Secure Navigation of Robots in Adversarial Environments

Gianluca Bianchin, Yin-Chen Liu, and Fabio Pasqualetti

***Abstract*—We study the problem of navigating a robot in an adversarial environment, where the objective is to perform localization and trajectory planning despite the malicious and unknown action of an attacker. We consider robots with single integrator dynamics, equipped with a Global Navigation Satellite System (GNSS) sensor and a Radio Signal Strength Indicator (RSSI) sensor that provides relative positioning information with respect to a group of radio stations, each with limited communication range. The attacker can simultaneously spoof the sensor readings and send falsified control inputs to the robot, so as to deviate its trajectory from the nominal path. We demonstrate the existence of attacks that escape detectability, and illustrate a method for their systematic design. Conversely, we show that appropriate control design and waypoints selection allow the trajectory planner to ensure attack detectability or secure navigation. More generally, our results show that trajectory planning in nominal and adversarial settings are substantially different, and that careful trajectory design is required to ensure resilience to attacks.**

***Index Terms*—Cyber-physical systems, autonomous systems, fault detection, security, trajectory planning.**

I. Introduction and Problem Setup

MOBILE robots have been used in a broad range of civilian and military operations thanks to their autonomous capabilities, flexibility, and wide range of engineering applications. Autonomous robots rely on sensors to measure their states and use this information to make decisions and to generate control commands to send to their actuators. Despite the tremendous advances in the development of more reliable sensing and communication devices, sensory data and communication channels can be accidentally and maliciously compromised, thus undermining the effectiveness of autonomous operations in critical and adversarial applications. To the best of our knowledge, tools to study the effects of attacks on the trajectories and to design controls that securely steer the robot to a desired final configurations are still critically lacking.

Manuscript received March 1, 2019; revised May 6, 2019; accepted May 30, 2019. Date of publication June 6, 2019; date of current version June 21, 2019. This work was supported in part by the Army Research Office under Award 71603NSYIP, in part by the National Science Foundation under Award CNS1646641, and in part by the Office of Naval Research under Award N00014-19-1-2264. Recommended by Senior Editor F. Dabbene. *(Corresponding author: Gianluca Bianchin.)*

The authors are with the Department of Mechanical Engineering, University of California at Riverside, Riverside, CA 92507 USA (e-mail: gianluca@engr.ucr.edu; yliu@engr.ucr.edu; fabiopas@engr.ucr.edu).

Digital Object Identifier 10.1109/LCSYS.2019.2921753

This letter focuses on robots with integrator dynamics,

$$\dot{x}_{\mathrm{n}} = u_{\mathrm{n}}, \tag{1}$$

where $x_{\mathrm{n}} : \mathbb{R}_{\geq 0} \to \mathbb{R}^2$ denotes the robot position in a two-dimensional space, and $u_{\mathrm{n}} : \mathbb{R}_{\geq 0} \to \mathbb{R}^2$ denotes the nominal control input that actuates the robot velocity. The input u_{n} is a design parameter that is used to plan the robot trajectory between two desired positions. We assume that u_{n} is piecewise continuous and $\|u_{\mathrm{n}}\| \leq u_{\max}$ at all times, with $u_{\max} \in \mathbb{R}_{>0}$.

We consider robots equipped with two noiseless sensors: a GNSS receiver that provides an absolute measure of the position, and a RSSI sensor that provides a measure of the relative distance between the robot and n_{b} radio stations. Let $b_i \in \mathbb{R}^2$ and $r_i \in \mathbb{R}_{>0}$ denote the position of the i-th station with respect to an absolute reference frame and its coverage range, respectively, with $b_i \neq b_j$ if $i \neq j$. We assume that the robot can measure its distance from the i-th station only when its position is within the communication range defined by r_i. The sensor readings are

$$y_{\mathrm{n}}^{\mathrm{GNSS}} = x_{\mathrm{n}}, \text{ and } y_{\mathrm{n},i}^{\mathrm{RSSI}} = \|x_{\mathrm{n}} - b_i\|^2, \tag{2}$$

where $i \in \Omega$ and

$$\Omega(x_{\mathrm{n}}) = \{i : i \in \{1, \dots, n_{\mathrm{b}}\} \text{ and } \|x_{\mathrm{n}} - b_i\| \leq r_i\}.$$

Although our results can be extended to include different classes of sensors, we focus on GNSS and RSSI sensors because they are available in many practical applications [1].

We assume that the robot operates in an adversarial environment, where adversaries can simultaneously spoof the GNSS readings and override the nominal input u_{n} with a compromised attack input. To distinguish between the nominal measurements and those obtained in the presence of attacks, we denote the dynamics of the robot under attack as

$$\dot{x} = u, \tag{3}$$

where $x : \mathbb{R}_{\geq 0} \to \mathbb{R}^2$ represents the attacked robot position and $u : \mathbb{R}_{\geq 0} \to \mathbb{R}^2$ denotes the attacked control input, which also obeys the bound on maximum velocity $\|u\| \leq u_{\max}$. The sensor readings in the presence of attacks are

$$y^{\mathrm{GNSS}} = x + u^{\mathrm{GNSS}}, \quad \text{and} \quad y_i^{\mathrm{RSSI}} = \|x - b_i\|^2, \tag{4}$$

where $u^{\mathrm{GNSS}} : \mathbb{R}_{\geq 0} \to \mathbb{R}^2$ denotes the GNSS spoofing signal, and $i \in \Omega(x)$. We assume that the RSSI readings are not compromised by the attacker, and that the nominal and attacked initial positions satisfy $x_{\mathrm{n}}(0) = x(0)$.

In this letter, we study the competing objectives of the attacker and of the trajectory planner, summarized as follows:

(i) The attacker aims to design the attack inputs (u, u^{GNSS}) so that the deviation between the robot nominal trajectory and the actual (attacked) trajectory is maximized, while maintaining undetectability (as defined below).

(ii) The trajectory planner seeks for a nominal control input u_{n} to guarantee that, in the absence of attacks, u_{n} allows the robot to reach a desired final state, and, in the presence of attacks, the measurements y^{GNSS} and y_i^{RSSI} allow the robot to detect the attack.

The actions of the attacker and of the trajectory planner can be interpreted in terms of two sequential phases. In the first phase, the trajectory planner designs the nominal control input u_{n} and the control horizon T to satisfy objective (ii). In the second phase, the attacker designs the attacks (u, u^{GNSS}) given the nominal input u_{n} and the nominal model (1)-(2) to satisfy objective (i). We stress that in our settings the nominal control input u_{n} is replaced with the input u by the attacker in the second phase, and thus the choice of the trajectory planner is irreversible and cannot be changed in the second phase.

Related Work: Despite their popularity, GNSS-based localization techniques are subject to a number of well-known vulnerabilities that are typically associated with the lack of appropriate encryption [2]. Existing methods to detect and identify GNSS spoofing attacks are based on filtering techniques to reveal compromised streams of sensory data [3], [4]. Differently, in this letter we focus on characterizing the detectability of attacks modifying both the measurements and the inputs to the system, and on the problem of designing nominal control inputs to restrict or prevent undetectable attacks against this class of cyber-physical systems. Although the security of cyber-physical systems is an extensively-studied topic (see e.g., [5]), most of the available methods are applicable to static systems or systems with linear dynamics [6], [7]. Few exceptions are [8]–[11], which are however restricted to particular classes of nonlinear dynamics, and to attacks modifying the system measurements only. A secure trajectory planning problem has been studied also in our early work [12]. Differently from [12], in this letter we focus on single integrator dynamics that allow us to derive more stringent conditions and explicit controls, and on the possibility of having multiple radio stations.

Paper Contribution: The contribution of this letter is threefold. First, we characterize the class of undetectable attacks against robots with single integrator dynamics operating on a plane. We show how to design undetectable attacks, and demonstrate that attacks can exist only when the robot is located in certain regions of the plane. Second, we formulate and solve an optimization problem that captures the attacker's goal of maximally deviating the robot trajectory from the nominal path. We characterize the form of optimal undetectable attacks, we provide algorithms for their design, and we study the set of positions that are reachable by the attacker. Third, we formalize the trajectory planner's goal of designing secure control inputs, that is, inputs that allow the detection of any attack action. We show that secure control inputs exist only between certain subsets of states, and we illustrate through an example how the trajectory planner can leverage the layout of the radio stations to plan trajectories that are secure.

II. Undetectable Attacks

We start by formalizing the notion of undetectable attack.

Definition 1 (Undetectable Attack): The attack (u, u^{GNSS}), with $u \neq u_{\text{n}}$, is undetectable if $\Omega(x) = \Omega(x_{\text{n}})$ at all times and

$$y^{\text{GNSS}} = y_{\text{n}}^{\text{GNSS}}, \quad \text{and} \quad y_i^{\text{RSSI}} = y_{\text{n},i}^{\text{RSSI}},$$

for all $i \in \Omega(x_{\text{n}})$. Otherwise, the attack is detectable.

Loosely speaking, an attack is undetectable if the measurements generated by the attacked trajectory are compatible with their nominal counterparts and with the nominal dynamics at all times. On the other hand, when Definition 1 is not satisfied, then the attack is readily detected by comparison between the actual and nominal measurements. In particular, an attack is detectable if the stations visited by the nominal and attacked trajectories differ, that is, $\Omega(x(t)) \neq \Omega(x_{\text{n}}(t))$ for some t.

Remark 1 (Undetectability With GNSS Sensor Only): In scenarios where GNSS is the only sensor for detection, an adversary can deliberately alter the control input and remain undetected (under the constraint $\Omega(x) = \Omega(x_{\text{n}})$). To see this, we note that the effect of any attack u can be canceled from the GNSS readings by selecting $u^{\text{GNSS}} = p_{\text{n}} - p$. Thus, secure trajectories can exist only if the robot has redundant measurement in addition to the GNSS readings.

A. Characterization of Undetectable Attacks

Let $\rho_{\text{n},i} = x_{\text{n}} - b_i$ and $\rho_i = x - b_i$ denote the robot nominal and attacked positions relative to the i-th station, and

$$\mathcal{R}(x_{\text{n}}) = \begin{bmatrix} \rho_{\text{n},i_1} & \cdots & \rho_{\text{n},i_s} \end{bmatrix},$$

where $\Omega(x_{\text{n}}) = \{i_1, \ldots, i_s\}$. Let $\text{Rank}(M)$ denote the rank of the matrix M. In the following result we characterize the existence and general expression of undetectable attacks.

Theorem 1 (Undetectable Attacks): There exist undetectable attacks (u, u^{GNSS}) with $u \neq u_{\text{n}}$ only if

$$\text{Rank}(\mathcal{R}(x_{\text{n}}(t))) < 2, \tag{5}$$

for some time t. Moreover, when $\text{Rank}(\mathcal{R}(x_{\text{n}}(t))) \neq 0$ at all times, every undetectable attack satisfies

$$u^{\text{GNSS}} = x_{\text{n}} - x, \ \text{ and } u = v_{\text{r},i}\rho_i + w, \tag{6}$$

for all $i \in \Omega(x_{\text{n}}(t))$, where $v_{\text{r},i} = u_{\text{n}}^{\mathsf{T}}\rho_{\text{n},i}/\|\rho_i\|^2$, $w^{\mathsf{T}}\rho_i = 0$.

Proof: We prove (5) by contrapositive, that is, we show that if $\text{Rank}(\mathcal{R}(x_{\text{n}}(t))) \geq 2$ for all t then every undetectable attack satisfies $u = u_{\text{n}}$ at all times. Let u denote an undetectable attack, and consider the time instant $\tau = 0$. From the assumption $x_{\text{n}}(0) = x(0)$ we obtain $\rho_i(0) = \rho_{\text{n},i}(0)$ and $\rho_j(0) = \rho_{\text{n},j}(0)$ for all $i, j \in \Omega(x_{\text{n}}(\tau))$. Moreover, from undetectability of u, we have $y_i^{\text{RSSI}} - y_{\text{n},i}^{\text{RSSI}} = 0$ and therefore

$$\dot{y}_i^{\text{RSSI}} - \dot{y}_{\text{n},i}^{\text{RSSI}} = u^{\mathsf{T}}\rho_i - u_{\text{n}}^{\mathsf{T}}\rho_{\text{n},i} = 0,$$

for all $i \in \Omega(x_{\text{n}})$ or, equivalently,

$$(u(\tau) - u_{\text{n}}(\tau))^{\mathsf{T}}\begin{bmatrix} \rho_i(\tau) & \rho_j(\tau) \end{bmatrix} = 0,$$

Since $\text{Rank}(\mathcal{R}(x_\text{n}(\tau))) \geq 2$, $\rho_i(\tau)$ and $\rho_j(\tau)$ are linearly independent, and thus $u(\tau) = u_\text{n}(\tau)$ and $x(\tau^+) = x_\text{n}(\tau^+)$. To conclude, we iterate the above reasoning for all $\tau \geq 0$, which yields $u = u_\text{n}$, and shows the implication.

(Expression of Undetectable Attacks): By substituting (2) and (4) into Definition 1 we obtain $x+u^\text{GNSS} = x_\text{n}$, from which $u^\text{GNSS} = x_\text{n} - x$ follows. Next, we take the time derivative of $y_i^\text{RSSI} - y_{\text{n},i}^\text{RSSI} = 0$ and substitute (1) and (3) to obtain

$$\dot{y}_i^\text{RSSI} - \dot{y}_{\text{n},i}^\text{RSSI} = u^\mathsf{T}\rho_i - u_\text{n}^\mathsf{T}\rho_{\text{n},i} = 0, \tag{7}$$

which implies that u can be decomposed as $u = v_{\text{r},i}\rho_i + w$, with $w^\mathsf{T}\rho_i = 0$ and $v_{\text{r},i} = u_\text{n}^\mathsf{T}\rho_{\text{n},i}/\|\rho_i\|^2$, which shows the claimed result and concludes the proof. ∎

Theorem 1 suggests that the existence of undetectable attacks depends on $\Omega(x_\text{n})$, and thus on the set of radio stations visited by the nominal trajectory. In particular, the theorem implies that undetectable attacks can exist only under three circumstances. First, $|\Omega(x_\text{n}(t))| = 0$ for some $t \in \mathbb{R}_{\geq 0}$ (in this case, any attack is undetectable as also discussed in Remark 1). Second, $|\Omega(x_\text{n}(t))| = 1$ for some $t \in \mathbb{R}_{\geq 0}$ (i.e., there exists a time t such that the nominal trajectory visits a single radio station). Third, $|\Omega(x_\text{n}(t)| > 1$ and the robot position $x_\text{n}(t)$ is collinear with the coordinates of all available radio stations for some $t \in \mathbb{R}_{\geq 0}$ (i.e., there exists a time t such that $x_\text{n}(t)$ and b_i for all $i \in \Omega(x_\text{n}(t))$ are collinear). Further, the theorem provides a systematic way to design undetectable attacks when the attacker knows the nominal input. In particular, the signal w in equation (6) can be arbitrarily selected by an attacker and it does not affect detectability. Finally, we emphasize that the theorem characterizes the existence of undetectable attacks in relation to the nominal path followed by the robot. As we will later demonstrate in this letter (see Section III), the existence of undetectable attacks can be further refined by appropriately designing the nominal control inputs. Fig. 1 illustrates the regions of the plane where undetectable attacks can exist.

Remark 2 (Condition Rank($\mathcal{R}(x_n(t))) = 0$): In the particular situation where $\text{Rank}(\mathcal{R}(x_\text{n}(t))) = 0$ for some t, we necessarily have $|\Omega(x_\text{n}(t))| = 1$ and $x_\text{n}(t) = b_i$, that is, the nominal position of the robot overlaps with the position of the (unique) radio station. In fact, these circumstances and under the assumption of non-overlapping radio stations ($b_i \neq b_j$, if $i \neq j$) we either have $\rho_{\text{n},i}(t) = 0$ or $\rho_{\text{n},j}(t) = 0$. In this case, undetectability imposes no constraints on the attack input. In fact, whenever $x_\text{n}(t) - b_i = 0$, any bounded $u(t)$ satisfies the notion of undetectability in Definition 1.

B. Design of Optimal Undetectable Attacks

We now illustrate how an attacker can design optimal undetectable attacks, that is, attacks that maximize the deviation between the nominal and attacked trajectories while maintaining undetectability. We focus on the case $|\Omega(x_\text{n})| = 1$, and formalize the problem as follows:

$$\begin{aligned} \delta^* = \max_{w} \ & \|x(T) - x_\text{n}(T)\|, \\ \text{subject to } & \dot{x} = u, && \text{(8a)} \\ & u = v_\text{r}\ \rho + w, && \text{(8b)} \\ & \|u\| \leq u_\text{max}, && \text{(8c)} \end{aligned}$$

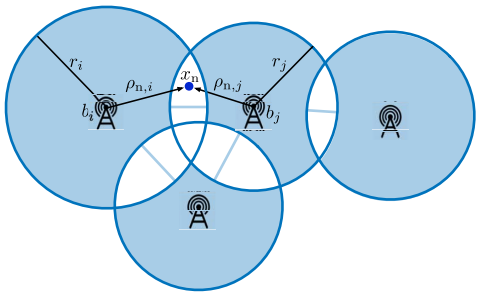

Fig. 1. Areas shaded in blue denote the regions where $\text{Rank}(\mathcal{R}(x_\text{n})) < 2$, that is, the regions where undetectable attacks can exist.

where $T \in \mathbb{R}_{\geq 0}$ denotes the control horizon of u_n, constraint (8b) ensures undetectability of the attack, and the expression for v_r and w are listed[1] in Theorem 1. Let e_i denote the i-th canonical vector of appropriate dimension. The following result characterizes the general expression of optimal attacks.

Theorem 2 (Optimal Undetectable Attacks): Let w^* be an optimal solution to the maximization problem (8). Then,

$$w^* = \gamma\sqrt{u_\text{max}^2 - v_\text{r}^2\|\rho\|^2}\,\frac{\tilde{w}}{\|\tilde{w}\|}, \tag{9}$$

where $\gamma : [0, T] \to \{-1, 0, 1\}$, and $\tilde{w}$ is any vector that satisfies $\tilde{w}^\mathsf{T}\rho = 0$. Moreover, let the nominal input be decomposed as $u_\text{n} = \alpha\rho_\text{n} + z$, with $\rho_\text{n}^\mathsf{T} z = 0$ and $\alpha \in \mathbb{R}$. Then, the optimal deviation δ^* satisfies

$$\delta^* = 2\|x_\text{n}(T)\|\sin(\theta_T/2),$$

where

$$\theta_T = \int_0^T \frac{\|w^*\| - \|z\|}{\|r\|}dt.$$

Proof: To show (9), we use the Pontryagin's Maximum Principle [13] to derive optimality conditions for the optimization problem (8). In particular, we rewrite $w = \sigma\frac{\tilde{w}}{\|\tilde{w}\|}$, where $\sigma \in \mathbb{R}$ and $\tilde{w}^\mathsf{T}\rho = 0$, and consider the Hamiltonian

$$\mathcal{H}(t, x, w, \lambda) = \lambda^\mathsf{T}(v_\text{r}\ \rho + \sigma\frac{\tilde{w}}{\|\tilde{w}\|}),$$

where $\lambda : [0, T] \to \mathbb{R}^2$, with the additional constraint $\|u\|^2 \leq u_\text{max}^2$ or, equivalently, $v_\text{r}^2\|\rho\|^2 + \sigma^2 \leq u_\text{max}^2$. Notice that the Hamiltonian is a function of time because of the dependence on v_r. By application of the Maximum Principle [14], the optimal control input at all times minimizes the Hamiltonian over the set of bounded attack inputs $U(t) = \{\sigma : v_\text{r}^2\|\rho\|^2 + \sigma^2 \leq u_\text{max}^2\}$, that is, the optimal σ^* satisfies

$$\begin{aligned} \sigma^* &= \arg\min_{\sigma\in U(t)} \mathcal{H}(t, x, w, \lambda) = \arg\min_{\sigma\in U(t)} \left(\frac{\sigma}{\|\tilde{w}\|}\lambda^\mathsf{T}\tilde{w}\right) \\ &= -\sqrt{u_\text{max}^2 - v_\text{r}^2\|\rho\|^2}\,\text{sign}(\lambda^\mathsf{T}\tilde{w}), \end{aligned}$$

where sign denotes the sign function, which proves (9).

To show the given expression for δ^*, we observe that the ratio $\|w\|/|\rho\|$ is the tangential velocity of the attacked trajectory; similarly, $\|z\|/\|r\|$ is the tangential velocity in the nominal trajectory. Thus, the angle between the vectors $x_\text{n}(T)$

[1] In the remainder, we omit the subscript i when $|\Omega(x_\text{n})| = 1$.

and $x(T)$ can be obtained by integrating the instantaneous difference between the two tangential velocities as

$$\theta_T = \int_0^T \frac{\|w^*\|}{\|\rho\|} - \frac{\|z\|}{\|r\|} dt = \int_0^T \frac{\|w^*\| - \|z\|}{\|r\|} dt, \qquad (10)$$

where we used $\|r\| = \|\rho\|$ since the attack is undetectable. To conclude, we note that when $\|x_n(T)\| = \|x(T)\|$ the deviation in trajectory can be related to the angular deviation by means of the following geometric relationship

$$\|x_n(T) - x(T)\| = 2\|x_n(T)\| \sin(\theta_T/2),$$

which shows the given expression for δ^* and concludes the proof. ■

From Theorem 2, optimal attacks are of the form of a feedback controller (that depends on the instantaneous values of u_n, x_n, and x through $\tilde{w}$), which switches abruptly between two (time-varying) expressions, and where the switching instants are determined by the function γ. Next, we propose an algorithm to determine the optimal switching times of the function γ. To this aim, we choose by convention the vector $\tilde{w}$ that minimizes the counterclockwise angle between ρ and $\tilde{w}$. Our method is illustrated in Algorithm 1 and relies on the following rationale to identify the control input that leads to optimal deviations: if the counterclockwise angle between $\rho(t)$ and $-\rho_n(T)$ is smaller than π, then $\gamma(t) = 1$; if such angle is larger than π, then $\gamma(t) = -1$; if such angle equals to zero, then $\gamma(t) = 0$. Finally, we observe that when the angle between the vectors $\rho(t)$ and $-\rho_n(T)$ equals to π, either choice $\gamma(t) = 1$ or $\gamma(t) = -1$ will result in an optimal solution of (8). In Algorithm 1, we let $\gamma(t) = 1$ in this situation. We formalize the optimality of Algorithm 1 in the following theorem.

Theorem 3 (Optimality of Algorithm 1): Let w be the output of Algorithm 1. Then, w is a solution to (8).

Proof: Let w be the output of Algorithm 1, let γ denote the corresponding switching function, and let u_n be decomposed as in Theorem 2. Let

$$\varphi_T := \int_0^T \frac{\|w\| - \|z\|}{\|r\|} dt,$$

denote the angular deviation between nominal and attacked trajectories, obtained by integrating the difference between the tangential velocities, and recall that every optimal attack satisfies $w^* = \gamma\sqrt{u_{\max}^2 - v_r^2\|\rho\|^2}\frac{\tilde{w}}{\|\tilde{w}\|}$, and $\delta^* = 2\|x_n(T)\| \sin(\theta_T/2)$, where θ_T is defined in (10). To show that w is a minimizer of (8), we equivalently show that $\varphi_T = \theta_T$. We prove this statement by contradiction, and distinguish among two cases.

(Case 1): $|\theta_T| > |\varphi_T|$. By replacing the integral expressions, we obtain

$$\left|\int_0^T \frac{\|w^*\| - \|z\|}{\|r\|} dt\right| > \left|\int_0^T \frac{\|w\| - \|z\|}{\|r\|} dt\right|,$$

which implies that there must exist t such that $\|w^*(t)\| > \|w(t)\|$. Since both w and w^* satisfy (9) and $|\gamma| = 1$ at all times, the above relationship results in a contradiction.

(Case 2): $|\theta_T| < |\varphi_T|$. We first observe that the scenario $|\theta_T| < |\varphi_T| < \pi$ immediately results in a contradiction, since w^* is, by assumption, an optimal solution to (8) and thus a minimizer of $|\theta_T - \pi|$. On the other hand, $|\varphi_T| > \pi$ is also a contradiction since in the algorithm $w(t) = 0$ whenever $\varphi_t = \pi$, which shows the result and concludes the proof. ■

Algorithm 1: Optimal Solutions to (8)

Input: $x_n(T)$, $u_{\max}$
Output: w solution to (8)
repeat
  Measure instantaneous values of x, x_n, u_n;
  $\phi \leftarrow$ Angle between $x(t)$ and $-x_n(T)$;
  if $\phi = 0$ **then**
    $\gamma \leftarrow 0$;
  else if $0 < \phi \leq \pi$ **then**
    $\gamma \leftarrow 1$;
  else
    $\gamma \leftarrow -1$;
  $v_r \leftarrow u_n^\mathsf{T}\rho_n/\|\rho\|^2$;
  $w \leftarrow \gamma\sqrt{u_{\max}^2 - v_r^2\|\rho\|^2}\frac{\tilde{w}}{\|\tilde{w}\|}$;
until $x_n = x_n(T)$;
return w

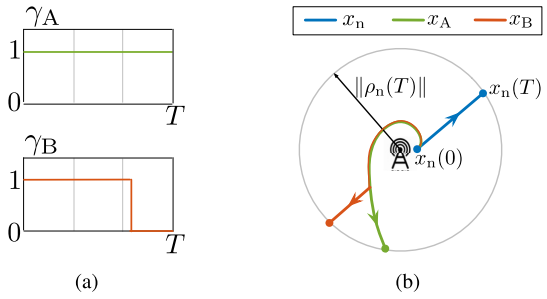

Fig. 2. (a) Suboptimal and optimal switching functions, and (b) corresponding trajectories. The circle shows that $\|\rho_n(T)\| = \|\rho_A(T)\| = \|\rho_B(T)\|$.

Fig. 2 illustrates optimal trajectories resulting from Algorithm 1, and shows a comparison between optimal attack trajectories and suboptimal attacks obtained when $\gamma = 1$ at all times. It is worth noting that the control law described in Algorithm 1 is of feedback type, that is, the instantaneous value of the control inputs v_r, w, and u are computed by using the current (measured) values of x, x_n, and u_n. Thus, differently from [12], optimal undetectable attacks can be cast by using instantaneous measurements of u_n and x_n, and without the full knowledge of the nominal open loop signals.

In the following remark, we discuss the set of positions that can be reached by a robot under attack.

Remark 3 (Reachable Positions Under Attacks): The set of positions that can be reached by an (undetectable) attacker depend only on the choice of u_n performed by the trajectory planner, and can be computed as follows. Let x_1 and x_{-1} denote the trajectories resulting from (3) with an undetectable attack input of the form (9) with $\gamma = 1$ and $\gamma = -1$ at all times, respectively. Moreover, let $\text{arc}(x_1(T), x_{-1}(T))$ be the arc of a circle that is centered at the station position b with radius $\|\rho_n(T)\|$, and containing $x_n(T)$ (see Fig. 3 for an illustration). For every $\bar{x} \in \text{arc}(x_1(T), x_{-1}(T))$, there exists a control input that steers the robot from $x(0)$ to $x(T) = \bar{x}$. In fact, it can be shown that the output of Algorithm 1 with $x_n(T) = -\bar{x}$ reaches the desired final state $\bar{x}$ (see Fig. 3).

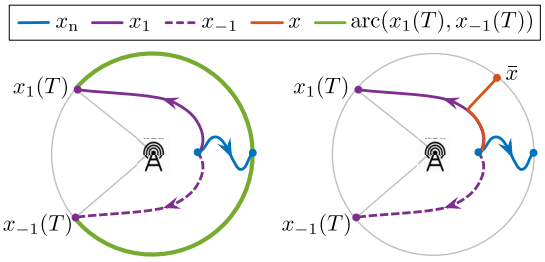

Fig. 3. Positions reachable by attackers (left) and example of attack trajectory obtained from Algorithm 1 (right).

III. Secure Navigation

This section is devoted to the characterization and design of secure trajectories. We consider scenarios where undetectable attacks can exist (see Theorem 1) and focus on the problem of designing nominal control inputs that ensure that every attack is detected. We say that a trajectory x_n is *secure* if, for all attacks u, one of the following mutually exclusive conditions is satisfied:

(C1) $x = x_n$ at all times; or

(C2) if $x \neq x_n$ at some time, then the attack is detectable.

A control input is *secure* if the resulting trajectory is secure.

Theorem 4 (Secure Control Inputs): Let $|\Omega(x_n)| = 1$ at all times. The control input u_n is secure if and only if the following conditions hold simultaneously:

(1) there exists a function $\kappa : \mathbb{R}_{\geq 0} \to \{-1, 1\}$ satisfying

$$u_n = \kappa \frac{\rho_n}{\|\rho_n\|} u_{\max}, \tag{11}$$

(2) the trajectory ρ_n satisfies $\rho_n \neq 0$ at all times.

Proof (Only if): To prove that (C1)-(C2) imply (1)-(2), we equivalently show that if (1)-(2) do not simultaneously hold, then there exists an undetectable attack that violates (C1)-(C2). We distinguish among two cases.

(Case 1): There exists a time instant τ such that (11) does not hold, that is, $u_n^\mathsf{T}(\tau)\rho_n(\tau) < u_{\max}$. Consider the attack input u satisfying $u^\mathsf{T}\rho = u_n^\mathsf{T}\rho_n$ at all times, and $\|u(\tau)\| = u_{\max}$. By construction, u is undetectable (see Theorem 1) and satisfies $u \neq u_n$, which violates (C1) and (C2).

(Case 2): There exists τ such that $\rho_n(\tau) = 0$. Under this assumption, every u satisfying $u^\mathsf{T}\rho = u_n^\mathsf{T}\rho_n$ at all times and $u(\tau) \neq u_n(\tau)$ is undetectable. In fact, whenever $\rho_n = 0$ undetectability imposes no constraints on the attack input, and concludes the proof of the implication.

(If): Assume the two conditions (1)-(2) hold. If the attack input does not satisfy $u^\mathsf{T}\rho = u_n^\mathsf{T}\rho_n$, then the attack is detectable and (C2) is verified. On the other hand, assume u is undetectable, that is, $u^\mathsf{T}\rho = u_n^\mathsf{T}\rho_n$ (and thus $\|\rho\| = \|\rho_n\|$) at all times. Then,

$$u_{\max}\|\rho_n\| = |u_n^\mathsf{T}\rho_n| = |u^\mathsf{T}\rho| \leq u_{\max}\|\rho\|,$$

where we substituted (11) and used the triangle inequality. Since $\|\rho\| = \|\rho_n\|$, exact equality must hold and the vectors u and ρ are linearly dependent with $\|u\| = u_{\max}$ at all times. To conclude, we note that $u = -u_n$ results in a violation of the undetectability assumption $u^\mathsf{T}\rho = u_n^\mathsf{T}\rho_n$, therefore $u = u_n$ at all times, which shows (C1) and concludes the proof. ■

Theorem 4 provides an explicit characterization of secure control inputs: it shows that every secure input has maximum magnitude at all times, and its direction is parallel to vector ρ_n. Two significant implications follow from Theorem 4. First, the result shows that appropriate control design prevents the existence of undetectable attacks. Second, it shows that whenever the nominal quantities do not satisfy conditions (1)-(2), then undetectable attacks always exist. Further, we note that this result extends the conclusions of Theorem 1 by showing that condition (5) is also sufficient for the existence of undetectable attacks for general choices of u_n.

Remark 4 (Game-Theoretic Interpretation of the Results): The security problem considered in this letter can equivalently be studied in a game-theoretic framework, and, specifically, as a *Stackelberg game* [15]. In fact, the secure trajectories in Theorem 4 can be viewed as the strategies that maximize the payoff of the trajectory planner, which anticipates the fact that the attacker will adopt its best response. This strategy is open-loop and independent of the attacker's action, which takes place subsequently. The undetectable attacks in Theorem 2, instead, can be viewed as the best response of the attacker given the strategy of the trajectory planner and the attacker's objectives. The attacker's strategy is of feedback form, because the best response of the attacker depends on the strategy of the trajectory planner to maintain undetectability and maximize the payoff. We remark that alternative formulations of the problem are also possible, where, for instance, the actions of the trajectory planner and the attacker occur concurrently.

Next, we focus on characterizing the set of initial and final positions that can be reached via secure trajectories. To this aim, we chose the coordinate system so that $b = 0$ and $x_n = \rho_n$, and let sign() be the sign function, with $\mathrm{sign}(0) = 0$.

Theorem 5 (Reachable Positions via Secure Control Inputs): Let $|\Omega(x_n)| = 1$ and u_n be a secure control input. Then, for all $T \in \mathbb{R}_{\geq 0}$,

$$x_n(T) \in \mathcal{S}(x_n(0)),$$

where $\mathcal{S}(x_n(0)) = \{x : x = \alpha x_n(0), \alpha \in \mathbb{R}_{>0}\}$. Moreover, for any $\bar{x}_n \in \mathcal{S}(x_n(0))$ the secure control input (11) with

$$\kappa = \mathrm{sign}(\|\bar{x}_n\| - \|x_n(0)\|), \tag{12}$$

steers the robot from $x_n(0)$ to $x_n(T) = \bar{x}_n$, with $T = \frac{\|\bar{x}_n\|^2}{4u_{\max}}$.

Proof (Reachable Set): We first show that for every secure control input the quantity x_1/x_2 is time-invariant, that is, $\frac{d}{dt}\frac{x_1}{x_2} = 0$. By expanding the time derivative we obtain

$$\dot{x}_1 x_2^{-1} - x_1 x_2^{-2}\dot{x}_2 = 0,$$

where we substituted (1) and (11). Next, we prove that $\alpha > 0$. Assume, by contradiction, that $x(T) = \alpha_T x_n(0)$, and $\alpha_T < 0$. By continuity of x_n, there exists $\tau \in [0, T)$ such that $x(\tau) = \alpha_\tau x_n(0)$, with $\alpha_\tau = 0$. But this violates the assumption that u is secure (condition (2) in Theorem 4), which contradicts the assumption and proves the claim.

(Expression for Secure Control Input): Let u_n be as in (11) and let $n := \|x_n\|^2$. Then, by substituting (11), we obtain

$$\dot{n} = \frac{d}{dt} x_n^\mathsf{T} x_n = 4\dot{x}_n x_n = 4\kappa u_{\max},$$

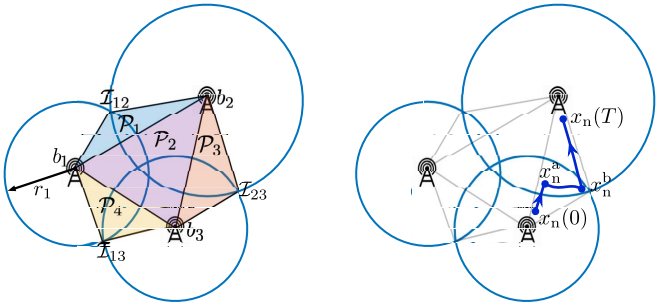

Fig. 4. (Left) Notation used in Example 1. (Right) Intermediate waypoints.

Moreover,

$$n(T) = \|x_{\rm n}(T)\|^2 = \int_0^T 4\kappa u_{\max} dt$$
$$= \text{sign}(\|\bar{x}_{\rm n}\| - \|x_{\rm n}^0\|) u_{\max} T = \|\bar{x}_{\rm n}\|^2,$$

where we have substituted the expression for κ and T. To conclude, we note that since $x_{\rm n}(T) \in \mathcal{S}(x_{\rm n}(0))$ and $\bar{x}_{\rm n} \in \mathcal{S}(x_{\rm n}(0))$, we necessarily have $x_{\rm n}(T) = \bar{x}_{\rm n}$, which shows the claimed result and concludes the proof. ■

Theorem 5 shows that the set of configurations that are reachable via secure trajectories from x_0 are described by the line passing through the points x_0 and the origin of the reference frame. We conclude this section by illustrating how the above results can be combined when $|\Omega(x_{\rm n})| \geq 2$.

Example 1 (Secure Navigation): Consider the scenario illustrated in Fig. 4, consisting of $n_{\rm b} = 3$ radio stations. For all $i, j \in \{1, 2, 3\}$, we let $\mathcal{I}_{ij} = \{x : \|x - b_i\| = r_i \text{ and } \|x - b_j\| = r_j\}$ denote the intersection points between the circles that identify the communication ranges of the radio stations. Further, let

$$\mathcal{P}_1 = \{b_1, \mathcal{I}_{12}, b_2\}, \quad \mathcal{P}_2 = \{b_1, b_2, b_3\},$$
$$\mathcal{P}_3 = \{b_2, \mathcal{I}_{23}, b_3\}, \quad \mathcal{P}_4 = \{b_1, b_3, \mathcal{I}_{13}\},$$

denote the polygons that originate from the locations of the RSSI stations (i.e., b_i) and the intersection points (i.e., $\mathcal{I}_{ij}$), see Fig. 4(a) for an illustration. As an illustrative example, consider any initial position $x_{\rm n}(0) \in \{x : x \in \mathcal{P}_3 \text{ and } \Omega(x) = \{3\}\}$, that is, any initial position that is located in the polygon $\mathcal{P}_3$ and within the communication range of station 3. Moreover, consider any final position $\bar{x}_{\rm n} \in \{x : x \in \mathcal{P}_3 \text{ and } \Omega(x) = \{2\}\}$, that is, any final position that is located in the polygon $\mathcal{P}_3$ and within the communication range of station 2 (see Fig. 4(b)). Moreover, define the sets

$$\chi_{\rm n}^{\rm A} := \{x : x = \alpha x_{\rm n}(0), \alpha \in \mathbb{R}_{>0}, \text{ and } \Omega(x) = \{2, 3\}\},$$
$$\chi_{\rm n}^{\rm B} := \{x : x = \alpha \bar{x}_{\rm n}, \alpha \in \mathbb{R}_{>0}, \text{ and } \Omega(x) = \{2, 3\}\},$$

which describe the positions that are reachable from $x_{\rm n}(0)$ and $\bar{x}_{\rm n}$, respectively, and that belong to the intersection between the communication ranges of stations 2 and 3. Notice that these sets are nonempty since the intersection between the communication ranges of stations 2 and 3 is non-empty. Now, let $x_{\rm n}^{\rm a} \in \chi_{\rm n}^{\rm A}$ and $x_{\rm n}^{\rm b} \in \chi_{\rm n}^{\rm B}$. Then, a secure control input from $x(0)$ to $\bar{x}$ is as follows:

(i) Apply the secure control input given by (11) with $\kappa = \text{sign}(\|x_{\rm n}^{\rm a}\| - \|x_{\rm n}(0)\|)$ until $x_{\rm n} = x_{\rm n}^{\rm a}$;

(ii) Apply any control input $u_{\rm n}$ that satisfies $\Omega(x_{\rm n}) = \{2, 3\}$ until $x_{\rm n} = x_{\rm n}^{\rm b}$;

(iii) Apply the secure control input given by (11) with $\kappa = \text{sign}(\|\bar{x}_{\rm n}\| - \|x_{\rm n}^{\rm b}\|)$ until $x_{\rm n} = \bar{x}_{\rm n}$;

We note that the geometry of the problem and Theorem 5 guarantee the existence of the secure control input defined in steps (i)-(iii). An illustration of the trajectory resulting from the above algorithm is presented in Fig. 4.

IV. Conclusion

In this letter we consider the problem of designing the trajectories of a robot when its measurements are maliciously compromised by an attacker. We demonstrate the existence of undetectable attacks in relation to the region of the plane where the robot is located, and present an efficient algorithm to cast optimal undetectable attacks. Conversely, we show how a trajectory planner can leverage the layout of the radio stations to design control inputs that allow the detection of any attack. Our results demonstrate that appropriate control design can enhance the security of robots operating in adversarial environments. Interesting aspects that require further investigation include the design of optimal attacks with multiple radio stations, the study of the effects of noise, and the assessment of the performance of the method in an experimental setup.

References

[1] M. Jun and R. D'Andrea, "Path planning for unmanned aerial vehicles in uncertain and adversarial environments," in *Cooperative Control: Models, Applications and Algorithms*. Boston, MA, USA: Springer, 2003, pp. 95–110.

[2] D. P. Shepard, T. E. Humphreys, and A. A. Fansler, "Evaluation of the vulnerability of phasor measurement units to GPS spoofing attacks," *Int. J. Crit. Infrastruct. Protect.*, vol. 5, nos. 3–4, pp. 146–153, 2012.

[3] A. Broumandan, A. Jafarnia-Jahromi, V. Dehghanian, J. Nielsen, and G. Lachapelle, "GNSS spoofing detection in handheld receivers based on signal spatial correlation," in *Proc. ION Position Location Navig. Symp.*, Apr. 2012, pp. 479–487.

[4] X. Jiang, J. Zhang, B. J. Harding, J. J. Makela, and A. D. Domínguez-García, "Spoofing GPS receiver clock offset of phasor measurement units," *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 3253–3262, Aug. 2013.

[5] Y. Z. Lun, A. D'Innocenzo, F. Smarra, I. Malavolta, and M. D. D. Benedetto, "State of the art of cyber-physical systems security: An automatic control perspective," *J. Syst. Softw.*, vol. 149, pp. 174–216, Mar. 2019.

[6] F. Pasqualetti, F. Dörfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Trans. Autom. Control*, vol. 58, no. 11, pp. 2715–2729, Nov. 2013.

[7] Y. Mo and B. Sinopoli, "Secure control against replay attacks," in *Proc. Allerton Conf. Commun. Control Comput.*, Monticello, IL, USA, Sep. 2010, pp. 911–918.

[8] J. P. Hespanha and S. D. Bopardikar, "Output-feedback linear quadratic robust control under actuation and deception attacks," in *Proc. Amer. Control Conf.*, Philadelphia, PA, USA, Jul. 2019.

[9] Y. Shoukry, P. Nuzzo, N. Bezzo, A. L. Sangiovanni-Vincentelli, S. A. Seshia, and P. Tabuada, "Secure state reconstruction in differentially flat systems under sensor attacks using satisfiability modulo theory solving," in *Proc. IEEE Conf. Decis. Control*, Dec. 2015, pp. 3804–3809.

[10] Q. Hu, D. Fooladivanda, Y. H. Chang, and C. J. Tomlin, "Secure state estimation and control for cyber security of the nonlinear power systems," *IEEE Trans. Control Netw. Syst.*, vol. 5, no. 3, pp. 1310–1321, Sep. 2018.

[11] 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 Trans. Autom. Control*, vol. 64, no. 3, pp. 1162–1169, Mar. 2019.

[12] Y.-C. Liu, G. Bianchin, and F. Pasqualetti, "Secure trajectory planning against undetectable spoofing attacks," *arXiv preprint arXiv:1902.10869*, 2019.

[13] I. M. Gelfand and S. V. Fomin, *Calculus of Variations*. New York, NY, USA: Courier Corporat., 2000.

[14] R. F. Hartl, S. P. Sethi, and R. G. Vickson, "A survey of the maximum principles for optimal control problems with state constraints," *SIAM Rev.*, vol. 37, no. 2, pp. 181–218, 1995.

[15] T. Başar and G. J. Olsder, *Dynamic Noncooperative Game Theory*, 2nd ed. Philadelphia, PA, USA: SIAM, 1999.