caused by friction. For simplicity, let $d_k = [d_{1,k}, \dots, d_{n-1,k}]$, $v_k = [v_{1,k}, \dots, v_{n,k}]$, and $u_k = [u_{1,k}, \dots, u_{n,k}]$. The leader vehicle is indexed by n .

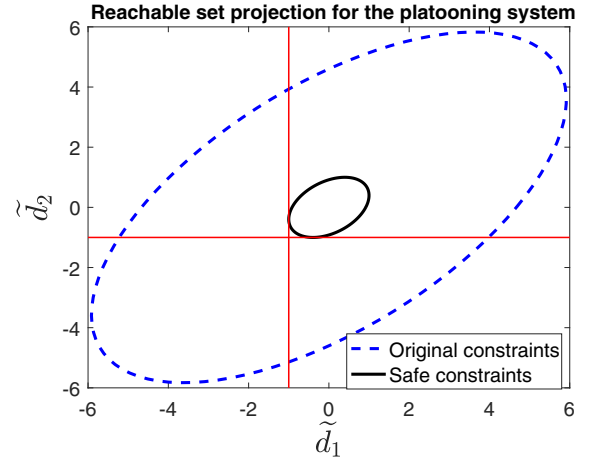


Fig. 2. Projection of the reachable set in the \tilde{d}_1 - \tilde{d}_2 plane for the platooning example with unsafe hyperplanes $\tilde{d}_i \leq -1$ for $i = 1, 2$. Notice that with the original constraints the unsafe states can be reached, but imposing the safe constraints obtained using our results ensures that no disturbance or attack will drive the system to unsafe states.

Here we consider a simple control strategy that combines an ACC and a secondary control (e.g., CACC) $u_{i,k} = \tilde{u}_{i,k} + w_{i,k}(d_k, v_k)$, where $\tilde{u}_{i,k}$ corresponds to a forward-and-reverse-looking proportional-derivative control [27]

$$\begin{aligned} \tilde{u}_{i,k} &= k_{p,i}(-d_{i-1,k} + d_{i-1}^*) + k_{p,i}(d_{i,k} - d_i^*) \\ &\quad + k_{d,i}(v_{i-1,k} - v_{i,k}) + k_{d,i}(v_{i,k} - v_{i+1,k}), \end{aligned} \quad (20)$$

to maintain a desired distance d_i^* between vehicles and $w_{i,k}(d_k, v_k)$ is the secondary control strategy that relies on a communication network. However, an adversary can potentially intercept and modify that communication causing adverse effects such as making the vehicles crash.

Because of the physical constraints in the acceleration/deceleration of each vehicle, we assume that the secondary control action is bounded according to

$$\underline{w}_i \leq w_{i,k} \leq \bar{w}_i. \quad (21)$$

for $\underline{w}_i < \bar{w}_i$. As mentioned above, we can impose virtual constraints in control actions in order to avoid unsafe states.

A. Experiments

Suppose we have a platoon of three vehicles where d_i^* and v^* are the desired separation distance and desired velocity, respectively. We introduce a change of variable $\tilde{d}_{i,k} = d_{i,k} - d_i^*$ and $\tilde{v}_{i,k} = v_{i,k} - v^*$, such that $\tilde{d}_{i,k} = 0$ implies that the desired reference is achieved, i.e., $d_i = d_i^*$. Therefore, let $x_k = [\tilde{d}_{1,k}, \tilde{d}_{1,k}, \tilde{v}_{1,k}, \tilde{v}_{2,k}, \tilde{v}_{3,k}]^T$ such that $G = [0_{2 \times 3}^T, \Delta_t I_3^T]^T$ and

$$F = \begin{bmatrix} 1 & 0 & -\Delta_t & -\Delta_t & 0 \\ 0 & 1 & 0 & 0 & \Delta_t \\ k_{p,1} & 0 & (1 + \beta_1) - k_{d,1} & k_{d,1} & 0 \\ -k_{p,2} & k_{p,2} & k_{d,2} & (1 + \beta_2) - 2k_{d,2} & k_{d,2} \\ 0 & -k_{p,3} & 0 & k_{d,3} & (1 + \beta_3) - k_{d,3} \end{bmatrix}.$$

The dynamic system with the ACC and the secondary control w_k is of the form $x_{k+1} = Fx_k + Gw_k$. Notice that $\tilde{d}_{i,k} = -d_i^*$ corresponds to the case when the distance $d_{i,k} = 0$,

which means that the pair of vehicles i and $i+1$ have crashed. Since we want to avoid crashes, we define the unsafe states

$$\mathcal{D} = \{x \in \mathbb{R}^n \mid -x_1 \geq d_1^* \cup -x_2 \geq d_2^*\}.$$

The simulation parameters are $d_i^* = d^* = 1$ m, $v^* = 60$ km/h, $\Delta_t = 0.5$ s, $\beta_i = -0.1$, $k_{p,i} = 0.2$, $k_{d,i} = 0.3$. The initial acceleration constraints are $-w_1 = \bar{w}_1 = 1.1$ m/s², $-w_2 = \bar{w}_2 = 0.9$ m/s², and $-w_3 = \bar{w}_3 = 1.05$ m/s², such that $\gamma = (1.2, 0.8, 1.1)$ (the square of the absolute bounds). Figure 2 illustrates the projection in the $\tilde{d}_1 - \tilde{d}_2$ space of the minimum volume ellipsoid that approximates the reachable set using Theorem 1 for the given acceleration constraints.

However, notice that unsafe states may be reached and thus there are inputs that can lead the vehicles to crash. Applying Theorem 2, we are able to find the set of constraints that will keep the reachable states outside of the dangerous states, as depicted in Fig. 2. The new set of bounds (safe constraints) is $\gamma = (0.03, 0.05, 0.03)$, such that any sensor or actuator attack that affects the control input will not be able to cause the vehicles to crash.

VI. CONCLUSION

In this paper we have taken a new approach to limiting the capabilities of an attacker by imposing artificial limits on the control inputs that drive the system. Whether caused by manipulation of the control inputs themselves, or an indirect effect of sensor or system manipulations, these actuator bounds restrict the states that can be reached. We derive methods based on convex optimization to quantify the reachable states given known actuator bounds and also methods to design new bounds to avoid the reachable set from entering a set of states determined to be unsafe or dangerous. Through the example of a platoon, we show how dangerous states can be determined (e.g., a crash of adjacent vehicles) and avoided. Also, the platoon example demonstrates that security through actuator bounds might come at the cost of conventional performance metrics such as settling time. In future work, we will analyze the balance between stabilizing the system and securing it with the proposed bounds.

REFERENCES

- [1] J. Giraldo, E. Sarkar, A. A. Cardenas, M. Maniatakos, and M. Kantarcioglu, "Security and privacy in cyber-physical systems: A survey of surveys," *IEEE Design & Test*, vol. 34, no. 4, pp. 7–17, Aug 2017.
- [2] A. Cardenas, S. Amin, B. Sinopoli, A. Giani, A. Perrig, and S. Sastry, "Challenges for securing cyber physical systems," in *Workshop on Future Directions in Cyber-physical Systems Security*, 2009.
- [3] R. Langner, "Stuxnet: Dissecting a cyberwarfare weapon," *IEEE Security & Privacy*, vol. 9, no. 3, pp. 49–51, 2011.
- [4] K. Zetter, "Inside the cunning, unprecedented hack of ukraine's power grid," *Wired Magazine*, 2016.
- [5] S. Checkoway, D. McCoy, B. Kantor, D. Anderson, H. Shacham, S. Savage, K. Koscher, A. Czeskis, F. Roesner, T. Kohno et al., "Comprehensive experimental analyses of automotive attack surfaces," in *USENIX Security Symposium*. San Francisco, 2011.
- [6] F. Pasqualetti, F. Dorfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Transactions on Automatic Control*, vol. 58, pp. 2715–2729, 2013.

- [7] Y. Mo, E. Garone, A. Casavola, and B. Sinopoli, "False data injection attacks against state estimation in wireless sensor networks," in *Decision and Control (CDC), 2010 49th IEEE Conference on*, 2010, pp. 5967–5972.
- [8] C. Kwon, W. Liu, and I. Hwang, "Security analysis for cyber-physical systems against stealthy deception attacks," in *American Control Conference (ACC), 2013*, 2013, pp. 3344–3349.
- [9] F. Miao, Q. Zhu, M. Pajic, and G. J. Pappas, "Coding sensor outputs for injection attacks detection," in *Decision and Control (CDC), 2014 IEEE 53rd Annual Conference on*, 2014, pp. 5776–5781.
- [10] C.-Z. Bai and V. Gupta, "On kalman filtering in the presence of a compromised sensor: Fundamental performance bounds," in *American Control Conference (ACC), 2014*, 2014, pp. 3029–3034.
- [11] C. Z. Bai, F. Pasqualetti, and V. Gupta, "Security in stochastic control systems: Fundamental limitations and performance bounds," in *American Control Conference (ACC), 2015*, 2015, pp. 195–200.
- [12] C. M. Ahmed, C. Murguia, and J. Ruths, "Model-based attack detection scheme for smart water distribution networks," in *Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security*, ser. ASIA CCS '17, 2017, pp. 101–113.
- [13] C. Murguia and J. Ruths, "Cusum and chi-squared attack detection of compromised sensors," in *proceedings of the IEEE Multi-Conference on Systems and Control (MSC)*, 2016.
- [14] E. Rothstein Morris, C. Murguia, and M. Ochoa, "Design-time quantification of integrity in cyber-physical-systems," in *Proceedings of the 2017 ACM SIGSAC Workshop on Programming Languages and Analysis for Security*, (accepted), 2017.
- [15] A. Cárdenas, S. Amin, Z. Lin, Y. Huang, C. Huang, and S. Sastry, "Attacks against process control systems: Risk assessment, detection, and response," in *Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security*, 2011, pp. 355–366.
- [16] S. Tarbouriech, G. Garcia, J. M. G. da Silva Jr, and I. Queinnec, *Stability and stabilization of linear systems with saturating actuators*. Springer Science & Business Media, 2011.
- [17] C. Murguia, N. van de Wouw, and J. Ruths, "Reachable sets of hidden cps sensor attacks: Analysis and synthesis tools," in *proceedings of the IFAC World Congress*, 2016.
- [18] N. D. That, P. T. Nam, and Q. P. Ha, "Reachable set bounding for linear discrete-time systems with delays and bounded disturbances," *Journal of Optimization Theory and Applications*, vol. 157, pp. 96–107, 2013.
- [19] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*, ser. Studies in Applied Mathematics. Philadelphia, PA: SIAM, 1994, vol. 15.
- [20] A. B. Kurzhanski and I. Vaynshteyn, *Ellipsoidal calculus for estimation and control*. Laxenburg, Austria : IIASA ; Boston : Birkhauser Boston, 1997.
- [21] J. Lofberg, "Yalmip : a toolbox for modeling and optimization in matlab," in *Computer Aided Control Systems Design, 2004 IEEE International Symposium on*, 2004, pp. 284–289.
- [22] J. F. Sturm, "Using sedumi 1.02, a matlab toolbox for optimization over symmetric cones," *Optimization Methods and Software*, vol. 11, no. 1-4, pp. 625–653, 1999.
- [23] C. Suthaputthakun, Z. Sun, and M. Dianati, "Applications of vehicular communications for reducing fuel consumption and co 2 emission: The state of the art and research challenges," *IEEE Communications Magazine*, vol. 50, no. 12, 2012.
- [24] D. Swaroop and J. K. Hedrick, "String stability of interconnected systems," *IEEE transactions on automatic control*, vol. 41, no. 3, pp. 349–357, 1996.
- [25] J. Ploeg, D. P. Shukla, N. van de Wouw, and H. Nijmeijer, "Controller synthesis for string stability of vehicle platoons," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 2, pp. 854–865, 2014.
- [26] S. Öncü, J. Ploeg, N. van de Wouw, and H. Nijmeijer, "Cooperative adaptive cruise control: Network-aware analysis of string stability," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 4, pp. 1527–1537, 2014.
- [27] 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*. ACM, 2015, pp. 167–178.