On the Detection of Deception Attacks and Resilience in Turn-Based Command Governors for Dynamically Decoupled Systems ⋆

Ayman El Qemmah * **Alessandro Casavola** *
Francesco Tedesco *

* *Dipartimento di Elettronica, Informatica e Sistemistica, Università degli Studi della Calabria, Via Pietro Bucci, Cubo 42-c, Rende (CS), 87036, Italy (e-mail:* {*elqemmah.ay,a.casavola,ftedesco*}*@dimes.unical.it).*

Abstract: The distributed Command Governor (CG) approach is an effective solution that can be employed to supervise and coordinate a set of dynamically decoupled systems, connected via data links, to ensure local and global constraints satisfaction. Here the problem of attack detection is investigated under the assumption that the agents are subject to deception attacks. In this type of cyber-attack, typically an adversary can inject data into the communication link between agents, replacing legitimate commands with an altered version. This can render coupling constraints with neighbors infeasible, leading to constraint violations. Under mild assumptions, sufficient conditions for identifying attack-free scenarios are introduced. Building on this analysis, a decentralized CG scheme that adapts coupling constraints to account for estimation uncertainties, is proposed. Eventually, the effectiveness of the proposed supervision architecture is assessed via a simulative example involving a set of autonomous vehicles under local and connectivity keeping constraints.

Keywords: Distributed Command Governor, Deception Attack Detection, Multi-Agent Systems, Noncompliant Agents, Resilient Distributed Command Governor.

1. INTRODUCTION

Command Governors (CG) (**?**) are effective tools for enforcing constraints on pre-compensated systems, particularly in industrial applications where peripheral constraint handling is required. Early CG-based formulations have been expressed in a centralized formulation, with a single supervisory unit solving the optimal control problem for a set of systems. More recently, to address the limitations posed by the use of a centralized unit in complex networked systems, recent research has focused on distributed approaches (**??**). These methods involve distributing the optimal control problem among network agents, each solving a local optimization problem. Nonetheless, network-enabled data sharing and enhanced connectivity introduce multiple vulnerabilities for malicious actors, resulting in repercussions that vary from performance degradation to instability.

Among the most common strategies employed to compromise the communication network are denial-of-service (DoS) attacks and deception attacks. In DoS attacks (**?**), the assailant makes the data inaccessible, potentially by disrupting the wireless network. These attacks can be executed by an adversary who compromises the availability of data, for example, by jamming the communication channel. Deception attacks, which involve data integrity, are more insidious. The adversary injects false data into the networked system by either substituting or altering the authentic data, thus making resilient supervision critically important. Detection necessitates the capacity to identify anomalous behaviors, which is crucial for implementing suitable mitigation strategies and ensuring global viability.

While several approaches have been proposed to address deception attacks in predictive control strategies, most assume dynamically coupled systems and exploit inter-agent feedback interconnections to detect noncompliant behaviors (**?**). In contrast, the setting considered in this work involves dynamically decoupled agents, where such interconnections are absent. Although some methods have been developed for dynamically decoupled systems, these typically rely on cooperative strategies, leveraging abnormalities in local cost functions or negotiation processes to infer the presence of malicious behaviors (**?**).

In this paper, we focus instead on a non-cooperative distributed framework, specifically the Turn-Based CG (TB-CG) supervision architecture. Within this context, we analyze the vulnerability of a distributed CG scheme to deception attacks and develop a detection strategy that enables agents to identify tampered admissible commands.

⋆ This work has been partially supported by the research project - ID:20225MESZB "Resilient Optimization in Distributed Cyber-Physical Control Problems subject to Adversarial Behaviour" granted by the Italian Ministry of University and Research (MUR), within the PRIN 2022 program, and European Union - Next Generation EU.

Furthermore, a robust decentralized supervision approach is proposed, suitable for scenarios in which the communication channel is unreliable or cannot be fully trusted, and is therefore not used as a critical component of the supervision strategy. An illustrative example of vehicles operating in a 2D space subject to connectivity keeping constraints is presented to assess the effectiveness of the proposed approach.

2. PRELIMINARIES

2.1 System Description

Consider $N \in \mathbb{N}$ dynamically decoupled agents, with each i-th subsystem modeled as a linear time-invariant discrete-time dynamical system

$$\Sigma_i : \begin{cases} x_i(t+1) &= \Phi_i x_i(t) + G_i g_i(t), \\ y_i(t) &= H_i^y x_i(t), \\ c_i(t) &= H_i^c x_i(t) + L_i g_i(t), \end{cases} \quad (1)$$

where: $t \in \mathbb{Z}_+$, $x_i(t) \in \mathbb{R}^{n_i}$ is the state vector (which includes the controller states under dynamic regulation), $g_i(t) \in \mathbb{R}^{p_i}$ is the manipulable command vector, which, in the absence of constraints (and no CG), would coincide with the desired reference $r_i(t) \in \mathbb{R}^{p_i}$, assuming each system is regulated by a local controller that ensures stability and satisfactory closed-loop performance, i.e. Φ_i are strictly *Schur* matrices. Furthermore, $y_i(t) \in \mathbb{R}^{p_i}$ is the output vector related to the tracking performance ($y_i(t) \approx r_i(t)$), and $c_i(t) \in \mathbb{R}^{z_i}$ is the constrained vector that collects all constrained local variables. From a global perspective, to describe (coupling) constraints among states of different subsystems, the vector c_i in (1) is allowed to depend on the aggregate state and manipulable commands vectors $x = [x_1^T, \dots, x_N^T]^T \in \mathbb{R}^n$, with $n = \sum_{i=1}^N n_i$, and $g = [g_1^T, \dots, g_N^T]^T \in \mathbb{R}^p$, with $p = \sum_{i=1}^N p_i$. To ensure constraints satisfaction, the global constraints vector $c(t) \in \mathbb{R}^z$, with $z = \sum_{i=1}^N z_i$, needs to satisfy, $\forall t \in \mathbb{Z}_+$, the following set membership condition

$$c(t) \in \mathcal{C},\ \mathcal{C} := \{c \in \mathbb{R}^z : f_h(c) \le 0,\ h = 1, \dots, n_q\}, \quad (2)$$

where the compact and convex set $\mathcal{C}$ is described by a list of n_q inequalities. Considering that a centralized approach (**?**) poses challenges regarding scalability and, critically, renders the system vulnerable to a single point of failure, more recent CG-based strategies have focused on a distributed formulation, where it is assumed that neighboring agents are capable to share information each other in a network. In view of this, the above constraints structure can be associated to an undirected connected graph $\Gamma(\mathcal{A}, \mathcal{E})$, whose set of nodes coincides with the set of agents $\mathcal{A}$ and where

$$\mathcal{E} := \{(i,j) : f_h \text{ in (2) involves both } i \text{ and } j \text{ for at least } h \in \{1, \dots, n_q\}\}. \quad (3)$$

Then, from the perspective of the i-th agent, the quantity $c_{[i]} := [c_i^T, c_{i_1}^T, c_{i_2}^T, \dots, c_{i_h}^T, \dots, c_{i_{N_i}}^T]^T \in \mathbb{R}^{z_{[i]}+1}$, with $z_{[i]} := \sum_{i \in \mathcal{N}_i} z_i$, which collects all local constrained variables, becomes relevant. In a similar way, let us recast each function f_h in (2) involving agent i and its neighbors only as $f_{[i]}^K : \mathbb{R}^{z_{[i]}+1} \to \mathbb{R}$. Then, the constraint set (2) can be seen as a combination of compact and convex sets defined as

$$\mathcal{C}_{[i]} := \{c_{[i]} \in \mathbb{R}^{z_{[i]}+1} : f_{[i]}^h(c_{[i]}) \le 0,\ h = 1, \dots, n_{q_i}\}. \quad (4)$$

Then, the distributed CG design problem can be stated as

Problem *1. Given systems* (1) *and constraints* (2), ***locally*** *determine, at each time step t and for each agent $i \in \mathcal{A}$, a suitable command signal $g_i(t)$ that is the best approximation of $r_i(t)$ and such that its application does not lead to the violation of constraints, i.e. $c(t) \in \mathcal{C}$, $\forall t \in \mathbb{Z}_+$.*

2.2 Turn-Based Distributed Command Governor

In **?**, the above stated **Problem** *1* has been solved by grouping the agents into L *turns*, where a turn $\mathcal{T} \subset \mathcal{A}$ is a *subset of non-neighboring nodes*, i.e. $\forall i, j \in \mathcal{T}$ such that $i \neq j$, $j \notin \mathcal{N}_i$ and $\bigcup_{l=1}^L \mathcal{T}_l = \mathcal{A}$. In order to introduce the TB-CG solution, it is useful to define both the steady-state of the constrained variable c_i as $c_{g_i} := H_i^c (I_{n_i} - \Phi_i)^{-1} G_i g_i$ and the virtual evolutions of $c_i(k, x_i, g_i)$ along the virtual time k, starting from the initial state x_i, as

$$c_i(k, x_i, g_i) := H_i^c (\Phi_i^k x_i + \left(\sum_{\tau=0}^{k-1} \Phi_i^{k-\tau-1} G_i g_i \right)) + L_i g_i, \quad (5)$$

when a constant command sequence $g_i(t) \equiv g_i$ is applied to Σ_i starting at time $k = 0$. The TB-CG algorithm operates such that at any given time t, only the agents belonging to the active turn, $\mathcal{T}_{l(t)}$, are allowed to update their local commands $g_i(t)$. The remaining agents, those in $\mathcal{A} \setminus \mathcal{T}_{l(t)}$, have to keep applying the local command they have applied at time $t-1$. Notice that for a static configuration of the network the turns $\mathcal{T}_l$ are *L-periodic*, i.e. $\mathcal{T}_{l(t+L)} = \mathcal{T}_{l(t)}$. For this reason, after each agent in $\mathcal{T}_{l(t)}$ received from its neighbors the values of their states $x_{[i]}(t)$ and of their previously applied commands $g_{[i]}(t-1)$, the global constraints could be satisfied if the local command $g_i(t)$ were selected as

$$\begin{aligned} g_i(t) &= \arg\min_{g_i} ||g_i - r_i(t)||_{\Psi_i}^2 \\ &\text{s.t. } g_i(t) \in \mathcal{V}_i(\bar{x}_i(t)), \\ \mathcal{V}_i(\bar{x}_i(t)) &:= \left\{ g_i \in \mathbb{R}^{p_i} \ : \ \begin{matrix} c_{[i]}(k, \bar{x}_i(t), \bar{g}_i(t)) \in \mathcal{C}_{[i]}, \\ \forall k \in \mathbb{Z}_+ \end{matrix} \right\}, \end{aligned} \quad (6)$$

where the quantities $\bar{x}_i \in \mathbb{R}^{n_{[i]}+1}$ and $\bar{g}_i \in \mathbb{R}^{n_{[i]}+1}$, collect, respectively, all states and commands of agent i and its neighbors. As demonstrated in **?**, instead of ensuring constraints satisfaction for all $k \ge 0$, it suffices to introduce steady-state constraints with a margin $\delta > 0$ and evaluate virtual predictions within a specific finite *admissibility index* k_0. As a result,

$$\mathcal{V}_i(\bar{x}_i(t)) := \left\{ g_i \in \mathcal{W}_i^\delta \ : \ \begin{matrix} c_{[i]}(k, \bar{x}_i(t), \bar{g}_i(t)) \in \mathcal{C}_{[i]}, \\ k = 0, \dots, k_0 \end{matrix} \right\}, \quad (7)$$

with

$$\mathcal{W}_i^\delta := \{g_i \in \mathbb{R}^{p_i} \ : \ c_{\bar{g}_i} \in \mathcal{C}_{[i]} \sim \mathcal{S}_\delta\} \quad (8)$$

representing the set of steady-state admissible references, where $\mathcal{S}_\delta := \{x \in \mathbb{R}^n : \|x\| \le \delta\}$ is a ball with radius δ. The parameter δ influences the finite determinateness of $\mathcal{V}_i(\bar{x}_i(t))$, which in turn determines the value of the integer k_0. An inverse relationship exists, such that decreasing δ leads to an increased value of k_0.

3. PROBLEM FORMULATION

Given the reliance of multi-agent systems on communication networks for inter-agent information exchange, external entities may compromise data integrity. We consider scenarios in which an attacker manipulates the admissible commands exchanged between agents, potentially causing

neighbors to deviate from nominal behavior and appear noncompliant. It is important to note that, in general, such non-compliant behavior may arise either from malicious agents or from external false data injection. While a malicious agent typically seeks to influence the behavior of other agents for its own benefit, deception attacks are designed to induce constraint violations and overall system disruption, making them particularly insidious. Under potential deception attacks, false data are injected into the communication channel. Specifically, when admissible commands are targeted

Assumption *1.* The attackers can modify and intercept data in transit (**?**), and each agent Σ_i can potentially receive, from at least one neighboring agent $j \in \mathcal{N}_i$,

$$\bar{g}_j(t) = g_j(t) + \Delta g_j(t). \tag{9}$$

Furthermore, we assume the following regarding the capabilities of the attacker

Assumption *2.* The attackers are assumed to possess knowledge on the operation of the system (**?**), and can inject data such that the altered commands remain admissible with respect to coupling constraints (4).

The previous assumption is motivated by the attacker's intent to remain undetected. Injecting false data that leads to non-admissible commands could trigger a simple admissibility check and reveal the tampering. Specifically, since the computation of the command by agent j relies on information from agent i, a violation of coupling constraints based on this data would indicate either an altered received command or a corrupted command originating from agent i. In both cases, the attack would no longer be purely deceptive, undermining the attacker's objective. For this reason, we assume an attacker with strategic behavior. While this assumption may appear strong, it is not unrealistic in practice (**?**). It adheres to Kerckhoffs' principle (**?**), which states that security should not depend on secrecy. Analyzing system behavior under such "worst-case" attacks provides a useful benchmark. Accordingly, a considerable number of studies in the literature operate under the assumption of complete attacker knowledge, e.g. **??**. Moving from the aforementioned considerations, the problem to be addressed in this distributed setting can be formally stated as follows

Problem *2. Design a detection policy such that, under Assumptions 1 and 2, at each time step t, each agent $i \in \mathcal{A}$ can determine whether the admissible commands received from its neighbors have been compromised.*

4. MULTI-LAYER DISTRIBUTED SUPERVISION ARCHITECTURE FOR ATTACK DETECTION

In this section, the main goal is to provide valuable insights into attack detection in the context of TB-CG distributed supervision scheme to be able to formally state a solution to solve **Problem** *2*. Such a policy can serve as the basis for defining a resilient distributed algorithm. In particular, it can facilitate the design of robust strategies that preserve overall system feasibility under deception, as well as mitigation strategies that isolate affected agents. Given that agents are assumed to be dynamically decoupled in this work, methodologies akin to those detailed in **?**, which leverage the inherent coupling in the system dynamics to detect noncompliant behaviors, are not directly applicable. While some approaches address attack detection in decoupled systems (**?**), these typically rely on cooperative strategies, where detection could be facilitated by observing anomalous patterns in the negotiation process. However, focusing on the above introduced TB-CG and considering its non-cooperative nature, in the absence of direct state estimation/measurement, distinguishing between normally operating agents and those under attack becomes inherently challenging, especially under Assumption 2, which ensures the constraint-admissibility of altered commands. Therefore, to address this challenge, we assume implicit communications, i.e. the availability of onboard sensors enabling state measurement of the other agents (**?**). This capability is a common feature in many application, such as in autonomous vehicle applications, and enables detection of non-compliant agents whose behavior deviates from expected dynamics.

Assumption *3.* At each time instant t, every agent owns the capability to measure the state $x_j(t)$ of all its neighboring agents $\mathcal{N}_i$.

In cases where the state is not fully available and it has to be reconstructed from a noisy output, an approach using a Luenberger observer, as proposed in **?**, can be followed. Given the adoption of a TB-CG strategy, the admissible commands of all relevant neighboring agents of agent i are held constant. This property allows agent i, when its turn occurs, to verify the integrity of the received admissible command. By analyzing the state of system at the previous and current time steps, a detection module pertaining to agent i can reveal any inconsistency between the received command and the actual system evolution. More formally

Proposition *1.* Under Assumption 3, agent i can determine the following quantity

$$\tilde{w}_j^i(t) = x_j(t) - \Phi_j x_j(t-1), \tag{10}$$

based on the measured information $x_j(t)$ and $x_j(t-1)$ of a neighboring agent j.

4.1 Ideal case

In the ideal case, the computed value $\tilde{w}_j^i(t)$ using (10) would coincide with the real value applied by agent j, i.e. $\tilde{w}_j^i(t) = G_j g_j(t-1)$ and, if the received command $\bar{g}_j(t)$ is unaltered, then

$$\underbrace{x_j(t) - \Phi_j x_j(t-1)}_{\tilde{w}_j^i(t)} - G_j \bar{g}_j(t) = 0. \tag{11}$$

Then, any noncompliance with respect to the received values $\bar{g}_j(t)$ can be detected if

$$\|\tilde{w}_j^i(t) - G_j \bar{g}_j(t)\| > 0, \forall j \in \mathcal{N}_i. \tag{12}$$

Consequently, a *Deception Detection Module* (DDM) may compute the set $\tilde{\mathcal{N}}_i(t) \subseteq \mathcal{N}_i$ of neighbors satisfying the detection condition. Then, to compensate the effects of the attack, the injected data can be computed as

$$G_j \Delta g_j(t) = \tilde{w}_j^i(t) - G_j \bar{g}_j(t), \forall j \in \tilde{\mathcal{N}}_i(t). \tag{13}$$

To induce resilience, a higher-level module can be introduced within the supervision architecture to construct the filtered admissible vector $g_{[i]}(t-1)$.

Remark 1. Although the estimated command-based quantities $\bar{w}_{[i]}$ would suffice in this ideal scenario, the proposed detection policy enables adaptation of the supervision strategy to broader cases. For instance, it can be used to detect malicious or faulty agents, enabling their isolation and removal from the network (**?**).

4.2 Non-Ideal case

In a more realistic setting, due to possible model uncertainties, $\tilde{w}_j^i(t) \approx G_j g_j(t-1)$, $\forall j \in \mathcal{N}_i$. In this case $\tilde{w}_j^i(t)$

can be seen as contained in a ball centered in the real (unknown) admissible command $G_j g_j(t-1)$, i.e.

$$\tilde{w}_j^i(t) \in \mathcal{O}(G_j g_j(t-1)), \tag{14}$$

with $\mathcal{O}(G_j g_j) := \{x \in \mathbb{R}^{n_i} : \|G_j g_j - x\| \leq r\}$, with $r \in \mathbb{R}_{>0}$. Moreover, suppose that

Assumption *4.* The magnitude of the estimation uncertainty is strictly smaller than that of any possible injected packet, i.e. there exists a vector $\Delta g_{\min} \in \mathbb{R}^{p_i}$ such that

$$\begin{aligned}\|\bar{g}_j(t) - g_j(t-1)\| > \|\Delta g_{\min}\|, \forall \Delta g_j(t) \in \mathbb{R}^{p_i} \;\Rightarrow \\ G_j \bar{g}_j(t) \notin \mathcal{O}(G_j g_j(t-1)),\ \forall t \in \mathbb{Z}_+.\end{aligned} \tag{15}$$

Remark 2. Although Assumption 4 may appear restrictive, it is necessary to ensure the detectability of deception attacks. If this assumption does not hold, some attacks may remain undetected, as their effects become indistinguishable from estimation uncertainty. These scenarios give rise to *stealthy attacks*, where injected deviations are small enough to evade detection (**?**).

Then, analogously to the ideal case, based solely on the deviation of the predicted state from the actual state using the received admissible command $\bar{g}_j(t)$, it is possible to detect any noncompliance if

$$\|\tilde{w}_j^i(t) - G_j \bar{g}_j(t)\| > \delta_{j,\mathrm{s}}, \forall j \in \mathcal{N}_i, \tag{16}$$

with $\delta_{j,\mathrm{s}} > r$. However, contrary to the ideal case, if (16) is not satisfied, it does not imply an attack-free scenario. In view of these considerations, it is useful to introduce the following result

Lemma *1.* The following condition

$$\|x_j(t) - \Phi_j x_j(t-1) - G_j \bar{g}_j(t)\| \leq \delta_{j,\mathrm{s}}, \forall j \in \mathcal{N}_i, \tag{17}$$

is necessary but not sufficient to rule out attack scenarios.

Proof. Consider a time instant $t' \in \mathbb{Z}_{>0}$ where agent i receives $\bar{g}_j(t')$ from agent $j \in \mathcal{N}_i$. Assume $G_j \bar{g}_j = \tilde{w}_j^i + G_j \tilde{g}_j^\epsilon$, with $\|\tilde{g}_j^\epsilon\| \leq \delta_{j,s}$ and that $\tilde{w}_j^i(t') \neq G_j g_j(t'-1)$, where $g_j(t'-1)$ is the true admissible command. Based on (10), it would result that

$$\|\tilde{w}_j^i(t') - G_j \bar{g}_j(t')\| \leq \delta_{j,s} \tag{18}$$

is satisfied, and thus the altered admissible command $\bar{g}_j(t')$ is directly used by agent i. □

Remark 3. A direct consequence of the latter results is that under some attack occurrences, the fulfillment of the condition (17) does not guarantee constraints satisfaction. Let us consider the case where the true command $g_j(t'-1)$ saturates the coupling constraints, i.e., there exists a prediction step $\bar{k} \in \{0, \dots, k_0\}$ such that

$$c_i^c(k, \bar{x}_i(t'), g_i, g_{[i]}(t'-1)) \in \partial \mathcal{C}_{[i]}^c,\ k = \bar{k}, \dots, k_0, \tag{19}$$

where $g_{[i]}(t'-1) = [g_{i_1}(t'-1), \dots, g_j(t'-1), \dots, g_{i_{\mathcal{N}_i}}(t'-1)]^T$ for some admissible commands $g_i \in \mathcal{V}_i(\bar{x}_i(t'))$. Suppose that the $\hat{g}_{[i]}(t'-1) = [g_{i_1}(t'-1), \dots, \bar{g}_j(t'), \dots, g_{i_{\mathcal{N}_i}}(t'-1)]^T$, built using the altered admissible command, is such that it artificially relaxes the constraint, i.e. it satisfies

$$c_i^c(k, \bar{x}_i(t'), g_i, \hat{g}_{[i]}(t'-1)) \in \mathrm{Int}(\mathcal{C}_{[i]}^c), \tag{20}$$

for the same commands g_i, with $k = 0, \dots, k_0$. Agent i then computes its command $g_i(t')$ by solving the local optimization problem (6) assuming that the received commands $\hat{g}_{[i]}(t'-1)$ are reliable. Due to the non-cooperative nature of the TB-CG scheme, it may select a $g_i(t')$ such that

$$c_i^c(k, \bar{x}_i(t'), g_i(t'), \hat{g}_{[i]}(t'-1)) \in \partial \mathcal{C}_{[i]}^c, \tag{21}$$

i.e. the agent may be misled into saturating artificially loosened constraints. However, since agent j actually applies $g_j(t'-1)$, the true evolution of the constrained variable may result in

$$c_i^c(k, \bar{x}_i(t'), g_i(t'), g_{[i]}(t'-1)) \notin \mathcal{C}_{[i]}^c,\ k = 0, \dots, k_0, \tag{22}$$

i.e., a violation of the coupling constraints. This confirms that the condition (17) is not sufficient to guarantee attack-free scenarios.

In order to derive sufficient conditions for identifying attack-free scenarios, it is necessary to characterize the dynamics of the estimation error (Proposition 1). While this behavior may, in general, be system-dependent and estimated offline, we adopt a simplified yet representative assumption in this work, that enables the formulation of sufficient detection conditions for identifying attack-free scenarios. Defining the estimation error as $e_j^i(t) := \tilde{w}_j^i(t) - G_j g_j(t)$, we make the following assumption

Assumption *5.* The components of the sequence of estimation errors $\{e_j^i(t)\}_{t \geq 0}$ associated to agent i are contracting, i.e. the evolution of the error satisfies

$$e_j^i(t) = \alpha(t) e_j^i(t-1),\ \alpha(t) \in [0,1], \forall j \in \mathcal{N}_i, \forall t \in \mathbb{Z}_+\,. \tag{23}$$

Then, sufficient conditions to ensure attack-free condition can be stated as follows

Lemma *2.* Let $i \in \mathcal{T}_{l(t \mod L)}$, computing $\delta_j^i(t) := \tilde{w}_j^i(t) - G_j \bar{g}_j(t)$. Under Assumption 5, the sufficient conditions to certify attack-free scenario are

$$\begin{cases} \mathrm{sign}(\delta_j^i(t)) = \mathrm{sign}(\delta_j^i(t-L)), \\ |\delta_j^i(t)| \quad \leq |\delta_j^i(t-L)|. \end{cases} \tag{24}$$

Proof. Under the attack-free scenario, i.e. $\bar{g}_j(t) = g_j(t-1)$, the error $\delta_j^i(t)$ stems purely from the modeling or estimation uncertainty, resulting in $\delta_j^i(t) = e_j^i(t)$. Then, $\delta_j^i(t)$ would evolve according to (23), which ensures satisfaction of conditions (24). Now consider a deception attack where the received command is altered as in (9), with $\Delta g_j(t) \neq 0$. In this case, the deviation becomes

$$\begin{aligned}\delta_j^i(t) &= \tilde{w}_j^i(t) - G_j\left(g_j(t-1) - \Delta g_j(t)\right) \\ &= e_j^i(t) - G_j \Delta g_j(t).\end{aligned} \tag{25}$$

Suppose, for the sake of contradiction, that $\delta_j^i(t) = \alpha(t)\delta_j^i(t-1)$ still holds. Then

$$e_j^i(t) - G_j \Delta g_j(t) = \alpha(t) e_j^i(t-L). \tag{26}$$

By considering (23), it follows that

$$\begin{aligned}\alpha'(t) e_j^i(t-L) - G_j \Delta g_j(t) = \alpha(t) e_j^i(t-L) \;\Rightarrow \\ G_j \Delta g_j(t) = (\alpha'(t) - \alpha(t)) e_j^i(t-L).\end{aligned} \tag{27}$$

This means that the injected perturbation needs to be perfectly aligned with the estimation error $e_j^i(t-L)$ and scaled accordingly. From (15), we have that the injected packet lies outside the uncertainty ball, which implies that

$$\|G_j \Delta g_j(t)\| > \|e_j^i(t)\|. \tag{28}$$

Now, by considering (23) and combining (27)-(28) one arrives to

$$\|G_j \Delta g_j(t)\| = |\alpha'(t) - \alpha(t))| \|e_j^i(t-L)\| > \alpha'(t) \|e_j^i(t-L)\|, \tag{29}$$

which implies that $|\alpha'(t) - \alpha(t))| > \alpha'(t)$, that is a contradiction, since $\alpha'(t) \in [0,1]$. □

Following the results of Lemma 2, if conditions (24) hold true, then the received command $\bar{g}_j(t)$ can be safely

used. On the other hand, if either of these conditions is violated, it indicates an inconsistency attributable to a deception attack. The rationale behind Lemma 2 is that any injected data disrupting the contraction pattern of the estimation error necessarily violates conditions (24). This result highlights how prior knowledge of the dynamics of the estimation error can be effectively leveraged to design customized and provably sufficient detection conditions of attack-free scenarios.

5. RESILIENCE WITH ADAPTIVE ROBUSTNESS

While the previous analysis provides valuable insights into attack detection, the design of strategies that ensure constraint feasibility under adversarial conditions often requires complex mechanisms. Although the primary goal of this work is to examine the TB-CG strategy and demonstrate detection under realistic assumptions, in this section we propose a robust decentralized approach based on implicit communication. Specifically, coupling constraints are conservatively tightened from the outset to account for the estimation error introduced by relying on estimated—rather than received—admissible commands. To this end, the set $\mathcal{V}_i(\bar{x}_i(t))$ defined in (7) can be seen as $\mathcal{V}_i(\bar{x}_i) := \mathcal{V}_i^l(x_i) \cap \mathcal{V}_i^c(\bar{x}_i)$, where

$$\mathcal{V}_i^c(\bar{x}_i(t)) := \left\{ g_i \in \mathcal{W}_i^{\delta,c} \; : \; \begin{array}{c} c_i^c(k, \bar{x}_i(t), g_i, g_{[i]}(t)) \in \mathcal{C}_{[i]}^c, \\ k = 0, \dots, k_0 \end{array} \right\}. \quad (30)$$

The virtual predictions of the c_i^c-variable for all possible neighboring admissible command perturbation sequence realizations $\{\tilde{g}_{[i]}^\epsilon(l) \in \mathcal{D}^{\delta_{[i]}}\}_{l=0}^k$ can be defined as

$$\begin{aligned} \bar{c}_i^c(k, \bar{x}_i, g_i, \tilde{w}_{[i]}, \tilde{g}_{[i]}^\epsilon(\cdot)) := \bigcup_{\tilde{g}_{[i]}^\epsilon(\cdot) \in \mathcal{D}^{\delta_{[i]}}} \{ H_{\mathcal{N}_i}^c (\Phi_{\mathcal{N}_i}^k \bar{x}_i \\ + \sum_{i=0}^{k-1} \Phi_{\mathcal{N}_i}^{k-i-1} I_{\mathcal{N}_i} \begin{bmatrix} G_i g_i \\ \tilde{w}_{[i]} + G_j \tilde{g}_{[i]}^\epsilon(i) \end{bmatrix}) \}, \end{aligned} \quad (31)$$

where $H_{\mathcal{N}_i}^c = diag(H_i^c, H_{i_1}^c, \dots, H_{i_{\mathcal{N}_i}}^c)$ and similarly for $\Phi_{\mathcal{N}_i}$, $G_{\mathcal{N}_i}$ and where $L_i = 0$. By linearity,

$$\bar{c}_i^c(k, \bar{x}_i, g_i, \tilde{w}_{[i]}, \tilde{g}_{[i]}^\epsilon(\cdot)) = c_i^c(k, \bar{x}_i, g_i, \tilde{w}_{[i]}) + \hat{c}_i^c(k, \tilde{g}_{[i]}^\epsilon(\cdot)), \quad (32)$$

where

$$\hat{c}_i^c(k, \tilde{g}_{[i]}^\epsilon(\cdot)) := \bigcup_{\tilde{g}_{[i]}^\epsilon(\cdot) \in \mathcal{D}^{\delta_{[i]}}} \{ \sum_{i=0}^{k-1} \Phi_{\mathcal{N}_i}^{k-i-1} I_{\mathcal{N}_i} \begin{bmatrix} 0 \\ G_j \tilde{g}_{[i]}^\epsilon(i) \end{bmatrix} \}, \quad (33)$$

with $\hat{c}_i^c(k, \tilde{g}_{[i]}^\epsilon(\cdot)) \subseteq \mathcal{B}^{\Delta^{\delta_{[i]}}}$. In order to ensure constraints fulfillment, a "worst case" approach can be followed, considering only the perturbance-free evolutions $c_i^c(k, \bar{x}_i, g_i, \tilde{w}_{[i]})$. In this case, the set $\mathcal{C}_i^c$ in (30) becomes dependent on the prediction step, i.e.

$$\begin{aligned} \mathcal{C}_{[i]}^{k,c} := \mathcal{C}_{[i]}^c \sim \bar{\mathcal{B}}_k^{\Delta^{\delta_{[i]}}}, \\ \bar{\mathcal{B}}_k^{\Delta^{\delta_{[i]}}} = \bar{\mathcal{B}}_{k-1}^{\Delta^{\delta_{[i]}}} \oplus H_{\mathcal{N}_i}^c \Phi_{\mathcal{N}_i}^{k-i-1} \mathcal{B}^{\Delta^{\delta_{[i]}}}, \end{aligned} \quad (34)$$

where, under explicit communication and attack-free scenario, $\mathcal{B}^{\Delta^{e_{[i]}}}$ could be adaptively adjusted based on error terms $\delta_{[i]}(t)$. The set $\mathcal{W}_i^{\delta,c}$ changes as

$$\mathcal{W}_i^{\delta,c} := \{ g_i \in \mathbb{R}^{p_i} \; : \; c_{\bar{g}_i} \in \mathcal{C}_{[i]}^c \sim \bar{\mathcal{B}}_\infty^{\Delta^{\delta_{[i]}}} \sim \mathcal{S}_\delta \}, \quad (35)$$

where $\bar{\mathcal{B}}_\infty^{\Delta^{\delta_{[i]}}}$, the Hausdorff limit of the set sequence $\bar{\mathcal{B}}_k^{\Delta^{\delta_{[i]}}}$, is convex and compact (**?**) and that, provided all $\mathcal{C}_{[i]}^{k,c}$ are non-empty, they are also convex and compact, satisfy the nesting condition $\mathcal{C}_{[i]}^{k,c} \subset \mathcal{C}_{[i]}^{k-1,c}$and make $\mathcal{C}_{[i]}^c \sim \bar{\mathcal{B}}_\infty^{\Delta^{\delta_{[i]}}}$a nonempty convex and compact set. As a consequence, constraints are always satisfied if

$$c_i^c(k, \bar{x}_{[i]}, g_i, \tilde{w}_{[i]}) \in \mathcal{C}_{[i]}^{k,c}. \quad (36)$$

6. ILLUSTRATIVE EXAMPLE: CONNECTIVITY KEEPING

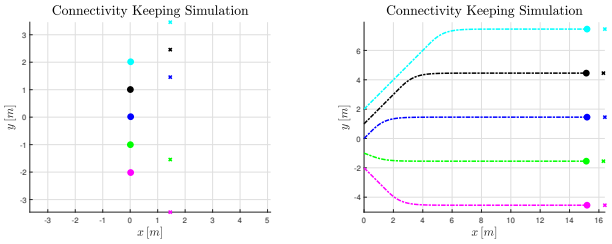
This section aims at validating the proposed method through simulation. Experiments were conducted using the CoGoV Toolbox (**?**), involving five vehicles operating in a 2D space subject to both local and convex connectivity keeping coupling constraints. For this purpose, the dynamics of the i-th vehicle are modeled as

$$\begin{cases} \ddot{x}_i^x(t) = -\dfrac{c}{m} \dot{x}_i^x(t) + T_{x,i}(t), \\ \ddot{x}_i^y(t) = -\dfrac{c}{m} \dot{x}_i^y(t) + T_{y,i}(t), \end{cases} \quad (37)$$

where $x_i^x(t)$, $x_i^y(t)$ represent the x and y positions in an earth-fixed reference frame. The models have been controlled, discretized and expressed as in (1). It is assumed that each agent i is subject to the following constraints

$$\begin{cases} |T_{l,i}(t)| \leq 1\,[N], |\dot{x}_i^l(t)| \leq 0.5\,[\dfrac{m}{s}], \\ \|x_i^l(t) - x_j^l(t)\|_\infty \leq d_{max}, \forall j \in \mathcal{N}_i, \forall t \in \mathbb{Z}_+, \end{cases} \quad (38)$$

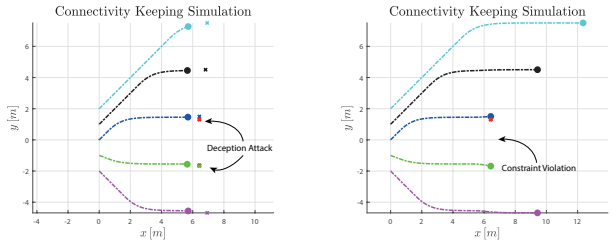
with $m = 10\,[Kg]$, $c = 1\,[\frac{Kg}{s}]$, $l = \{x, y\}$, $d_{max} = 3\,[m]$, $\mathcal{N}_{i_1} = \{i_2, i_3\}$, $\mathcal{N}_{i_2} = \{i_1, i_4\}$, $\mathcal{N}_{i_3} = \{i_1, i_5\}$, $\mathcal{N}_{i_4} = i_2$ and $\mathcal{N}_{i_5} = i_3$. The simulating scenario depicted in Figure 1(a) is considered, where agents, initially located close to each other, are instructed to reach references that violate the coupling constraints. Due to this choice, the measured distance between vehicles increases until reaching the maximum admissible distance allowing connectivity, as shown in Figure 1(b). Let us assume an attack



(a) Initial configuration of the connectivity keeping simulation. (b) Nominal behavior.

Fig. 1. Connectivity Simulation with no attacks.

occurs when the coupling constraints are saturated. In this case, the attacker injects data into the admissible command of agents i_1 and i_2, making agents believe that their respective neighbor is approaching (Figure 2(a)). Due to the non-cooperative nature of the TB-CG approach, agents interpret this as a margin to improve the local costs and compute an admissible command closer to the assigned references, which ultimately leads to infeasibility due to a violation of the connectivity keeping constraints. Consequently, agents will continuously apply their last computed command, resulting in a system-wide standstill (Figure 2(b)), demonstrating the ability of the attacker to disrupt nominal behavior on the entire system by simply manipulating the admissible command of a single agent. In contrast, the proposed approach enhances resilience to deception attacks by operating in a fully decentralized manner. Agents conservatively modify and adapt the coupling constraints to account for uncertainty in the estimation process, maintaining overall feasibility (Figure 3).

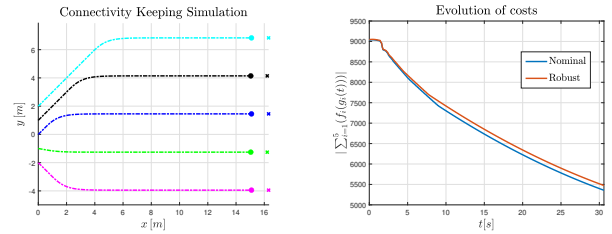


(a) Deception attack on data exchanged between agents i_1 and i_2.

(b) Disruption of overall nominal behavior caused by the deception attack.

Fig. 2. Connectivity Simulation with a false data injection attack.

However, as illustrated in Figure 3(b), due to conservative constraint-tightening, agents approach their references less closely and incur a higher overall cost compared to the nominal case, reflecting a trade-off between feasibility and performance.



(a) Behavior with robust approach.

(b) Evolution of cost in nominal and robust case.

Fig. 3. Trajectories of the vehicles the proposed robust approach with $\alpha(t) = 1, \forall t \in \mathbb{Z}_+$.

7. CONCLUSIONS

In this paper, the problem of deception attack detection in the context of the Turn-Based CG strategy has been investigated. In particular, formal insights into deception attacks detection, under both ideal and non-ideal conditions, have been provided. In the non-ideal case, sufficient conditions for identifying attack-free scenarios were established under mild assumptions. Overall, it has been demonstrated how theoretical characterization of the dynamics of the estimation error can be leveraged to design detection policies and, thus, allow for the development of inherently resilient supervision strategies. Moreover, a robust decentralized approach was also proposed, tightening coupling constraints to ensure feasibility without relying on communication. Future research endeavors could investigate compensation mechanisms and/or scenarios with partial state availability or measurements from other agents.