Affine Transformation-based Perfectly Undetectable False Data Injection Attacks on Remote Manipulator Kinematic Control with Attack Detector

Jun Ueda[1] and Jacob Blevins[1]

***Abstract*— This paper demonstrates the viability of perfectly undetectable affine transformation attacks against robotic manipulators where intelligent attackers can inject multiplicative and additive false data while remaining completely hidden from system users. The attacker can implement these communication line attacks by satisfying three Conditions presented in this work. These claims are experimentally validated on a FANUC 6 degree of freedom manipulator by comparing a nominal (non-attacked) trial and a detectable attack case against three perfectly undetectable trajectory attack Scenarios: scaling, reflection, and shearing. The results show similar observed end effector error for the attack Scenarios and the nominal case, indicating that the perfectly undetectable affine transformation attack method keeps the attacker perfectly hidden while enabling them to attack manipulator trajectories.**

***Index Terms* – False data injection attack, Jacobian velocity control, Affine transformation**

I. Introduction

In the era of Industry 4.0, virtually all modern devices are interconnected via the internet, facilitating the exchange of sensor measurements, control commands, and other vital information for monitoring and controlling complex systems through computer networks [1]. The cybersecurity of these communication channels is increasingly under threat from various adversaries, raising significant concerns [2] [3]. A typical networked control system is illustrated in Fig. 1, wherein a physical plant such as an industrial robotic manipulator is connected to the internet, allowing it to be controlled remotely by the user. The controller receives measurements from the plant via a communication channel and computes control commands that are sent back to the plant via another communication channel, forming a closed-loop system for remote control of the manipulator.

One notable form of cybersecurity threat is the False Data Injection Attack (FDIA), wherein adversaries tamper with sensor, control, or monitoring signals within the communication lines to adversely affect the behavior, performance, and stability of the targeted system [4]. Rather than gaining unauthorized access to the controller or the plant (including its local, low-level embedded controllers), conducting FDIA on communication lines is likely a more efficient and secretive approach for adversaries looking to disrupt operations.

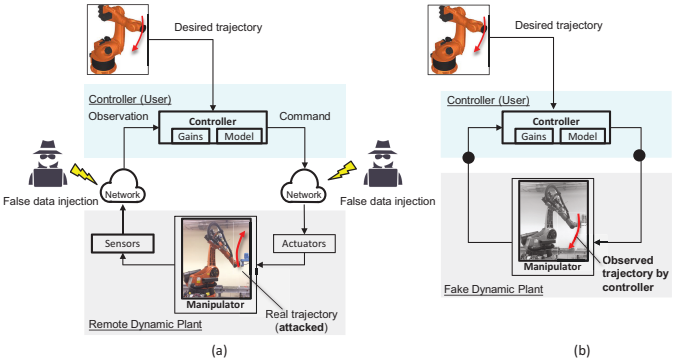


Fig. 1: Conceptual diagram of false data injection attack (FDIA) on remote manipulator: (a) Attacked networked control system with coordinated FDIA on the commands and observables. (b) Plant dynamics as perceived by the controller, indistinguishable from the nominal plant behavior and thus undetectable.

There are three types of FDIAs: detectable, undetectable, and stealthy [4]. Detectable FDIAs are attacks in which the controller can observe an active injection attack; this form of attack is well studied and can be easily rejected or compensated for. Current literature for detectable attacks suggests isolating observed anomalies to determine if there is an attack and to understand its type and the affected channels. If the attack is detected, a fault can be thrown [4]. Other methods for defending against detectable attacks include comparing sensor values against pre-determined models or expected values and adapting the control input as if the attack is a disturbance [5]. Additionally, adaptive control can be used to estimate a scalar multiplicative attack constant and reject it in the control law [6]. Studies on attacks to physical systems have been previously implemented in [7] wherein researchers defend against detectable attacks to UAVs via error analysis with a sliding mode compensation scheme.

Undetectable and stealthy attacks are characterized by attacks which are more difficult for the operator to detect. In the case of an undetectable attack, the attacked signals coincide with those that are within regular operating range, thus faults and regular detectors fail [8]. Perfectly undetectable attacks are those where there is no change in observed states, but the system is still attacked. A relative of the undetectable attack is the stealthy attack. Stealthy attacks are those where an attack detector is explicitly taken into account; therefore, ensuring

Manuscript received: April 11, 2024; Revised July 19, 2024; Accepted August 19, 2024.

This paper was recommended for publication by Editor Jaydev P. Desai upon evaluation of the Associate Editor and Reviewers' comments. This work was supposed in part by NSF CMMI Grant 2112793.

[1]Jun Ueda and Jacob Blevins are with the George W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0405, USA. (e-mail: jun.ueda@me.gatech.edu, jacob.blevins@gatech.edu).

Digital Object Identifier (DOI): see top of this page.

that the attacker does not set off an alarm [4].

The significance of this research is the formulation of a generalized FDIA that involves coordinated multiplicative and additive data injections on both control commands and observables, allowing attackers to create perfectly undetectable attacks against kinematic manipulator systems. As far as the authors are aware, current literature studies either additive or multiplicative FDIA on control commands or observables, and often not for perfectly undetectable attacks [9], [10]. Covert attack literature has been aimed at both commands and observables and may include multiplication and addition, but just for stealthy FDIAs [11]. By taking remote manipulator kinematic control as a representative example, which is also considered in FDIA literature [6], this paper contributes a specific structure of the plant dynamics, from commands to observables, that allow for a range of perfectly undetectable FDIAs, *regardless* of the type of feedback control and attack detector. The mathematical analysis of this proof is given and tested on a FANUC manipulator.

II. Perfectly Undetectable FDIA from the Controller's Perspective

A. FDIA on networked control systems

In game theory, an attacker's strategy for a perfectly undetectable attack aims to maximize impact while minimizing detection risk [12], [13]. This approach requires deep understanding of the defender's measures and the security landscape [14]. The attacker's utility function, which quantifies their strategy and risk tolerances, guides decisions to optimize attack effectiveness with minimal exposure [15].

When FDIA on a networked control system is concerned, as shown in Fig. 1(a), common attack strategies include sensor spoofing, destabilization, and performance degradation, each aiming to elicit responses from the plant that deviate from those observed under normal operation. While FDIA on control commands is chosen based on specific objectives, an intelligent attacker would avoid using solely command attacks that lead to performance degradation [6] or immediate instability [5]. Such command attacks, due to their noticeable impact on observables, are easily detectable.

Instead, an intelligent attacker would opt for an attack that is difficult to detect or remains perfectly undetected by the users [4], [11], [16], [17]. Perfectly undetectable attacks from the plant's perspective are defined in the literature, as shown in Appendix I for interested readers' reference. In this definition, it is assumed that the detector receives ground truth observables without being compromised by FDIA.

B. Perfectly undetectable FDIA via coordinated attacks on commands and observables

This paper adopts methods in [6] for FDIA on networked manipulator kinematic control, and we define perfectly undetectable attacks from the perspective of the controller, detailing specifications related to the networked control architecture illustrated in Fig. 1. Illustrated in Fig. 2, typical Jacobian-based joint velocity control is where $\boldsymbol{q} = [\theta_1, \theta_2, \cdots,]^T \in R^n$ is a joint angle vector, $\boldsymbol{r}$ is the end-point position and orientation in the 3D space, typically, $\boldsymbol{r} \in R^6$, $\boldsymbol{r} = f(\boldsymbol{q})$ is forward kinematics, and $\boldsymbol{J}(\boldsymbol{q}) = \frac{\partial f}{\partial \boldsymbol{q}^T}$ is the Jacobian matrix. With the desired end-effector trajectory $\boldsymbol{r}_d$, commonly, a simple kinematic control law utilizing Jacobian transpose is given as:

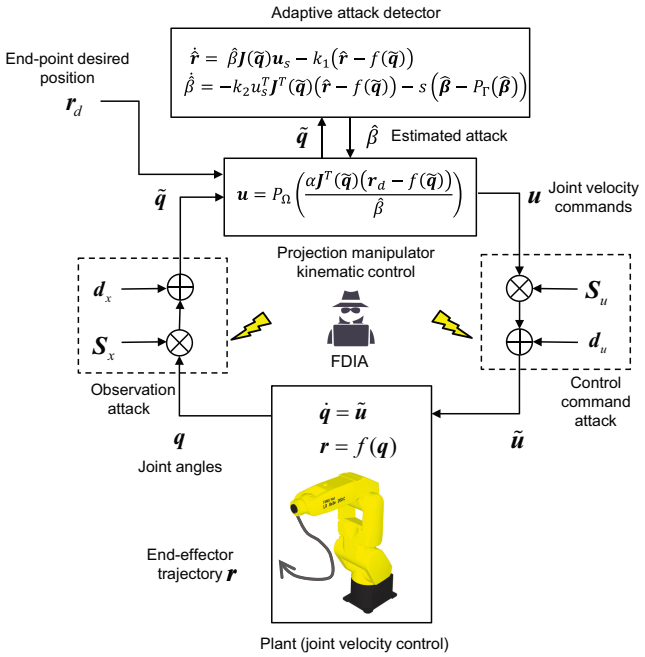


Fig. 2: FDIA on manipulator joint angles: Attack detector and control scheme adopted from [6].

$$\boldsymbol{u} = \boldsymbol{J}(\boldsymbol{q})^T(\boldsymbol{r}_d - \boldsymbol{r}), \tag{1}$$

where $\boldsymbol{u} \in R^m$ is a control velocity command. Variations of control implementations will be discussed in Section V.

A generalized form of FDIA that involves coordinated multiplicative and additive data injections into both control commands and observables is represented in Fig. 2 by introducing affine transformations. $\tilde{\boldsymbol{q}}$ is a compromised observables vector resulting from the attack, $\boldsymbol{S}_x \in R^{n\times n}$ represents an arbitrary transformation, such as scaling, reflection, and shearing, and $\boldsymbol{d}_x \in R^n$ represents a translation introducing an offset. This paper assumes $\boldsymbol{S}_x$ and $\boldsymbol{d}_x$ are constants. The affine transformation attack to the observables is

$$\tilde{\boldsymbol{q}} = \boldsymbol{\alpha}(\boldsymbol{q}) = \boldsymbol{S}_x\boldsymbol{q} + \boldsymbol{d}_x, \tag{2}$$

where $\boldsymbol{\alpha}$ is a static observables attack function. Similarly, to the control command, $\tilde{\boldsymbol{u}}$ is a compromised (attacked) control command vector resulting from the attack by a static attack function, $\boldsymbol{\beta}$. When $\boldsymbol{\beta}$ is linear affine transformation, $\boldsymbol{S}_u \in R^{m\times m}$ represents an arbitrary transformation and $\boldsymbol{d}_u \in R^m$ represents a translation introducing an offset. This paper assumes that $\boldsymbol{S}_u$ and $\boldsymbol{d}_u$ are constants.

$$\tilde{\boldsymbol{u}} = \boldsymbol{\beta}(\boldsymbol{u}) = \boldsymbol{S}_u\boldsymbol{u} + \boldsymbol{d}_u. \tag{3}$$

Most studies have considered either additive or multiplicative FDIA on control commands or observables. For example,

in [6], a multiplicative attack on the control command with a scalar and constant attack parameter β was considered, i.e., $\boldsymbol{S}_u = \beta \boldsymbol{I}^{m\times m}, \boldsymbol{d}_u = 0, \boldsymbol{S}_x = \boldsymbol{I}^{n\times n}, \boldsymbol{d}_x = 0$. Similarly, [9] studied both multiplicative and additive data injections, but the observables remained uncompromised. Representing FDIAs in the form of affine transformations with (2) and (3) allows for more generalized analyses. While this paper assumes constants for $\boldsymbol{S}_u, \boldsymbol{d}_u, \boldsymbol{S}_x,$ and $\boldsymbol{d}_x$ to satisfy the conditions presented in Section III-A, time-variant attacks will be considered in future publications. In comparison to time varying covert attacks, the linear affine transformation FDIA is simple for attackers to implement since it is a static attack, giving this form of attack greater significance.

This paper considers FDIA being undetectable when the user on the controller side is not able to distinguish the difference between the observed and predicted output estimated based on the nominal plant dynamics with the same control policy. The actual plant trajectory is altered by a control command attack. However, when the controller still observes the unaltered trajectory due to an observation attack, an attack detector in the controller may fail to detect the attack. An intelligent attacker coordinates both the control command attack and observation attack to make the attack perfectly undetectable. This definition is not in contradiction with (10) or those given in the literature [4], [16], [17].

Definition 1 (Perfectly undetectable FDIA from the controller's perspective): Let $\boldsymbol{x}(t, \boldsymbol{x}(0), u, \boldsymbol{\alpha}, \boldsymbol{\beta})$ denotes the state variables of a dynamic plant with the initial condition $\boldsymbol{x}(0)$, state-feedback controller $\boldsymbol{u} = \boldsymbol{k}(\boldsymbol{x})$, and FDIA attack functions $\boldsymbol{\alpha}(\boldsymbol{x})$,$\boldsymbol{\beta}(\boldsymbol{u})$. An FDIA is perfectly undetectable by the controller if:

$$\boldsymbol{x}(t, \boldsymbol{x}(0), \boldsymbol{u}, \boldsymbol{i_d}, \boldsymbol{i_d}) = \tilde{\boldsymbol{x}}(t, \boldsymbol{x}(0), \tilde{\boldsymbol{u}}, \boldsymbol{\alpha}, \boldsymbol{\beta}), \tag{4}$$

where $\tilde{\boldsymbol{x}} = \boldsymbol{\alpha}(\boldsymbol{x})$, $\tilde{\boldsymbol{u}} = \boldsymbol{\beta}(\boldsymbol{u})$, and $\boldsymbol{i_d}$ is the identify function (representing no attack).

Remark 1 (Undetectable affine transformation-based FDIAs). FDIAs illustrated in Fig. 2 are defined to be perfectly undetectable if: $\boldsymbol{q}(t, \boldsymbol{x}(0), \boldsymbol{u}, \boldsymbol{i_d}, \boldsymbol{i_d}, \boldsymbol{i_d}, \boldsymbol{i_d}) = \tilde{\boldsymbol{q}}(t, \boldsymbol{x}(0), \tilde{\boldsymbol{u}}, \boldsymbol{S}_x, \boldsymbol{d}_x, \boldsymbol{S}_u, \boldsymbol{d}_u)$ where $\boldsymbol{q} = \boldsymbol{i_d}(\boldsymbol{q})$, $\boldsymbol{u} = \boldsymbol{i_d}(\boldsymbol{u})$ are the identify functions.

It should be mentioned that simultaneous FDIA on the commands and observables is not necessarily a new concept. Papers on covert attacks in the literature introduced a similar structure in which the attacker implements an additional dynamic controller between the commands and observables as a stealthy attack [11]. In contrast, this paper formulates perfectly undetectable FDIAs in terms of affine transformations, as described in next section. Additionally, undetectable attacks may be implemented by using transmission zeros; however, specific manipulator dynamics in the following Section do non have zero dynamics as they are non-redundant [17]. For these reasons, this paper provides a simpler form of attack compared to covert and zero-dynamic undetectable attacks. Also, note that prior work assumed $\boldsymbol{x}(0) = 0$ for linearity assumptions. However, this paper derives the conditions including the one to match the observation of the initial condition for undetectable attacks. For the manipulator kinematic control, we do not assume $\boldsymbol{q}(0) = 0$ since $\tilde{\boldsymbol{q}}(0) = \boldsymbol{S}_x\boldsymbol{q}(0) + \boldsymbol{d}_x$ does not always hold.

Remark 2 (Exposure of the controller information and the desired trajectory to the attacker). Equation (4) may be achieved by the attacker without the knowledge of controller $\boldsymbol{k}$. Indeed, typical attack detector (such as an state observer or an adaptive algorithm [6]) attempts to detect an attack based on the observed dynamic relationship from $\boldsymbol{u}$ to $\tilde{\boldsymbol{x}}$, where the control scheme is commonly outside of the procedure. In other words, if an FDIA is implemented such that the observed dynamics of an attacked plant, the RHS of (4), is equivalent to the dynamics of the nominal plant, the LHS of (4), the attack cannot be detected regardless of the feedback control scheme and the desired trajectory. While this cannot be easily extended to general control systems, certain dynamic plants, including the one shown in Fig. 2, enable such FDIAs.

III. Realization of perfectly undetectable FDIA

A. Manipulator kinematic control and its attackability

Recall the manipulator nominal dynamics (i.e., $\tilde{\boldsymbol{u}} = \boldsymbol{u}$) with joint velocity control,

$$\dot{\boldsymbol{q}} = \boldsymbol{u}, \tag{5}$$

one would observe that, in many industrial motion control systems, the control of joint angles is fully decoupled across individual joints and exhibits linear (first-order) time-invariant dynamics. Control commands are represented by $\dot{\boldsymbol{q}} = \tilde{\boldsymbol{u}} = \boldsymbol{S}_u\boldsymbol{u} + \boldsymbol{d}_u$ and plant observables by $\tilde{\boldsymbol{q}} = \boldsymbol{S}_x\boldsymbol{q} + \boldsymbol{d}_x$. One can differentiate the latter, and plug in the former to achieve the resultant, observed attacked plant dynamics,

$$\dot{\tilde{\boldsymbol{q}}} = \boldsymbol{S}_x\boldsymbol{S}_u\boldsymbol{u} + \boldsymbol{S}_x\boldsymbol{d}_u, \tag{6}$$

yielding the following theorem.

Theorem 1: Perfectly undetectable FDIA on manipulator control system. The vector fields defined by (5) and (6) are equivalent and, thus, resultant observables are indistinguishable including the initial condition regardless of the control scheme if the following conditions are satisfied:

- Condition 1: $\boldsymbol{S}_x\boldsymbol{S}_u = \boldsymbol{I}^{n\times n}$,
- Condition 2: $\boldsymbol{q}(0) = \boldsymbol{S}_x\boldsymbol{q}(0) + \boldsymbol{d}_x$,
- Condition 3: $\boldsymbol{d}_u = 0$.

Proof: Substituting Conditions 1 and 3 into (6) yields (5). Evaluating $\tilde{\boldsymbol{q}} = \boldsymbol{S}_x\boldsymbol{q} + \boldsymbol{d}_x$ at $t = 0$ that must be identical to the nominal initial condition $\boldsymbol{q}(0)$ yields Condition 2. Note that if a different dynamic plant is considered, these conditions will change accordingly. ■

Remark 3 (Attack matrix selection). The attacker can utilize a range of $\boldsymbol{S}_x$ and $\boldsymbol{S}_u$ combinations such as scaling, reflection, shear, and rotation to satisfy Condition 1. In other words, this particular plant model is *highly attackable.* The attacker needs to know the initial posture $\boldsymbol{q}(0)$ to satisfy Condition 2. Also, $\boldsymbol{d}_u = 0$ is required as the nominal dynamics do not include an offset term.

B. Undetectable FDIA against adaptive attack detector

In [6], a specific case with $\boldsymbol{S}_u = \beta \boldsymbol{I}^{m\times m}, \boldsymbol{d}_u = 0, \boldsymbol{S}_x = \boldsymbol{I}^{n\times n}, \boldsymbol{d}_x = 0$ was considered with a scalar, constant attack parameter $0 < \beta \leq 1$. Note that $\beta = 1$ indicates no FDIA or normal operation. The efficiency of the adaptive attack detector integrated into the Jacobian velocity control with projection operators, illustrated in Fig. 2, was demonstrated. The adaptive attack detector is represented as:

$$\dot{\hat{\boldsymbol{r}}} = \hat{\beta}\boldsymbol{J}(\tilde{\boldsymbol{q}})\boldsymbol{u} - k_1(\hat{\boldsymbol{r}} - f(\tilde{\boldsymbol{q}})), \tag{7}$$

$$\dot{\hat{\beta}} = -k_2 \boldsymbol{u}^T \boldsymbol{J}(\tilde{\boldsymbol{q}})^T(\hat{\boldsymbol{r}} - f(\tilde{\boldsymbol{q}})) - s(\hat{\beta} - P_\Gamma(\hat{\beta})), \tag{8}$$

where $\hat{\beta}$ is the estimation of β. P_Γ is a projection operator [18] to maintain $\hat{\beta}$ within $\Gamma = \{\hat{\beta} \in R \mid \varepsilon \leq \hat{\beta} \leq 1\}, \varepsilon > 0$ via a mapping function s. Similarly, P_Ω is another projection operator to maintain joint variables within specified displacement and velocity limits. $\hat{\boldsymbol{r}}$ monitors if the remote manipulator moves following the control command $\boldsymbol{u}$. Note that (7) and (8) already include the compromised observable $\tilde{\boldsymbol{q}}$. The proof of convergence of $\hat{\beta} \to \beta, t \to \infty$ was given by the Lyapunov function analysis [6].

In the FDIA discussed in this paper, the assumption that measurements of ground truth obvervables, or that $\boldsymbol{S}_x = \boldsymbol{I}^{n\times n}, \boldsymbol{d}_x = 0$, no longer holds. From Condition 1, $\boldsymbol{S}_x = 1/\beta \boldsymbol{I}^{n\times n}$ for $\boldsymbol{S}_u = \beta$. Assume $\boldsymbol{d}_x$ is chosen appropriately.

Theorem 2: Undetectable FDIA by adaptive attack detector. For the attack detector given (7) and (8), $\hat{\beta} \to 1, t \to \infty$, if $\boldsymbol{S}$ and $\boldsymbol{S}_u$ are chosen as $\boldsymbol{S}_x = 1/\beta$ and $\boldsymbol{S}_u = \beta$.

Proof: Let $\boldsymbol{r}' = f(\tilde{\boldsymbol{q}})$ be the end-effector position of the manipulator that the controller reconstructs from the compromised observables $\tilde{q}$. If $\boldsymbol{r}'$ is identical to that expected from nominal control commands $\boldsymbol{u}$, the attack detector is unable to estimate β. Let $\tilde{\boldsymbol{r}} = \hat{\boldsymbol{r}} - f(\tilde{\boldsymbol{q}})$. From (7), $\dot{\tilde{\boldsymbol{r}}} = \hat{\beta}\boldsymbol{J}(\tilde{\boldsymbol{q}})\boldsymbol{u} - k_1\tilde{\boldsymbol{r}}$.

$$\begin{aligned} \dot{\tilde{\boldsymbol{r}}} &= \dot{\hat{\boldsymbol{r}}} - \frac{\partial f(\tilde{\boldsymbol{q}})}{\partial \tilde{\boldsymbol{q}}}\frac{d\tilde{\boldsymbol{q}}}{dt} = \dot{\hat{\boldsymbol{r}}} - \boldsymbol{J}(\tilde{\boldsymbol{q}})\dot{\tilde{\boldsymbol{q}}} \\ &= \hat{\beta}\boldsymbol{J}(\tilde{\boldsymbol{q}})\boldsymbol{u} - k_1\tilde{\boldsymbol{r}} - \boldsymbol{J}(\tilde{\boldsymbol{q}})\boldsymbol{u} \\ &= (\hat{\beta} - 1)\boldsymbol{J}(\tilde{\boldsymbol{q}})\boldsymbol{u} - k_1\tilde{\boldsymbol{r}}. \end{aligned} \tag{9}$$

The last equation is derived as follows: $\beta\dot{\tilde{\boldsymbol{q}}} = \dot{\boldsymbol{q}}$ since $\beta\tilde{\boldsymbol{q}} = \boldsymbol{q}$. Also, $\dot{\boldsymbol{q}} = \tilde{\boldsymbol{u}} = \beta\boldsymbol{u}$. Therefore, $\dot{\tilde{\boldsymbol{q}}} = \dot{\boldsymbol{q}}/\beta = \beta/\beta\boldsymbol{u} = \boldsymbol{u}$. Note that (10) is a special case when $\beta = 1$ of Equation (18) in [6] (β to be estimated has vanished in the adaptive control law), leading to $\hat{\beta} \to 1, t \to \infty$ regardless of β. ■

Since $\dot{\boldsymbol{r}}' = \boldsymbol{J}(\tilde{\boldsymbol{q}})\dot{\tilde{\boldsymbol{q}}} = \boldsymbol{J}(\tilde{\boldsymbol{q}})\boldsymbol{u}$, the controller perceives that $\boldsymbol{r}'$ is realized by following the nominal dynamics despite the actual trajectory $\boldsymbol{r} \neq \boldsymbol{r}'$, enabling a perfectly undetectable FDIA. Note that (10) is derived for any $\boldsymbol{S}_x$ and $\boldsymbol{S}_u$ satisfying $\boldsymbol{S}_x\boldsymbol{S}_u = \boldsymbol{I}^{n\times n}$ (Theorem 1, Condition 1) without a loss of generality, indicating that residual-based attack detectors such as (8) are susceptible to perfectly undetectable FDIAs.

Other attack detectors such as watermarking and moving target methods are perfectly ineffective against undetectable FDIAs. Watermarking is ineffective due to the attack's independence with respect to the control input, and it depends on a changing control input. Moving target detection requires plant dynamics to be able to be modified by the user, whereas the authors assume the plant is unchanging due to plant dynamics relying specifically on the physics of the system.

IV. Experiments

The perfectly undetectable attack game to validate the presented results is executed on a 6 degree-of-freedom nonredundant maniplator (FANUC lR Mate 200iD/7L) through robot control software RoboDK with MATLAB API. The MATLAB program calculates control commands and RoboDK performs the physical robot TCP/IP connection and command execution. Since the purpose of the experiment is to provide proof of the undetectable FDIA structure, the attacks are implemented through the same computer that operates the controller, not from an actual 3rd malicious party over the network. First, a nominal, no attack trial is demonstrated. Secondly, the attack detector, (7) and (8) in III-B is implemented on a detectable attack to compare the following undetectable scenarios against. The objective of the attacker is to perform the following scenarios:

- Scenario 1: Scaling attack
- Scenario 2: Reflection attack
- Scenario 3: Shear attack,

all while remaining completely undetectable according to Conditions 1–3. A smiley face is desired to be drawn by the manipulator's end effector, with $\boldsymbol{q}(0) = [0, -10, 10, 0, 0, 0]^T$ degrees. The attacks are implemented at $t = 0$ seconds, but this does not effect the results as long Conditions 1-3 are satisfied. The time step is 1 ms which allows the authors to estimate the system as continuous.

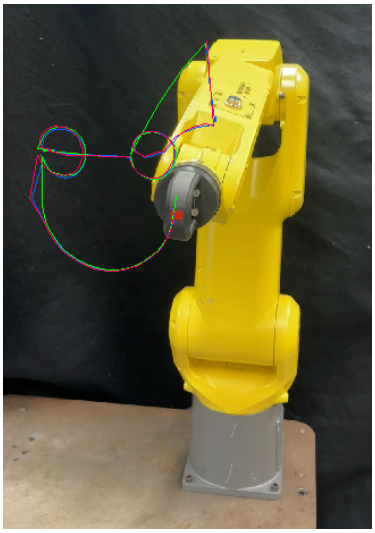
Nominal Trial: The nominal trial demonstrates the desired trajectory outcome when the system is not attacked. As such, $\beta = 1, \boldsymbol{S}_x = I^{6\times6}, \boldsymbol{S}_u = I^{6\times6}, \boldsymbol{d}_x = 0, \boldsymbol{d}_u = 0$ (no attack).

Detectable Attack: The chosen attack for the detectable trial is a scaling attack to the control input, where $\beta = 0.25, \boldsymbol{S}_x = \boldsymbol{I}^{6\times6}, \boldsymbol{S}_u = 0.25\boldsymbol{I}^{6\times6}, \boldsymbol{d}_x = [0, 30, -30, 0, 0, 0]^T, \boldsymbol{d}_u = 0$. Since this attack does not satisfy conditions 1–3, it is considered detectable.

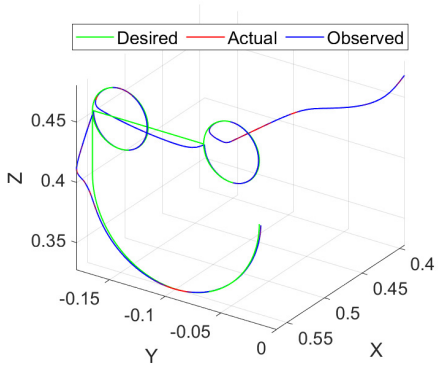
Scenario 1: The attacker in this scenario intends to implement a scaling attack on the manipulator. As such, the attacker chooses $\beta = 0.25, \boldsymbol{S}_x = 0.25\boldsymbol{I}^{6\times6}, \boldsymbol{S}_u = 4\boldsymbol{I}^{6\times6}, \boldsymbol{d}_x = [0, 30, -30, 0, 0, 0], \boldsymbol{d}_u = 0$. Conditions 1–3 are satisfied by these static variable choices.

Scenario 2: The attacker intends to reflect the manipulator trajectory. As such, the attacker chooses $\beta = -1, \boldsymbol{S}_x = -\boldsymbol{I}^{6\times6}, \boldsymbol{S}_u = -\boldsymbol{I}^{6\times6}, \boldsymbol{d}_x = [0, -20, 20, 0, 0, 0], \boldsymbol{d}_u = 0$. Conditions 1–3 are satisfied by these static variable choices. Note that $\beta = -1$ is outside the range of β defined by the attack detector in Section III-B so the attack detector will not be effective in this case; however, the desired, observed, and actual trajectories can still be analyzed to understand the effectiveness of the undetectable attack.

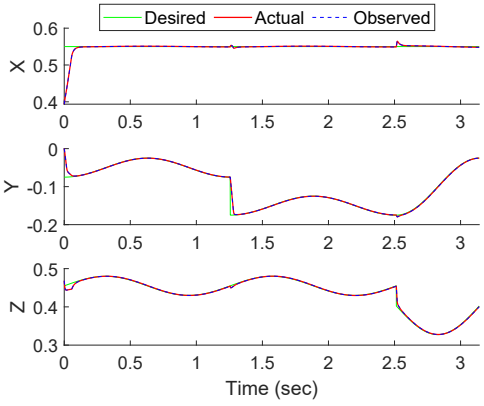
Scenario 3: Finally, the attacker intends to shear the manipulator trajectory. By shearing the trajectory, the attacker is manipulating joint values by varying quantities instead of all by the same quantity. $\boldsymbol{S}_x$ is a 6×6 matrix with ones across the diagonal and the first super diagonal and $\boldsymbol{S}_u$ is the inverse of $\boldsymbol{S}_x$. $\boldsymbol{d}_x = [-20, 10, 0, 0, 0, 0], \boldsymbol{d}_u = 0$. Conditions 1–3 are once again satisfied by these static variable choices. A test of attacker $q(0)$ mismatch is performed to understand the effects of utilizing incorrect initial conditions in the attack.



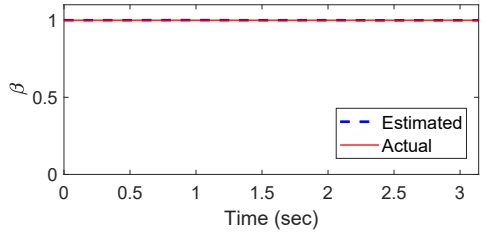
(a) Nominal FANUC Results (See below for legend)



(b) 3D Trajectory (Meters)

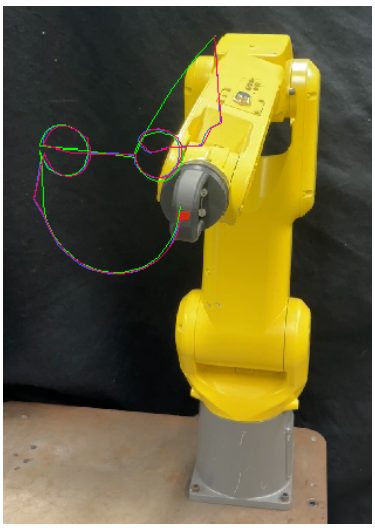


(c) End Effector Position (Meters)

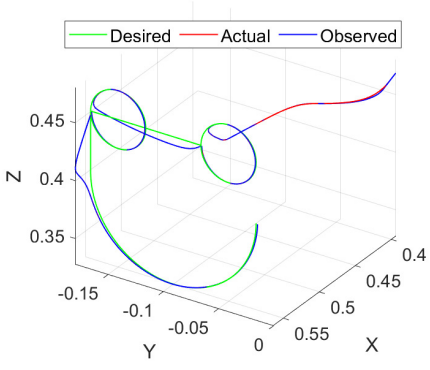


(d) Beta Estimation

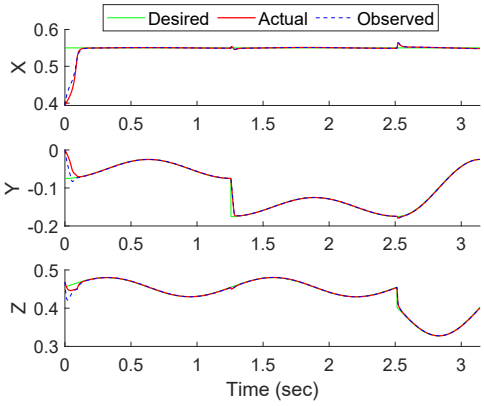
Fig. 3: Nominal Trial Results



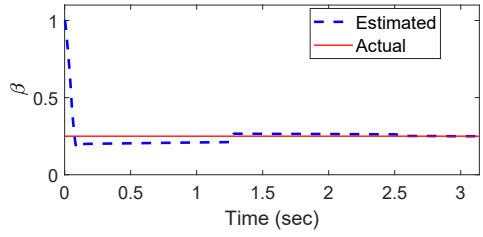
(a) Detectable Attack FANUC Run (See below for legend)



(b) 3D Trajectory (Meters)

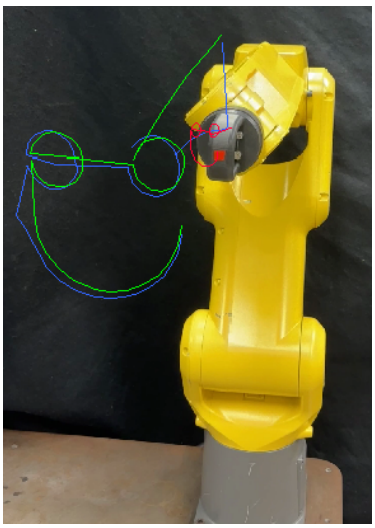


(c) End Effector Position (Meters)

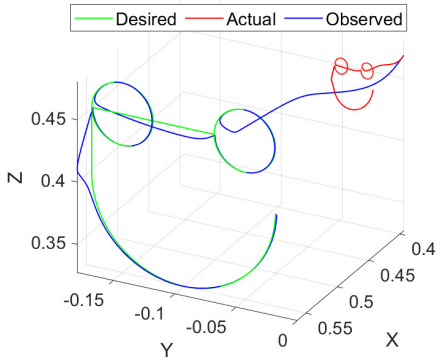


(d) Beta Estimation

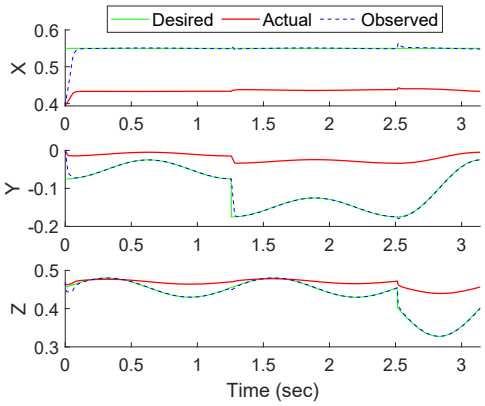
Fig. 4: Detectable Attack Results



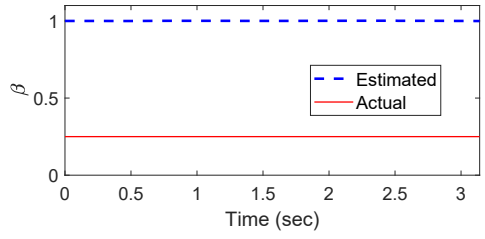
(a) Scenario 1 FANUC Results (See below for legend)



(b) 3D Trajectory (Meters)

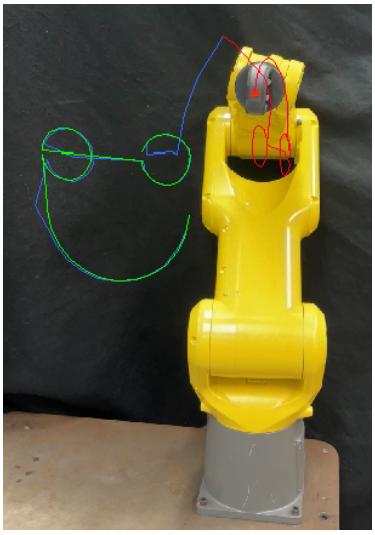


(c) End Effector Position (Meters)

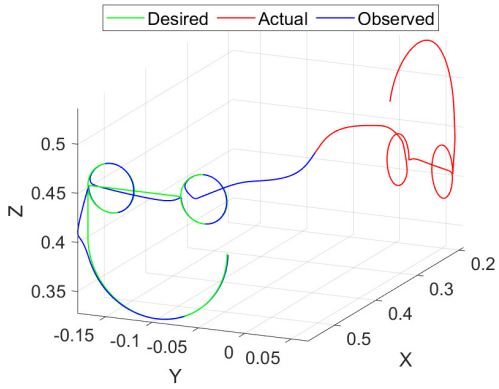


(d) Beta Estimation

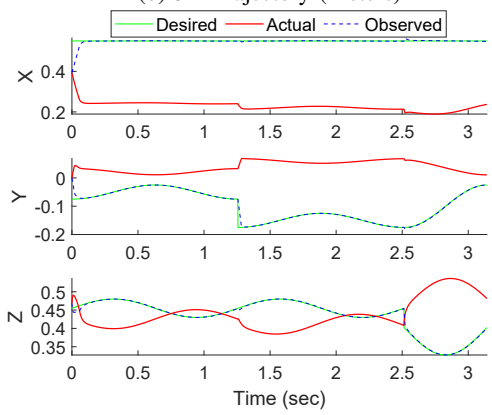
Fig. 5: Scenario 1: Scaling Attack Results



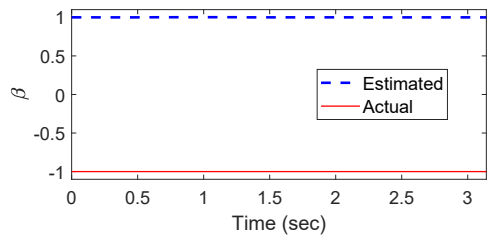
(a) Scenario 2 FANUC Run (See below for legend)



(b) 3D Trajectory (Meters)

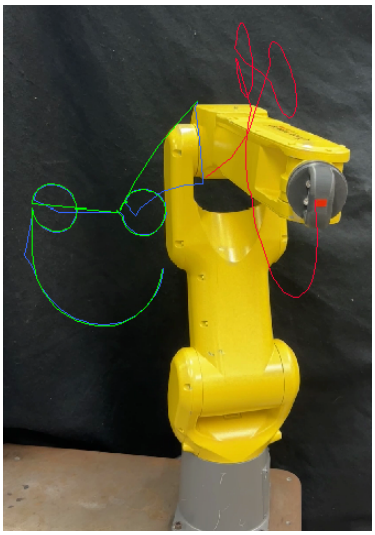


(c) End Effector Position (Meters)

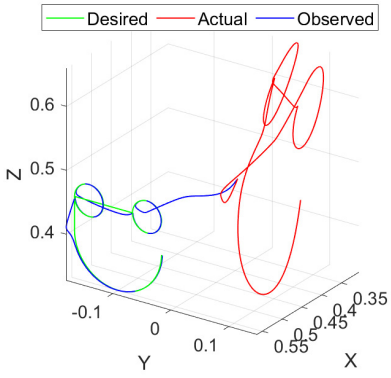


(d) Beta Estimation

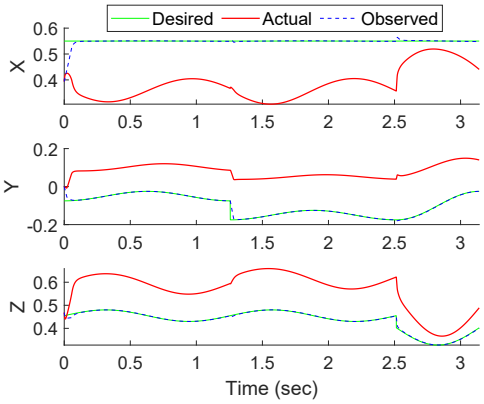
Fig. 6: Scenario 2: Reflection Attack Results



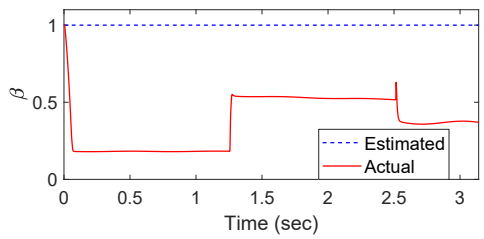
(a) Scenario 3 FANUC Run (See below for legend)



(b) 3D Trajectory (Meters)



(c) End Effector Position (Meters)



(d) Beta Estimation - Note: Actual β estimated by running detectable shear attack

Fig. 7: Scenario 3: Shear Attack Results

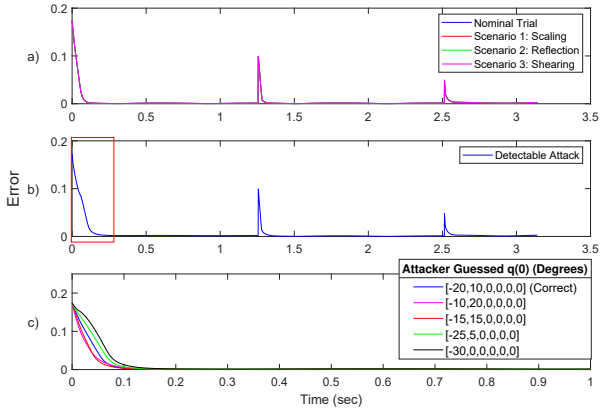


Fig. 8: End Effector Observed Position Errors (Meters)

V. Results and Discussions

Results are shown in Figs. 3-7. The nominal case demonstrates the smiley face drawn by the manipulator's end effector with no attack. As expected, the trajectory is accurately drawn via Jacobian transpose control with minimal error as demonstrated in Fig. 8a. The spikes in the error are the trajectories from the first eye to the other and from the second eye to the mouth. For the detectable case in Fig. 4, the attack detector described in Section III-B converges the $\hat{\beta}$ to its actual value as in Fig. 4d, and thus the manipulator draws the smiley face as desired even in the presence of an attack. Any robust control method such as a disturbance observer can be utilized to compensate this detectable attack. The average error in the nominal case is $0.0017\,m$, but the detectable case is a bit larger at $0.0023\,m$, as can be seen in Fig. 8b's red box. This larger initial error is due to the transient period of $\hat{\beta}$'s convergence.

For Scenarios 1-3, upon satisfaction of Conditions 1-3, with knowledge of initial conditions, the attacker can perform undetectable attacks to manipulate kinematic manipulator trajectories as in Figs. 5-7. According to the objective of the attacker, Scenarios 1-3 should have an identical observed error since the attacks are undetectable according to Remark 1, leaving the user observing seemingly perfect trajectory results. Figure 8a confirms this observation, as the observed errors agree with the nominal trial all at $0.0017\,m$ average error. Although the observed errors are demonstrating proper operating procedure, the experimental results in Figs. 5a, 6a, and 7a demonstrate the end effector drawing an unexpected smiley face. In Scenario 1, since $\beta = 0.25$, the actual trajectory is scaled to 25% of the desired trajectory. For Scenario 2, the actual trajectory is reflected since $\beta = -1$, and in Scenario 3, the actual trajectory is a sheared manipulation of the desired trajectory due to the shearing S_x and S_u matrices. Initial positions between the actual and desired trajectories are identical due to the choice of d_x according to Condition 1 for each Scenario. These results are as the attacker intended. The special case where the attacker mismatches the initial conditions, thus not satisfying Condition 2 is presented in Figure 8c for Scenario 3. In this case, 4 incorrect initial conditions are implemented in the attack. As one can see, the mismatch initially causes

noticeable error between the expected error and the actual errors in the range from $0.006m$ up to $0.016m$ during the transient period, but quickly flattens out to a mean error of $.000287m$ since the controller removes the disturbance. This initial condition mismatch is dangerous for the attacker since the observer could detect the attack during the transient period as Condition 2 is not satisfied. The attacker should ensure these manipulated trajectories remain within the manipulator's operating workspace and do not move into singularity so that no faults are triggered on the operator's side. These virtual attacks represent physical modification of hardware; shearing attacks are a rerouting of sensing and actuator wiring such that joints operate as if they are a different joint.

One of the primary requirements of the perfectly undetectable attack is that the plant, as seen by the controller, has a form of linear dynamics, i.e., individual joint velocity control. This is a common form of industrial robot control which is utilized in [6] and in this experimentation. Oftentimes, a user may apply methods to compensate for nonlinearities in plant dynamics at a local plant, e.g., feedback linearization or compute torque method at the plant [19] to simplify the remote controller. This architecture introduces a security hole by creating an easily attackable plant which may invite a variety of undetectable FDIAs. Note that the presence of nonlinearities in the plant does not fully prevent undetectable FDIAs; demonstration of perfectly undetectable attacks on nonlinear plants are currently being developed by the authors.

VI. Conclusion

This paper demonstrated that intelligent attackers can implement coordinated, perfectly undetectable FDIAs on control commands and observables in the form of affine transformations, allowing them to alter linear kinematic manipulator trajectories within their workspaces. Validation is provided through a nominal trial, a detectable case, and three Scenarios which show effective attacks on a manipulator. Future work includes extension to nonlinear systems and development of an attack detection method as the defender.

VII. Acknowledgements

The authors would like to thank Dr. Kiminao Kogiso, the University of Electro Communications, for valuable discussions, and Dr. Yinyan Zhang, Jinan University College of Cyber Security, for their assistance on robot kinematic control and attack detection programming.

Appendix I
Perfectly undetectable FDIA from the plant's perspective

Definition A1 (Perfectly undetectable FDIA from the plant's perspective) (Milosevic 2021 [4], [20]). Let $y(x(0), u, a)$ denote the response of the system for the initial condition $x(0)$, input $u(t)$, and attack signal $a(t)$. The attack is perfectly undetectable if

$$y(x(0), u, a) = y(x(0), u, 0), t \geq 0. \tag{10}$$

The attacker does not leave any traces in the measurements of y, and can impact the system's performance or behavior without being noticed by an attack detector that utilizes y for attack detection. Research showed that (10) can be achieved by zero dynamics attacks with the existence of transmission zeros [4], [16], [17]. In this definition, the detector receives ground truth observables without being compromised, i.e., $\boldsymbol{S}_x = \boldsymbol{I}^{n\times n}, \boldsymbol{d}_x = \boldsymbol{0}$, as a special case of Fig. 2.

References

[1] D. G. Pivoto, L. F. de Almeida, R. da Rosa Righi, J. J. Rodrigues, A. B. Lugli, and A. M. Alberti, "Cyber-physical systems architectures for industrial internet of things applications in industry 4.0: A literature review," *Journal of Manufacturing Systems*, vol. 58, 2021.

[2] W. Duo, M. Zhou, and A. Abusorrah, "A survey of cyber attacks on cyber physical systems: Recent advances and challenges," *IEEE/CAA Journal of Automatica Sinica*, vol. 9, no. 5, pp. 784–800, 2022.

[3] A. Botta, S. Rotbei, S. Zinno, and G. Ventre, "Cyber security of robots: A comprehensive survey," *Intelligent Systems with Applications*, vol. 18, p. 200237, 2023.

[4] H. Sandberg, V. Gupta, and K. H. Johansson, "Secure networked control systems," *Annual Review of Control, Robotics, and Autonomous Systems*, vol. 5, no. 1, pp. 445–464, 2022.

[5] K. Teranishi and K. Kogiso, "Control-theoretic approach to malleability cancellation by attacked signal normalization," *IFAC-PapersOnLine*, vol. 52, no. 20, pp. 297–302, 2019.

[6] Y. Zhang and S. Li, "Kinematic control of serial manipulators under false data injection attack," *IEEE/CAA Journal of Automatica Sinica*, vol. 10, no. 4, pp. 1009–1019, 2023.

[7] Y. Gu, K. Guo, C. Zhao, X. Yu, and L. Guo, "Fast reactive mechanism for desired trajectory attacks on unmanned aerial vehicles," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 8, 2023.

[8] F. Pasqualetti, F. Dörfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Transactions on Automatic Control*, vol. 58, no. 11, pp. 2715–2729, 2013.

[9] H. Zhu, L. Xu, Z. Bao, Y. Liu, L. Yin, W. Yao, C. Wu, and L. Wu, "Secure control against multiplicative and additive false data injection attacks," *IEEE Transactions on Industrial Cyber-Physical Systems*, 2023.

[10] S. Liu, G. Wei, Y. Song, and Y. Liu, "Extended kalman filtering for stochastic nonlinear systems with randomly occurring cyber attacks," *Neurocomputing*, vol. 207, pp. 708–716, 2016.

[11] C. Schellenberger and P. Zhang, "Detection of covert attacks on cyber-physical systems by extending the system dynamics with an auxiliary system," in *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, 2017, pp. 1374–1379.

[12] M. H. Manshaei, Q. Zhu, T. Alpcan, T. Bacşar, and J.-P. Hubaux, "Game theory meets network security and privacy," *ACM Comput. Surv.*, vol. 45, no. 3, jul 2013.

[13] S. Roy, C. Ellis, S. Shiva, D. Dasgupta, V. Shandilya, and Q. Wu, "A survey of game theory as applied to network security," in *2010 43rd Hawaii International Conference on System Sciences*, 2010, pp. 1–10.

[14] Q. Zhu and T. Başar, "Game-theoretic approach to feedback-driven multi-stage moving target defense," in *Decision and Game Theory for Security*. Springer International Publishing, 2013, pp. 246–263.

[15] J. Pawlick, E. Colbert, and Q. Zhu, "A game-theoretic taxonomy and survey of defensive deception for cybersecurity and privacy," *ACM Computing Surveys*, vol. 52, 12 2017.

[16] J. Milošević, A. Teixeira, K. H. Johansson, and H. Sandberg, "Actuator security indices based on perfect undetectability: Computation, robustness, and sensor placement," *IEEE Transactions on Automatic Control*, vol. 65, no. 9, pp. 3816–3831, 2020.

[17] Y. Mao, H. Jafarnejadsani, P. Zhao, E. Akyol, and N. Hovakimyan, "Novel stealthy attack and defense strategies for networked control systems," *IEEE Transactions on Automatic Control*, vol. 65, no. 9, pp. 3847–3862, 2020.

[18] G. Larchev, S. Campbell, and J. Kaneshige, "Projection operator: A step toward certification of adaptive controllers," *AIAA Infotech at Aerospace 2010*, 04 2010.

[19] J. J. Craig, *Introduction to Robotics: Mechanics and Control*, 4th ed. USA: Addison-Wesley Longman Publishing Co., Inc., 1989.

[20] S. Gracy, J. Milošević, and H. Sandberg, "Security index based on perfectly undetectable attacks: Graph-theoretic conditions," *Automatica*, vol. 134, p. 109925, 2021.