Detection of Stealthy Adversaries for Networked Unmanned Aerial Vehicles

Mohammad Bahrami and Hamidreza Jafarnejadsani

***Abstract*— A network of unmanned aerial vehicles (UAVs) provides distributed coverage, reconfigurability, and maneuverability in performing complex cooperative tasks. However, it relies on wireless communications that can be susceptible to cyber adversaries and intrusions, disrupting the entire network's operation. This paper develops model-based centralized and decentralized observer techniques for detecting a class of stealthy intrusions, namely zero-dynamics and covert attacks, on networked UAVs in formation control settings. The centralized observer that runs in a control center leverages switching in the UAVs' communication topology for attack detection, and the decentralized observers, implemented onboard each UAV in the network, use the model of networked UAVs and locally available measurements. Experimental results are provided to show the effectiveness of the proposed detection schemes in different case studies.**

SUPPLEMENTARY MATERIAL

Video: `https://youtu.be/lVT_muezKLU`

Code: `https://github.com/SASLabStevens/TelloSwarm`

I. INTRODUCTION

Teams of autonomous robots, particularly Unmanned Aerial Vehicles (UAVs), are of interest as their cooperation allows for distributed coverage, reconfigurability, and mobility in a wide range of applications such as search and rescue missions, surveillance, wildfire monitoring, and delivery. Networked UAVs rely on information exchange over a wireless communication network to coordinate and cooperate. However, numerous studies have shown the vulnerabilities of wireless networks to spoofing attacks, malicious intrusions, and Denial of Service (DoS), which raises security concerns in safety- and time-critical applications. This paper addresses the problem of detecting a class of *stealthy* attacks on the control system of networked UAVs in a setting where UAVs cooperate to achieve a particular formation, and their communication network is subject to topology switching.

A. Related work

Multi-UAV cooperation. Cooperation and coordination of a network of robots (UAVs) in the forms of swarming and formation control have been extensively studied [1], [2]. Formation control has been adopted in performing collaborative tasks such as coverage control for sensor networks [3], forest firefighting [4], and sensor planning for precision agriculture [5]. A team of UAVs can achieve formation by exchanging their spatial local information (e.g. position and velocity states) and in different settings such as leader-follower, leaderless consensus-based, and virtual structure approaches [2]. This paper focuses on consensus-based formation control of a team of UAVs coordinating their relative positions over a switching communication network.

Mohammad Bahrami and Hamidreza Jafarnejadsani are with the Department of Mechanical Engineering, Stevens Institute of Technology, Hoboken, NJ 07030, USA, {mbahrami,hjafarne}@stevens.edu.

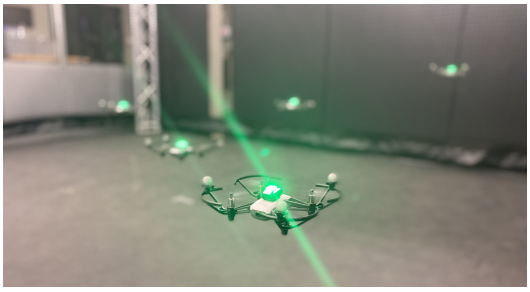


Fig. 1. The experimental setup for multi-UAV formation control using a motion capture system for positioning.

Adversary and cyber-attack detection for UAVs. The susceptibility of networked UAVs to cyber adversaries and attacks has been reported in [6], [7]. There are also different defense mechanisms against attacks and adversaries. Intrusion Detection Systems (IDS) that operate often based on the statistical characteristics of transmitted data can significantly limit possible attacks, thereby improving the network security [8]. However, IDS' invariance to the system's dynamics renders them incapable of detecting stealthy attacks devised based on networked UAVs' dynamical characteristics. Examples of such attacks are zero-dynamics attack [9], covert attack [10], and replay attack [11]. Challenges of IDS and also model-based monitoring frameworks in detecting stealthy attacks on a single UAV have been studied in [12], [13]. Redundancy in hardware and software has been adopted to detect and mitigate a class of attacks on a single UAV [14]. Secure planning against stealthy (GPS spoofing) attack has been studied using control theoretic [15] and reinforcement learning [16] approaches. Only a few studies have considered the detection problem of stealthy attacks on networked UAVs, which are encryption and encoding for detection of false data injection (FDI) attacks [17], information fusion for GPS spoofing detection [18], distributed observers for cyberattack detection and isolation [19], and finally learning-based methods to detect stealthy FDI attacks targeting a single UAV in cooperative localization [20] and to detect anomalies in swarming drones [21]. However, none of these studies have considered the detection problem of stealthy attacks, namely zero-dynamics attack (ZDA) and covert attack, on coordination of networked UAVs.

Statement of contributions. We study the detection of *stealthy*

attacks on a team of UAVs coordinating their relative positions, over a wireless network with switching communication links, to achieve formation. Stealthy attacks are designed based on the high-level coordination model (communication topology) of the UAVs to maximize the attack's effect on the cooperation performance as well as the attack's stealthiness in monitored signals.

The main contributions of this paper are as follows: We develop centralized and decentralized attack detection strategies against *stealthy* attacks, namely zero-dynamics attack (ZDA) and covert attack, that adversely affect the coordination objective of networked UAVs while remaining stealthy in the monitoring signals. We investigate switching communication links in terms of their role in attack detection and system reconfigurability. We demonstrate an experimental implementation of the stealthy attacks on a team of small UAVs and evaluate the effectiveness of our proposed methods in the timely detection of stealthy attacks.

II. Problem Formulation

A. Notations

We use $\mathbb{R}$, $\mathbb{R}_{>0}$, $\mathbb{R}_{\geq 0}$, $\mathbb{N}$, and $\mathbb{Z}_{\geq 0}$ to denote the set of reals, positive reals, non-negative reals, natural numbers, and non-negative integers, respectively. $\mathbf{1}_n$, $\mathbf{0}_n$, I_n and $\mathbf{0}_{n\times m}$ stand for the n-vector of all ones, the n-vector of all zeros, the identity n-by-n matrix, and the n-by-m zero matrix, respectively[1]. In addition, we use $\mathfrak{e}_i$ to denote the i-th canonical vector in $\mathbb{R}^n$. We use $\mathrm{col}(\cdot)$ and $\mathrm{diag}(\cdot)$ to denote the column and diagonal concatenation of vectors or matrices, and $\otimes$ to denote the Kronecker product. Finally, for any set $\mathcal{F}$, $|\mathcal{F}|$ denotes its cardinality.

B. Quadrotor's dynamics

We consider a team of N homogeneous unmanned aerial vehicles (quadrotor UAVs) that cooperate to achieve a geometric shape/formation in $\mathbb{R}^2$. Attached to the center of mass of each quadrotor, the body frame $\{\boldsymbol{\mathcal{B}}^i\}$ with unit axes $\{\vec{\mathfrak{b}}_1^i, \vec{\mathfrak{b}}_2^i, \vec{\mathfrak{b}}_3^i\}$, $i \in \{1,\dots,N\} =: \mathcal{V}$, whose position and orientation with respect to the inertial global frame $\{\boldsymbol{\mathcal{I}}\}$ with unit vectors $\{\vec{\mathfrak{e}}_x, \vec{\mathfrak{e}}_y, \vec{\mathfrak{e}}_z\}$ (see Figure 2a) are, respectively, determined by a vector $p_i = \mathrm{col}\left(p_i^x, p_i^y, p_i^z\right) \in \{\boldsymbol{\mathcal{I}}\}$, $\forall\, i \in \mathcal{V}$ and a rotation matrix $R_i(\psi_i, \phi_i, \theta_i) \in \mathrm{SO}(3)$ in the special orthogonal group with ψ_i, ϕ_i, and θ_i being the respective $z-x-y$ Euler angles. Then, the rigid body motion of the quadrotors follows [22]

$$\dot{p}_i = v_i, \qquad m\dot{v}_i = -mg\vec{\mathfrak{e}}_z + R_i f_i, \tag{1a}$$

$$\dot{R}_i = R_i\Omega_i^{\times}, \qquad J\dot{\Omega}_i = -\Omega_i \times J\Omega_i + \boldsymbol{\tau}_i, \tag{1b}$$

where $p_i \in \mathbb{R}^3$ and $R_i(\psi_i, \phi_i, \theta_i) \in \mathrm{SO}(3)$ are the position and orientation of the i-th quadrotor in the inertial frame $\{\boldsymbol{\mathcal{I}}\}$, m is the mass of the quadrotor, g is the gravitational acceleration, and finally $f_i \in \{\boldsymbol{\mathcal{B}}^i\}$ is the total thrust. Also, in the rotational dynamics, $\Omega_i \in \mathbb{R}^3$ is the angular velocity, $J \in \mathbb{R}^{3\times 3}$ is the inertia matrix, and $\boldsymbol{\tau}_i \in \mathbb{R}^3$ is the total torque, all expressed in respective body-fixed frames. Finally,

[1]We may omit the subscripts when clear from the context.

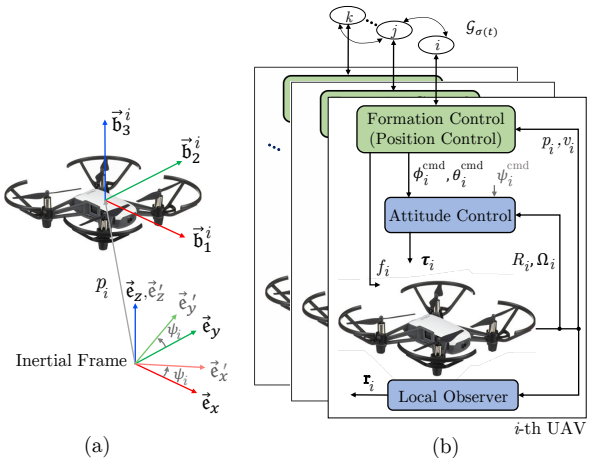


Fig. 2. (a) Illustration of reference frames. (b) The coordination control architecture.

the notation $\Omega_i^{\times}$ denotes the skew-symmetric matrix, such that $\Omega_i^{\times} r = \Omega_i \times r$ for any vector $r \in \mathbb{R}^3$ and the cross product $\times$.

C. Formation control

The cooperative control of quadrotor UAVs, shown in Figure 2b, follows a hierarchical structure, where at the high level, the UAVs coordinate with each other and their formation/position controller cooperatively generates the desired attitude/orientation and the desired total thrust for a low-level attitude controller. In this paper, we focus on 2D formation in the $x-y$ plane, for which cooperative control protocols will be designed based on a linearized model of the UAVs' transnational dynamics in (1a) around a hovering state and under small-angle approximations as follows [22]:

$$\ddot{p}_i^x = g(\Delta\theta_i \cos(\psi_i) + \Delta\phi_i \sin(\psi_i)), \tag{2a}$$

$$\ddot{p}_i^y = g(\Delta\theta_i \sin(\psi_i) - \Delta\phi_i \cos(\psi_i)), \tag{2b}$$

$$\ddot{p}_i^z = -g + f_i/m, \tag{2c}$$

in which $\Delta\theta_i$ and $\Delta\phi_i$ denote, respectively, the deviation of pitch and roll angles of the i-th quadrotor from their equilibrium point $\theta_i = \phi_i = 0$. Associated with each UAV, we define an intermediary frame $\{\boldsymbol{\mathcal{I}'}\}$ with unit axes $\{\vec{\mathfrak{e}}_x', \vec{\mathfrak{e}}_y', \vec{\mathfrak{e}}_z'\}$ and orientation $R_z(\psi_i)$ such that $p_i = R_z(\psi_i)p_i'$ for vectors $p_i \in \{\boldsymbol{\mathcal{I}}\}$ and $p_i' \in \{\boldsymbol{\mathcal{I}'}\}$ (see Figure 2a). Assuming all UAVs have consensus on a desired yaw angle $\psi_i = \psi^*$, $\forall\, i \in \mathcal{V}$, $\{\boldsymbol{\mathcal{I}'}\}$ will be the common reference frame of the UAVs in which the linearized equation of motion in (2) can be represented by

$$\ddot{p}_i^{x'} = +g\Delta\theta_i, \tag{3a}$$

$$\ddot{p}_i^{y'} = -g\Delta\phi_i, \tag{3b}$$

$$\ddot{p}_i^{z'} = -g + f_i/m, \tag{3c}$$

that shows the decoupled dynamics in the x_i', y_i', z_i' directions[2]. We let the reference commands for pitch and roll

[2]We will omit the superscript $'$ in the rest of paper for notational simplicity.

angles in (3a)-(3b) be

$$\theta_i^{\rm cmd} = +u_i^x/g, \quad \phi_i^{\rm cmd} = -u_i^y/g, \tag{4}$$

where u_i^x and u_i^y are the formation control inputs to be designed, respectively, in the x and y directions. It is also necessary to mention that $\theta_i^{\rm cmd}$ and $\phi_i^{\rm cmd}$ in (4) will be desired setpoints for each UAV's low-level (on-board) attitude controller, and that we use independent PID controllers to stabilize the altitude of quadrotors (z_i-dynamics in (2c)) around a desired hovering point. Therefore, the altitude dynamics in (2c) and the rotational dynamics in (1b) are dropped from the high-level state space of networked UAVs and the reduced-order planar dynamics is obtained by substituting (4) for $\Delta\theta_i$ and $\Delta\phi_i$ in (3a) and (3b) as follows:

$$\Sigma_i : \begin{cases} \dot{\boldsymbol{p}}_i(t) = \boldsymbol{v}_i(t) \\ \dot{\boldsymbol{v}}_i(t) = \boldsymbol{u}_i(t) \end{cases}, \quad i \in \mathcal{V} = \{1, \dots, N\}, \tag{5}$$

in which $\boldsymbol{p}_i(t) \coloneqq \mathrm{col}(p_i^x, p_i^y) \in \mathbb{R}^2$, and $\boldsymbol{v}_i(t) \coloneqq \mathrm{col}(\dot{p}_i^x, \dot{p}_i^x) \in \mathbb{R}^2$ are the stacked positions and velocities in the x and y directions, and $\boldsymbol{u}_i(t) \coloneqq \mathrm{col}(u_i^x, u_i^y) \in \mathbb{R}^2$ denotes their corresponding control input for each UAV.

Desired formation reference. We define a desired configuration (formation shape) by specifying a set of N desired setpoints $\boldsymbol{p}_1^*, \boldsymbol{p}_2^*, \dots, \boldsymbol{p}_N^*$ in $\mathbb{R}^2$ that form the desired relative positions[3] $\{\boldsymbol{p}_{ij}^* = \boldsymbol{p}_i^* - \boldsymbol{p}_j^* \in \mathbb{R}^2 \mid \forall i, j \in \mathcal{V},\ i \neq j\}$, all expressed in the UAVs' common frame. The formation references are transmitted to the UAVs from a ground control center. We follow the consensus-based formation settings [23] where the UAVs coordinate their relative positions to reach the desired relative positions $\boldsymbol{p}_{ij}^*$'s, which is formulated as

$$\lim_{t\to\infty} |\boldsymbol{p}_i(t) - \boldsymbol{p}_j(t) - \boldsymbol{p}_{ij}^*| = \boldsymbol{0}, \qquad \forall\, i, j \in \mathcal{V}, \tag{6a}$$

$$\lim_{t\to\infty} |\boldsymbol{v}_i(t)| = \boldsymbol{0}, \qquad \forall\, i \in \mathcal{V}. \tag{6b}$$

Inter-UAV communication. We model the switching inter-UAV communication by an undirected graph $\mathcal{G}_{\sigma(t)} = (\mathcal{V}, \mathcal{E}_{\sigma(t)})$, where the vertex set $\mathcal{V} = \{1, \dots, N\}$ represents the index set of N UAVs in (1) (with their respective reduced models in (5)), and the edge set $\mathcal{E}_{\sigma(t)} \subseteq \mathcal{V} \times \mathcal{V}$ represents the communication links such that an edge $(i, j) \in \mathcal{E}_{\sigma(t)}$ implies information exchange between the i-th and j-th UAV in a given active mode determined by the right-continuous switching signal $\sigma(t) : \mathbb{R}_{\geq 0} \to \mathcal{Q} := \{1, 2, \dots, q\}$, $q \in \mathbb{N}$, at time t, with $\mathcal{Q}$ being the finite index set of possible communication graphs. The UAVs' interaction is further encoded into a symmetric adjacency matrix $\mathsf{A}_{\sigma(t)} \coloneqq [a_{ij}^{\sigma(t)}] \in \mathbb{R}_{\geq 0}^{N\times N}$ such that $a_{ij}^{\sigma(t)} = a_{ji}^{\sigma(t)} = 1$ if an edge $(i, j) \in \mathcal{E}_{\sigma(t)}$, and $a_{ij}^{\sigma(t)} = a_{ji}^{\sigma(t)} = 0$, otherwise. Also, the set of the neighbors of the i-th node (UAV) in any active mode $\sigma(t)$ is defined as $\mathcal{N}_{\sigma(t)}^i = \{j \in \mathcal{V} \mid (i, j) \in \mathcal{E}_{\sigma(t)}\}$.

Throughout this paper, we assume the inter-UAV's communication graphs $\mathcal{G}_{\sigma(t)}$'s are connected in all modes.

[3]In the context of formation control, these reference states are called *formation states* [23] or *shape vectors* [24] depending on the design methods and their underlying assumptions.

To meet the formation constraints in (6), we use the following consensus-based distributed control protocol:

$$\boldsymbol{u}_i = \boldsymbol{u}_{{\rm n}_i} + \boldsymbol{u}_{{\rm a}_i}, \quad i \in \mathcal{V}, \tag{7}$$

$$\boldsymbol{u}_{{\rm n}_i} = -\alpha \sum_{j\in\mathcal{N}_{\sigma(t)}^i} a_{ij}^{\sigma(t)} (\boldsymbol{p}_i - \boldsymbol{p}_j - \boldsymbol{p}_{ij}^*) - \gamma \boldsymbol{v}_i \tag{8}$$

where $\boldsymbol{u}_{{\rm n}_i} \in \mathbb{R}^2$ denotes the nominal control input with $a_{ij}^{\sigma(t)}$ being the entry of the symmetric adjacency matrix associated with the UAVs' switching communication graph $\mathcal{G}_{\sigma(t)}$. Also, $\alpha \in \mathbb{R}_{>0}$ and $\gamma \in \mathbb{R}_{>0}$ are the control gains. $\boldsymbol{u}_{{\rm a}_i} \in \mathbb{R}^2$ is the vector-valued malicious signal injected in the control channel of the i-th UAV.

We let an unknown subset $\mathcal{A} = \{i_1, i_2, \dots\} \subset \mathcal{V}$ denote the set of UAVs subject to attack $\boldsymbol{u}_{{\rm a}_i} \neq \boldsymbol{0}$, which we refer to as compromised UAVs, and we refer to the rest of UAVs Σ_i's with $\boldsymbol{u}_{{\rm a}_i} = \boldsymbol{0},\ \forall\, i \in \mathcal{V} \setminus \mathcal{A}$, in (5) as uncompromised UAVs.

Network-level dynamics. Given (5), (7) and (8), the dynamics of the networked UAVs can be represented by

$$\Sigma : \dot{\mathbf{x}} = \mathbf{A}_{\sigma(t)}\mathbf{x} + \mathbf{B}_{\sigma(t)}^{\boldsymbol{F}}\mathbf{x}^* + \mathbf{B}_{\mathcal{A}}\mathbf{u}_{\rm a}, \quad \mathbf{x}(t_0) = \mathbf{x}_0, \tag{9}$$

in which, the system states and matrices are given by

$$\mathbf{x}(t) = \mathrm{col}\,(\boldsymbol{p}_1, \dots, \boldsymbol{p}_N, \boldsymbol{v}_1, \dots, \boldsymbol{v}_N) \in \mathbb{R}^{4N}, \tag{10a}$$

$$\mathbf{x}^* = \mathrm{col}\,(\boldsymbol{p}_1^*, \dots, \boldsymbol{p}_N^*, \boldsymbol{0}_2, \dots, \boldsymbol{0}_2) \in \mathbb{R}^{4N}, \tag{10b}$$

$$\mathbf{A}_{\sigma(t)} = A_{\sigma(t)} \otimes I_2,\ \mathbf{B}_{\sigma(t)}^{\boldsymbol{F}} = -\mathbf{A}_{\sigma(t)},\ \mathbf{B}_{\mathcal{A}} = B_{\mathcal{A}} \otimes I_2, \tag{10c}$$

$$A_{\sigma(t)} = \begin{bmatrix} 0_{N\times N} & I_N \\ -\alpha \mathsf{L}_{\sigma(t)} & -\gamma I_N \end{bmatrix},\ B_{\mathcal{A}} = \begin{bmatrix} 0 \\ \mathrm{B}_{\mathcal{A}} \end{bmatrix}, \tag{10d}$$

$$\mathrm{B}_{\mathcal{A}} = \begin{bmatrix} \mathfrak{e}_{i_1} & \mathfrak{e}_{i_2} & \dots & \mathfrak{e}_{i_{|\mathcal{A}|}} \end{bmatrix},\ \ \mathbf{u}_{\rm a} = \mathrm{col}\,(\boldsymbol{u}_{{\rm a}_i})_{i\in\mathcal{A}}, \tag{10e}$$

where $\mathsf{L}_{\sigma(t)}$ is the Laplacian matrix of graph $\mathcal{G}_{\sigma(t)}$, encoding the inter-UAVs' communication links, defined as $\mathsf{L}_{\sigma(t)} \coloneqq [l_{ij}^{\sigma(t)}] \in \mathbb{R}^{N\times N}$ with $l_{ii}^{\sigma(t)} = \sum_{j\neq i} a_{ij}^{\sigma(t)}$ and $l_{ij}^{\sigma(t)} = -a_{ij}^{\sigma(t)}$ if $i \neq j$. $\mathfrak{e}_i$ is the i-th vector of the canonical basis in $\mathbb{R}^N$ corresponding to the i-th UAV compromised by attack $\boldsymbol{u}_{{\rm a}_i},\ i \in \mathcal{A}$.

Network-level measurements. We define the system measurements $\mathbf{y}$ to be composed of the position of a set of UAVs $\mathcal{M}_{\rm p} = \{{\rm p}_1, {\rm p}_2, \dots\} \subset \mathcal{V}$ and/or the velocity of a set of UAVs $\mathcal{M}_{\rm v} = \{{\rm v}_1, {\rm v}_2, \dots\} \subset \mathcal{V}$ that are transmitted to a ground control center for monitoring. More precisely,

$$\mathbf{y} = \mathbf{C}\mathbf{x} - \mathbf{u}_{\rm s}, \quad \mathbf{C} = C \otimes I_2, \quad \mathcal{M} = \{\mathcal{M}_{\rm p}, \mathcal{M}_{\rm v}\}, \tag{11a}$$

$$C = \mathrm{diag}\,(C_{\rm p}, C_{\rm v}), \tag{11b}$$

$$C_{\rm p} = \mathrm{col}\big(\mathfrak{e}_{{\rm p}_1}^\top, \mathfrak{e}_{{\rm p}_2}^\top, \dots, \mathfrak{e}_{{\rm p}_{|\mathcal{M}_{\rm p}|}}^\top\big) \in \mathbb{R}^{|\mathcal{M}_{\rm p}|\times N}, \tag{11c}$$

$$C_{\rm v} = \mathrm{col}\big(\mathfrak{e}_{{\rm v}_1}^\top, \mathfrak{e}_{{\rm v}_2}^\top, \dots, \mathfrak{e}_{{\rm v}_{|\mathcal{M}_{\rm v}|}}^\top\big) \in \mathbb{R}^{|\mathcal{M}_{\rm v}|\times N}, \tag{11d}$$

where $\mathbf{u}_{\rm s} = \mathrm{col}\big(\mathbf{u}_{{\rm s}_1}, \mathbf{u}_{{\rm s}_2}, \dots, \mathbf{u}_{{\rm s}_{|\mathcal{M}|}}\big) \in \mathbb{R}^{2|\mathcal{M}|}$ denotes the vector-valued sensory attacks on the measurements.

Proposition II.1. (Formation convergence). *Assume that the formation configuration is feasible and that the communication graph is connected in each mode. Then, under the control protocol* (7)*, and in the absence of attacks, the states of the UAVs in* (5) *converge to the desired formation configuration in* (6)*.*

Proof. The proof follows a change of variables as in [25] and a convergence analysis similar to that in [9]. The details are omitted here due to space limitation. □

Note that the UAV's dynamics in (5) as well as the control protocol (8) for the x and y directions are decoupled. Thus, for notational simplicity, we may use the following

$$\Sigma : \dot{\mathrm{x}} = A_{\sigma(t)}\mathrm{x} + B^{\boldsymbol{F}}_{\sigma(t)}\mathrm{x}^{*} + B_{\mathcal{A}}\mathrm{u}_{\mathrm{a}}, \quad \mathrm{x}(t_0) = \mathrm{x}_0, \tag{12a}$$

$$\mathrm{y} = C\mathrm{x} - \mathrm{u}_{\mathrm{s}}, \tag{12b}$$

to represent the dynamics in (9) with its monitored states in (11b) in only one direction of the $x-y$ plane. Accordingly, $\mathrm{x} = \mathrm{col}\,(\mathrm{p}, \mathrm{v}) \in \mathbb{R}^{2N}$ in (12) denotes the stacked vector of all positions $\mathrm{p} = \mathrm{col}\,(\mathrm{p}_i)_{i\in\mathcal{V}}$ and velocities $\mathrm{v} = \mathrm{col}\,(\mathrm{v}_i)_{i\in\mathcal{V}}$ in one direction with their corresponding formation references $\mathrm{x}^{*} \in \mathbb{R}^{2N}$ as well as attack inputs $\mathrm{u}_{\mathrm{a}} \in \mathbb{R}^{|\mathcal{A}|}$ and $\mathrm{u}_{\mathrm{s}} \in \mathbb{R}^{|\mathcal{M}|}$, and other system matrices are given in (10d)-(10e) and (11b) with $B^{\boldsymbol{F}}_{\sigma(t)} = -A_{\sigma(t)}$.

D. Attack stealthiness

Motivated by the susceptibility of wireless networks to adversarial intrusions [6], [7], we study the worst-case scenario adversarial settings where an attacker leverages *a priori* system knowledge of the UAVs' coordination or a prerecorded sequence of sensory data to design sophisticated stealthy attacks implementable through actuator attacks $\boldsymbol{u}_{\mathrm{a}_i}(t)$'s, $i \in \mathcal{A}$ in (7) and sensor attacks $\mathbf{u}_{\mathrm{s}}(t)$ in (11a).

Here, *a priori* system knowledge refers to the initial configuration of the networked system (9) with the measurements (11) (or equivalently (12)), denoted by the tuple $\hat{\Sigma}(\hat{\mathbf{A}}_{\sigma(t)}, \hat{\mathbf{C}}, \sigma(t) = 1)$ with $\hat{\mathbf{A}}_{\sigma(t)}$ and $\hat{\mathbf{C}}$ being the approximations of their counterparts in (9) and (11). The amount of *a priori* system knowledge needed for designing *stealthy* attacks varies for different attacks [26], and will be quantified in Section III-A.

Stealthy attacks refer to a class of adversarial intrusions (cyber attacks [10], [11], [27]) $\boldsymbol{u}_{\mathrm{a}_i}$'s, $i \in \mathcal{A}$ in (7) and $\mathbf{u}_{\mathrm{s}}$ in (11a) that disrupt the system's normal operation while remain stealthy in the monitored measurements (11), that is

$$\mathbf{y}(\mathbf{x}_0, \mathbf{u}_{\mathrm{a}}, \mathbf{u}_{\mathrm{s}}, t) = \mathbf{y}^{\mathrm{n}}(\mathbf{x}^{\mathrm{n}}_0, \mathbf{0}, \mathbf{0}, t), \quad \forall\, t \in [t_0, t_d), \tag{13}$$

where $\mathbf{y}^{\mathrm{n}}(\mathbf{x}^{\mathrm{n}}_0, \mathbf{0}, \mathbf{0}, t) = \mathbf{C}\mathbf{x}^{\mathrm{n}}$ is the output associated with an attack-free system with the same dynamics as in (12a), and $\mathbf{x}_0$ and $\mathbf{x}^{\mathrm{n}}_0$ are the actual and a possible initial states, respectively. Also, t_0 is the initial time instant and t_d is the attack detection time instant, i.e., the time instant at which condition in (13) no longer holds and attacks lose their stealthiness.

E. Problem statement: attack detection

We consider the attack detection problem as a hypothesis testing problem with the null and alternative hypotheses

$$\mathcal{H}^{\circ} : \text{attack-free, vs. } \mathcal{H}^{\scriptscriptstyle 1} : \text{attacked}, \tag{14}$$

for which we present detection frameworks in Section III-B.

III. Observer Design and Analysis for Attack Detection

In this section, we characterize the models for stealthy attacks on the networked UAVs in (9) and develop centralized and decentralized detection schemes.

A. Realization of stealthy attacks

Given system in (9), let $\mathcal{M}$ in (11a) be a set of monitored states and let $\mathcal{A}$ be a set of compromised UAVs subject to attack $\boldsymbol{u}_{\mathrm{a}_i} \neq \mathbf{0}$ in (7). In what follows, we characterize stealthy attacks in terms of different realizations of (13).

Zero-dynamics attack (ZDA). ZDA refers to the class of attacks based on the zero dynamics of the system $(A_{\sigma(t)}, B_{\mathcal{A}}, C, \sigma(t) = 1)$ in (12) that are (nontrivial) state trajectories excited through input directions $B_{\mathcal{A}}$ and invisible at the output y, and that can be characterized by the rank deficiency of matrix pencil

$$P(\lambda_o) = \begin{bmatrix} \lambda_o I_N - A_1 & -B_{\mathcal{A}} \\ C & \mathbf{0} \end{bmatrix}, \tag{15}$$

for some $\lambda_o \in \mathbb{R}_{>0}$ [9].

Proposition III.1. *Assume the system in* (9) *in its initial active mode* $\sigma(t) = 1$ *has unstable zero dynamics, i.e., the matrix pencil* $P(\lambda_o)$ *in* (15) *is rank deficient for some* $\lambda^x_o,\ \lambda^y_o, \in \mathbb{R}_{>0}$*, and that the attacker's a priori knowledge of the system* $\hat{\Sigma}(\hat{\mathbf{A}}_{\sigma(t)}, \hat{\mathbf{C}}, \sigma(t) = 1) = (\mathbf{A}_{\sigma(t)}, \mathbf{C}, \sigma(t) = 1)$*. Then, there exists a stealthy attack policy*

$$\mathbf{u}_{\mathrm{a}} = \mathrm{col}\,(\boldsymbol{u}_{\mathrm{a}_i})_{i\in\mathcal{A}}\,,\ \boldsymbol{u}_{\mathrm{a}_i} = [\mathrm{u}^x_{\mathrm{a}_i}(0)e^{\lambda^x_o t} \quad \mathrm{u}^y_{\mathrm{a}_i}(0)e^{\lambda^y_o t}\ ]^{\top}, \tag{16}$$

in dynamics (9) *that causes part of system states exponentially deviate from the formation configuration in* (6) *while the condition in* (13) *holds. In this attack model, the measurement signals are not compromised, i.e.,* $\mathbf{u}_{\mathrm{s}} = \mathbf{0}$.

Proof. The proof follows from [27, Prop. II.4] and so is omitted here. □

It is noteworthy that the assumption on *a priori* system knowledge in Proposition III.1 can be relaxed. In the cases that only a subset of the system model as *a priori* is disclosed to the attacker, that is $\hat{\Sigma}(\hat{\mathbf{A}}_{\sigma(t)}, \hat{\mathbf{C}}, \sigma(t) = 1) \approx (\mathbf{A}_{\sigma(t)}, \mathbf{C}, \sigma(t) = 1)$, a ZDA can be realized that only affects the UAVs within the known subset of the system, which is known as local ZDA [26].

Covert attack [10]. Covert attacks are a class of intrusions through input channels $\mathbf{B}_{\mathcal{A}}$ whose covertness at the output is obtained by alteration of the measurement signals (11a) and whose realization requires perfect knowledge of the system i.e. $\hat{\Sigma}(\hat{\mathbf{A}}_{\sigma(t)}, \hat{\mathbf{C}}, \sigma(t) = 1) = (\mathbf{A}_{\sigma(t)}, \mathbf{C}, \sigma(t) = 1)$. Let attack policy $\mathbf{u}_{\mathrm{a}}(t) = \mathrm{col}\,(\boldsymbol{u}_{\mathrm{a}_i})_{i\in\mathcal{A}} : \mathbb{R}_{\geq 0} \mapsto \mathbb{R}^2$ in (9) be any continuous signal initiated at time instant $t_{\mathrm{a}} \in \mathbb{R}_{\geq 0}$. Then, the attack is covert and (13) holds if the attacker alters the measurement (11a) by

$$\mathbf{u}_{\mathrm{s}}(t) = \mathbf{C}\int_{t_{\mathrm{a}}}^{t} e^{\mathbf{A}_1(t-\tau)}\mathbf{B}_{\mathcal{A}}\mathbf{u}_{\mathrm{a}}(\tau)d\tau. \tag{17}$$

We refer to [27] for details of the derivation and proof.

Cooperative DoS and replay attack. It is shown that a denial-of-service (DoS), interfering in a UAVs' communication, causes unstable and unsafe flights [28]. We formulate a scenario where replay attacks[4] [11] are implemented in cooperation with a DoS in order to keep the DoS stealthy in the networked-level measurements (11a). Here, the cooperatively-stealthy DoS and replay attack take place when the UAVs have reached the formation configuration in (6) and thus are hovering only, giving the attacker the opportunity to record and store slow-varying measurements (11a) for a time interval $T_r \in \mathbb{R}_{>0}$ before starting the attacks $\mathbf{u}_\mathrm{a}$ and $\mathbf{u}_\mathrm{s}$ respectively in (9) and (11a) that is $\mathbf{u}_\mathrm{a}(t) = \mathbf{0}$ and $\mathbf{u}_\mathrm{s}(t) = \mathbf{0}$, $\forall\, t \in [t_0\ t_\mathrm{a})$ where $t_\mathrm{a} > T_r$. Then, upon starting a DoS at a time instant $t_\mathrm{a} \in \mathbb{R}_{>0}$, causing one or more UAVs lose their inter-communication and deviating from the equilibrium states (6), a concurrent replay attack $\mathbf{u}_\mathrm{s}$ in (11a) given by

$$\mathbf{u}_\mathrm{s}(t) = \mathbf{C}\mathbf{x}(t) - \mathbf{y}(t - nT_r),\ n \in \mathbb{N},\quad t \geq t_\mathrm{a}, \tag{18}$$

causes the stealthiness condition in (13) holds.

We note that *a priori* system knowledge is not required for the cooperative DoS and replay attack that is $\hat{\Sigma}(\hat{\mathbf{A}}_{\sigma(t)}, \hat{\mathbf{C}}, \sigma(t)) = \emptyset$.

B. Observer-based detection framework

The susceptibility of observers/monitors to the sophisticated attacks satisfying the stealthiness condition in (13) has been demonstrated in both deterministic and stochastic settings in [9], [10], [11]. Motivated by this challenge, we present centralized and decentralized detection schemes to address the attack detection problem formulated as in (14).

1) Centralized detection scheme: In the centralized detection scheme, we leverage switching links in the inter-UAVs' communication topology to generate model discrepancy rendering the stealthy attacks detectable in the measurements (11) monitored in a ground control center. Note that the UAVs' communication may be subject to switching connections in two senses. First, a communication link failure induced due to operation in uncertain environments, and second, a planned switch (addition or removal of connections) triggered for security and performance reasons. Regardless of the underlying causes of switching links in the inter-UAVs' communication, we investigate their effect on the detection of stealthy attacks.

Consider the dynamical system in (12), a centralized attack detection monitor (central monitor), derived based on the initial (normal) communication mode of UAVs ($\sigma(t) = 1$), is given by

$$\Sigma_\mathcal{O}^\mathcal{M}: \begin{cases} \dot{\hat{\mathrm{x}}} = A_{\sigma(t)}\hat{\mathrm{x}} + \mathrm{B}^{\boldsymbol{F}}_{\sigma(t)}\mathrm{x}^* + \mathrm{H}(\mathrm{y} - \hat{\mathrm{y}}), & \sigma(t) = 1, \\ \hat{\mathrm{y}} = C\hat{\mathrm{x}}, & \hat{\mathrm{x}}(0) = \mathbf{0}, \\ \mathrm{r}_0 = \mathrm{y} - \hat{\mathrm{y}}, & \text{central residual}, \end{cases} \tag{19}$$

where H is an observer gain such that $(A_{\sigma(t)} - HC)$ is stable in all modes and $\lim_{t\to\infty} \mathrm{r}_0 = \mathbf{0}$ in the absence of attacks [27]. Also, we let $\mathrm{r}_0^j(t) = C^j(\mathrm{x} - \hat{\mathrm{x}})$ denote the j-th component of the residual r_0 with C^j being the j-th row vector of matrix C in (11b). Then, in the absence of attacks, an upper bound on the residuals is obtained as follows:

$$|\mathrm{r}_0^j(t)| \leq \bar{k}_j e^{-\bar{\lambda}_j t}\bar{\omega} + \epsilon_0 =: \epsilon_0^j, \tag{20}$$

where $\bar{k}_j$ and $\bar{\lambda}_j$ are positive constants such that $|C^j e^{(A_1 - HC)t}| \leq \bar{k}_j e^{-\bar{\lambda}_j t}$, $\bar{\omega}$ is an upper bound such that $|\mathrm{e}(0)| = |\mathrm{x}(0) - \hat{\mathrm{x}}(0)| = |\mathrm{x}(0)| \leq \bar{\omega}$, and $\epsilon_0 \in \mathbb{R}_{>0}$ is a sufficiently small constant to account for measurement noises.

Given the central monitor (19) and its corresponding thresholds in (20), the hypothesis testing problem in (14) can be quantified either by

$$\mathcal{H}^\mathtt{o}: \text{attack-free}, \quad \text{if} \quad |\mathrm{r}_0^j(t)| \leq \epsilon_0^j,\ \ \forall\, j \in \mathcal{M}, \tag{21a}$$
$$\mathcal{H}^\mathtt{1}: \text{attacked}, \quad \text{if} \quad |\mathrm{r}_0^j(t)| > \epsilon_0^j,\ \ \exists\, j \in \mathcal{M}, \tag{21b}$$

or by

$$\mathrm{r}_0^\top \boldsymbol{\Sigma}_{\mathrm{r}_0}^{-1} \mathrm{r}_0 \underset{\mathcal{H}^\mathtt{1}}{\overset{\mathcal{H}^\mathtt{o}}{\lessgtr}} \mathtt{threshold}, \tag{22}$$

with $\boldsymbol{\Sigma}_{\mathrm{r}_0}$ being the covariance of the residual r_0 having a zero-mean Gaussian distribution in stochastic settings where a discretized version of (19) as a Kalman filter is used, together with χ^2 (chi-squared) tests, for attack detection [11].

As shown in [11], [27], χ^2 detectors, Kalman filters, and Luenberger-type observers/monitors fail in detecting the stealthy attacks that were presented in Section III-A provided the stealthiness condition (13) holds, causing a false validation of the null hypothesis (21a). Here, based on the results in [27, Th. III.3], we evaluate the effect of switching connections in inter-UAVs' communication on the violation of (13) and thus on the validation of the null hypothesis (21a). This procedure will be presented in Algorithm 2 in Section IV.

2) Decentralized detection scheme: In the decentralized detection scheme, a set of UAVs, equipped with on-board (local) monitors, leverage the information exchange with their neighbouring UAVs to locally detect the stealthy attacks on their neighbours. Upon attack detection, a local monitor triggers an inter-UAV communication switch and informs other local monitors as part of a contingency plan (see Algorithm 2).

Note that in the networked UAVs with a connected communication graph $\mathcal{G}_{\sigma(t)}$, any UAV has access to the states of itself as well as the position states of the set of its immediate neighbors $\mathcal{N}^i_{\sigma(t)}$ (cf. control protocol (8)). Accordingly, we define, for the i-th UAV in the network, a set of local measurements, indexed by set $\mathcal{M}^i$, as follows:

$$\mathcal{M}^i = \mathcal{N}^i_{\sigma(t)} \cup \{i\}, \qquad \sigma(t) = 1 \in \mathcal{Q}, \tag{23a}$$
$$\mathbf{y}_i = \mathbf{C}_i\mathbf{x}, \ \text{ and } \ \mathrm{y}_i = C_i\mathrm{x}, \quad \mathbf{C}_i = C_i \otimes I_2, \tag{23b}$$
$$C_i = \mathrm{diag}\left(\mathrm{C}_{\mathrm{p},i}, \mathfrak{e}_i^\top\right), \qquad \mathrm{C}_{\mathrm{p},i} = \mathrm{col}\left(\mathfrak{e}_j^\top\right)_{j\in\mathcal{M}^i}. \tag{23c}$$

where $\mathbf{x}$ and x are the system states in (9) and (12), respectively. Different from the networked measurements in (11), the local measurements in (23) are not transmitted

[4] A replay attack is the case that the attacker replays (periodically) a sequence of stored data as real-time measurements to conceal any deviation from a normal operation.

through compromised network channels to the control center for monitoring. Instead, they are locally available for each UAV, and thus are not subject to alterations by sensory attacks.

Given the local measurements (23) and dynamics (12), we define the local attack detector $\Sigma_{\mathcal{O}}^i$ for the i-th UAV as follows:

$$\Sigma_{\mathcal{O}}^i : \begin{cases} \dot{\hat{\mathrm{x}}}_i = A_{\sigma(t)}\hat{\mathrm{x}}_\mathrm{i} + \mathrm{B}_{\sigma(\mathrm{t})}^{\boldsymbol{F}}\mathrm{x}^* + \mathrm{H}^\mathrm{i}(\mathrm{y}_\mathrm{i} - \hat{\mathrm{y}}_\mathrm{i}), & \sigma(t) \in \mathcal{Q}, \\ \hat{\mathrm{y}}_\mathrm{i} = \mathrm{C}_\mathrm{i}\hat{\mathrm{x}}, & \hat{\mathrm{x}}(0) = \mathbf{0}, \\ \mathrm{r}_i = \mathrm{y}_i - \hat{\mathrm{y}}_\mathrm{i}, & \text{local residual}, \end{cases} \tag{24}$$

where $\hat{\mathrm{x}}_\mathrm{i}$ is the local estimation of x in (12), and H^i is an observer gain such that $(A_{\sigma(t)} - H^i C_i)$ is stable in all modes. Therefore, in the absence of attacks, $\lim_{t\to\infty} \mathrm{r}_i = \mathbf{0}$, and similar to the central monitor's, the j-th component of local residuals, r_i^j's, hold an upper bound (threshold) as follows:

$$|\mathrm{r}_i^j(t)| \leq \bar{k}_{i,j} e^{-\bar{\lambda}_{i,j} t}\bar{\omega} + \epsilon_i =: \epsilon_i^j, \tag{25}$$

where $\bar{k}_{i,j}$ and $\bar{\lambda}_{i,j}$ are positive constants such that $|C_i^j e^{(A_1 - H^i C_i^j)t}| \leq \bar{k}_{i,j} e^{-\bar{\lambda}_{i,j} t}$, $\bar{\omega}$ is an upper bound such that $|\mathrm{e}_i(0)| = |\mathrm{x}(0) - \hat{\mathrm{x}}_\mathrm{i}(0)| = |\mathrm{x}(0)| \leq \bar{\omega}$, and $\epsilon_i \in \mathbb{R}_{>0}$ is a sufficiently small constant to account for measurement noises.

Now, the hypothesis testing problem in (14) can be revisited and quantified using local residuals as follows:

$$\mathcal{H}^{\mathrm{o}} : \text{attack-free, if } |\mathrm{r}_i^j(t)| \leq \epsilon_i^j,\ \forall j \in \mathcal{M}^i,\ \forall i \in \mathcal{D}, \tag{26a}$$

$$\mathcal{H}^{\mathrm{1}} : \text{attacked,} \quad \text{if } |\mathrm{r}_i^j(t)| > \epsilon_i^j,\ \exists j \in \mathcal{M}^i,\ \exists i \in \mathcal{D}, \tag{26b}$$

where $\mathcal{D}$ is the set of all the UAVs equipped with a local detector as in (24).

Note that a successful attack detection using the hypothesis testing (26) does depend on the sensitivity of the local residuals, r_i's, to the stealthy attacks. In this regard, the following results characterize the capability of local detectors in detecting stealthy attacks.

Proposition III.2. *Consider dynamics* (12) *and let the i-th UAV be equipped with the local attack detector $\Sigma_{\mathcal{O}}^i$ in* (24) *and local measurements* (23)*. Then, stealthy ZDA and covert attacks are detectable in $\Sigma_{\mathcal{O}}^i$'s residual r_i if the set of compromised UAVs satisfies $\mathcal{A} \subseteq \mathcal{N}_{\sigma(t)}^i$, $\sigma(t) = 1 \in \mathcal{Q}$.*

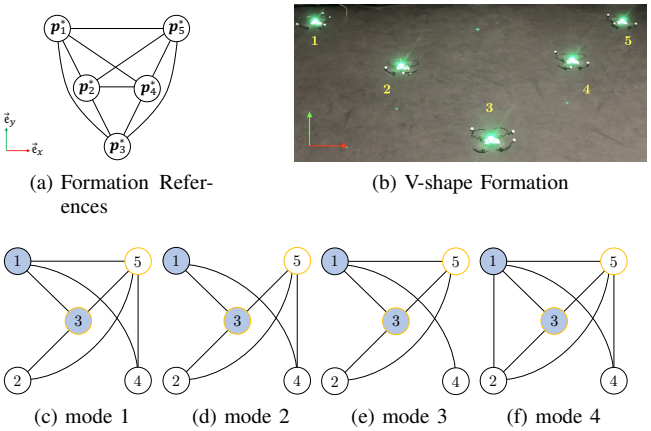
Proof. Omitted due to space limitation. □

Note that the i-th UAV's local monitor, $\Sigma_{\mathcal{O}}^i$, $i \in \mathcal{D}$ secures the networked UAVs against the stealthy attacks on its neighbors' set $\mathcal{N}_{\sigma(t)}^i$. Therefore, the problem of interest is to determine a set $\mathcal{D} \subseteq \mathcal{V}$ of local detectors $\Sigma_{\mathcal{O}}^i$'s, $i \in \mathcal{D}$ such that they cover the entire set $\mathcal{V}$ of UAVs.

Proposition III.3. *Consider the networked UAVs with the dynamics in* (12) *subject to stealthy attacks on a set of compromised UAVs $\mathcal{A} \subseteq \mathcal{V}$ and let the set*

$$\mathcal{D} := \{i \in \mathcal{V} \mid \bigcup_{i\in\mathcal{D}} \mathcal{N}_{\sigma(t)}^i = \mathcal{V},\ \sigma(t) = 1 \in \mathcal{Q}\}, \tag{27}$$

represent the set of UAVs equipped with local attack detectors $\Sigma_{\mathcal{O}}^i$'s in (24)*. Then, stealthy ZDA and covert attacks*



(a) Formation References (b) V-shape Formation

(c) mode 1 (d) mode 2 (e) mode 3 (f) mode 4

Fig. 3. Multi-UAV's formation and communication topology. (a) Formation references specifying a V-shape in the $x-y$ plane. (b) V-shape formation of UAVs. (c)-(f) Inter-UAV's communication graph $\mathcal{G}_{\sigma(t)}$ with four modes $\sigma(t) = \{1, 2, 3, 4\} =: \mathcal{Q}$. UAVs initially communicate in mode $\sigma(t) = 1$ and may switch to other modes $\sigma(t) = \{2, 3, 4\}$ if activated by a local detector. Blue nodes indicate the UAVs equipped with a local monitor and orange nodes specify the UAVs monitored by the ground control center.

undetectable in $\Sigma_{\mathcal{O}}^i$'s residual r_i, $\forall i \in \mathcal{D}$, is impossible, securing the entire network set $\mathcal{V}$ of UAVs against stealthy attacks.

Proof. Omitted due to space limitation. □

Algorithm 1 Attack detection by the i-th local monitor, $i \in \mathcal{D}$

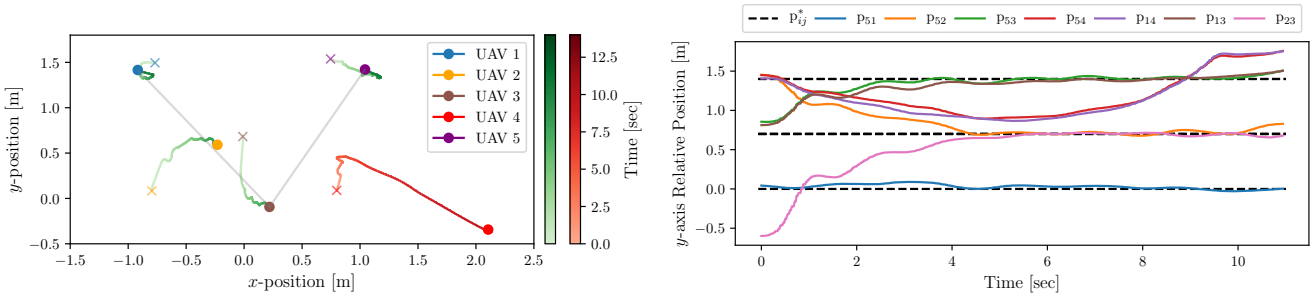
Input: $\Sigma_{\mathcal{O}}^i$, $i \in \mathcal{D}$ in (24) and (27), $\mathbf{y}_i$ in (23), ϵ_i^j in (25)

procedure LOCAL HYPOTHESIS TESTING (26)
  while $\mathcal{H}^{\mathrm{o}}$ in (26a) **do**
    Compute local residual r_i as in (24)
    Compute corresponding thresholds ϵ_i^j as in (25)
    if $|\mathrm{r}_i^j| > \epsilon_i^j$ **then**
      Reject the null hypothesis $\mathcal{H}^{\mathrm{o}}$ in (26a) ▷ Stealthy attack is locally detected.
      cooperate with other local detectors in $\mathcal{D}$ to run a contingency plan for the entire network.
    end if
  end while
end procedure

Algorithm 2 Topology switching for centralized attack detection

Inputs: local observer: $\Sigma_{\mathcal{O}}^i$, $i \in \mathcal{D}$ in (24) and (27), $\mathbf{y}_i$ in (23), ϵ_i^j in (25);
centralized observer: $\Sigma_{\mathcal{O}}^{\mathcal{M}}$ in (19), $\mathbf{y}$ in (11), ϵ_0^j in (20)

procedure CENTRAL HYPOTHESIS TESTING (21)
  Run Algorithm 1
  if $\mathcal{H}^{\mathrm{1}}$ in (26b) **then**
    Switch to a new comm. mode $\sigma(t) \in \mathcal{Q} \setminus 1$ ▷ Stealthy attack has been detected locally.
  end if
  while $\mathcal{H}^{\mathrm{o}}$ in (21a) **do**
    Compute central residual r_0 as in (19)
    Compute corresponding thresholds ϵ_0^j as in (20)
    if $|\mathrm{r}_0^j| > \epsilon_0^j$ **then**
      Reject the null hypothesis $\mathcal{H}^{\mathrm{o}}$ in (21b) ▷ Stealthy attack is detected globally at the control center.
    end if
  end while
end procedure



(a) UAVs' position in the $x-y$ plane.

(b) Coordination of UAVs in the y direction.

Fig. 4. Experiment 1: ZDA on UAVs $1, 4, 5$ and topology switching from mode 1 to 4. (a) UAVs' position in the x–y plane with the colorbars quantifying the time span. The $\times$ markers and the colored circles show, respectively, the UAVs' initial position and final position during the experiment. Finally, the gray lines visualize the V-shape formation achieved by the final position of the UAVs. (b) The relative positions of UAVs in the y direction corresponding to the inter-UAV communication links in mode $\sigma(t) = 1$, shown in Figure 3c. Also, the dashed lines, labeled by $\mathrm{p}^*_{ij},\ i, j \in \mathcal{V}$, denote the desired relative positions based on the formation references in Figure 3a.

It is worth mentioning that a trivial solution for (27) is $\mathcal{D} = \mathcal{V}$ that is all of the UAVs are equipped with a local detector, although this set can be optimally selected.

Given Propositions III.2 and III.3, one can verify that the networked UAVs can be secured against stealthy attacks using a set of local monitors, given by (24), that locally detect stealthy attacks, addressing problem (26). A procedure for this local hypothesis testing will be presented in Algorithm 1 in Section IV.

IV. Experimental Results

We conducted a set of experiments that serve two purposes. First, the evaluation of the stealthiness of intrusions/deception attacks, described in Section III-A, on the wireless communication network of a team of quadrotor UAVs in real-time practical settings. Second, the performance evaluation of the detection schemes presented in Section III-B.

A. Experimental Setup

Our experimental setup consists of a team of five homogeneous quadrotors (Tello Drones[5]), shown in Figure 1, flying in a $6\ \mathrm{m} \times 4\ \mathrm{m} \times 3\ \mathrm{m}$ flight area that is equipped with a VICON[6] motion capture system with 10 cameras. The VICON system provides the ground truth position and orientation of each UAV at 50 Hz for a central PC running Ubuntu 20.04 with ROS Noetic.

In our experiments, we use the VICON system's ground truth data available in the central PC to compute the high-level formation control commands that are sent to each UAV at 50 Hz and to run the central and local monitors, presented in Section III-B. The central PC transmits the high-level formation control commands to the UAVs through different Wi-Fi channels and the UAVs' on-board attitude controllers use the received control commands to stabilize and steer the UAVs to their desired pose (see Figure 2b). This connection setup allows us to replicate the peer-to-peer communication of the UAVs and also to implement stealthily intrusions on the Wi-Fi channels in a controlled setting.

[5]https://www.ryzerobotics.com/tello.

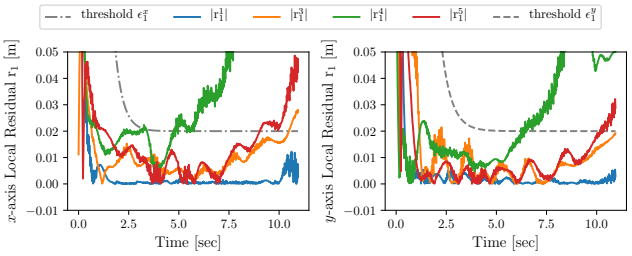
[6]https://www.vicon.com.

B. Results

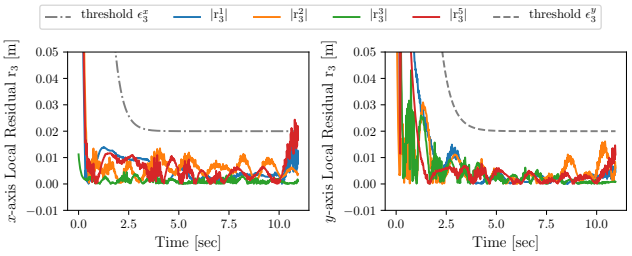
We performed several experiments that serve as a proof of concept of the real-world applicability of the proposed attack detection methods in multi-UAV cooperation settings. In these tests, the UAVs are tasked with achieving a V-shape formation. The spacial configuration of the V-shape formation and a picture of its real-world implementation are shown in Figures 3a and 3b, respectively.

In our experiments, the UAVs, indexed by $\mathcal{V} = \{1, 2, 3, 4, 5\}$, coordinate using the control protocol (8), initially in communication mode $\sigma(t) = 1$, shown in Figure 3c, to achieve the V-shape formation. The UAVs also have consensus on their yaw angle $\psi_i = \psi^* = 0,\ \forall i \in \mathcal{V}$ as well as their hovering altitude. We select the position of UAVs 3 and 5 as the network-level monitored states at the ground control center that is $\mathcal{M}_\mathrm{p} = \{3, 5\}$ and $\mathcal{M}_\mathrm{v} = \emptyset$ in (11). These measurements are used in the realization of the central monitor (attack detector $\Sigma^{\mathcal{M}}_{\mathcal{O}}$) in (19) and its residuals in (20). We also let $\mathcal{D} = \{1, 3\}$ in (27), that is UAVs 1 and 3, which have the respective set of neighbors $\mathcal{N}^1_{\sigma(t)=1} = \{3, 4, 5\}$ and $\mathcal{N}^3_{\sigma(t)=1} = \{1, 2, 5\}$, and the local measurements $\mathbf{y}_i (\text{or}\, \mathrm{y}_i),\ i \in \mathcal{D}$ in (23), are selected as the host UAVs for local monitors (attack detectors $\Sigma^i_{\mathcal{O}}$'s) in (24). Accordingly, the condition (27) holds which in turn guarantees the local monitors of UAVs 1 and 3 are sufficient to locally detect the stealthy attacks on the entire network of UAVs in a decentralized manner. Also, as described earlier in Sections II-C, the UAVs follow a decoupled dynamics in the x and y directions, and therefore we implement central and local monitors independently for the x- and y-direction dynamics based on the discretized models of (12), (19), and (24) with the sampling time $T_\mathrm{s} = 0.02$ sec.

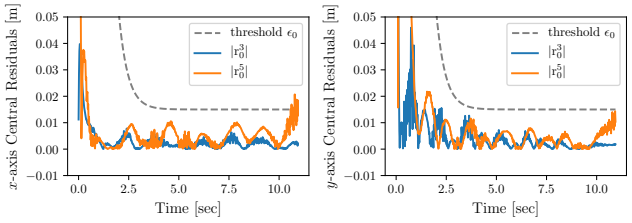
In the following, we present the results of attack detection through the centralized detection scheme with central (global) hypothesis testing (21) and the central monitor (19) as well as through the decentralized detection scheme with local hypothesis testing (26) and the local monitors of UAVs



(a) Residuals of local monitor $\Sigma_{\mathcal{O}}^1$ run on UAV 1. The stealthy ZDA is locally detected at $t = 3.22$ sec.

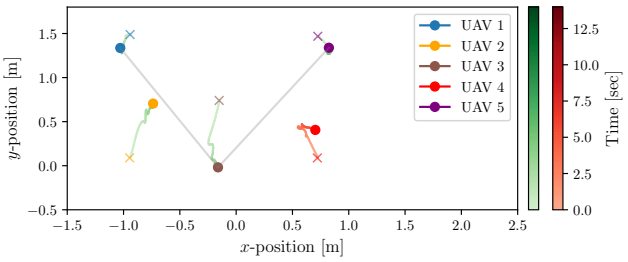


(b) Residuals of local monitor $\Sigma_{\mathcal{O}}^3$ run on UAV 3.

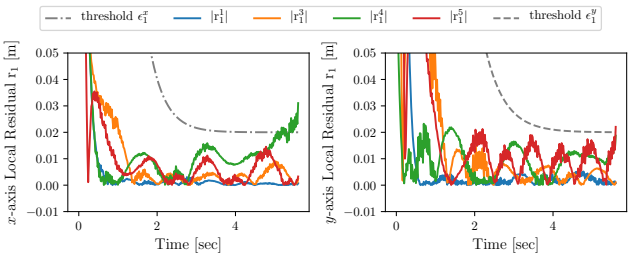


(c) Residuals of central monitor $\Sigma_{\mathcal{O}}^{\mathcal{M}}$ run on the control center.

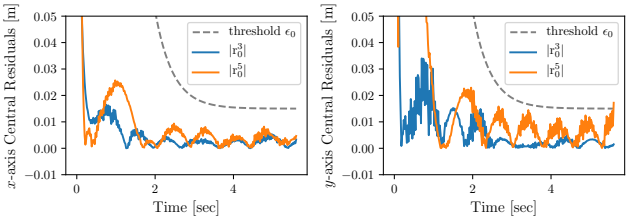
Fig. 5. Experiment 1: ZDA on UAVs $1, 4, 5$ and topology switching from mode 1 to 4, which is triggered by local monitor $\Sigma_{\mathcal{O}}^1$ at $t = 3.22$ sec. (a)-(b) The notation r_1^i, $i \in \{1, 3, 4, 5\}$ (r_3^i, $i \in \{1, 2, 3, 5\}$), denotes the residual of position estimation for the UAV 1's (3's) neighbors obtained by its local monitor $\Sigma_{\mathcal{O}}^1$ ($\Sigma_{\mathcal{O}}^3$) in the x and y directions with the respective thresholds ϵ_1^x (ϵ_3^x) and ϵ_1^y (ϵ_3^y) as given in (25). (c) The notation r_0^i, $i \in \{3, 5\}$, denotes the residual of position estimation for UAVs 3 and 5 by the central monitor $\Sigma_{\mathcal{O}}^{\mathcal{M}}$ in the x and y directions with the threshold ϵ_0 as given in (20).



(a) UAVs' position in the x–y plane. The central monitor $\Sigma_{\mathcal{O}}^{\mathcal{M}}$ detects the ZDA at $t = 5.6$ sec and ends the experiment.



(b) Residuals of local monitor $\Sigma_{\mathcal{O}}^1$ run on UAV 1. The stealthy ZDA is locally detected at $t = 5.08$ sec.



(c) Residuals of central monitor $\Sigma_{\mathcal{O}}^{\mathcal{M}}$ run on the control center. The stealthy ZDA is detected at $t = 5.6$ sec using Algorithm 2.

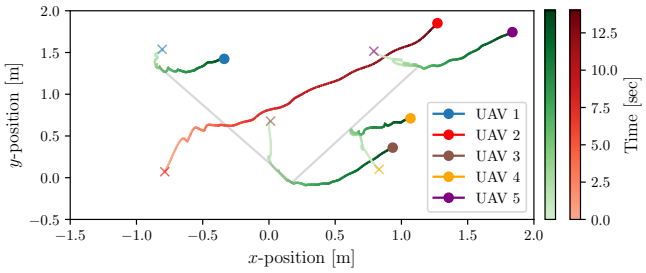
Fig. 6. Experiment 2: ZDA on UAVs $1, 4, 5$ and topology switching from mode 1 to 3, which is triggered by local monitor $\Sigma_{\mathcal{O}}^1$ at $t = 5.08$ sec. (a) UAVs' position in the $x-y$ plane with the same annotations as in Figure 4a. (b)-(c) The residuals of local monitor $\Sigma_{\mathcal{O}}^1$ and central monitor $\Sigma_{\mathcal{O}}^{\mathcal{M}}$ with the same annotations as in Figures 5a and 5c, respectively.

1 and 3. The procedure of the local hypothesis testing is presented in Algorithm 1 and that of the central hypothesis testing is presented in Algorithm 2.

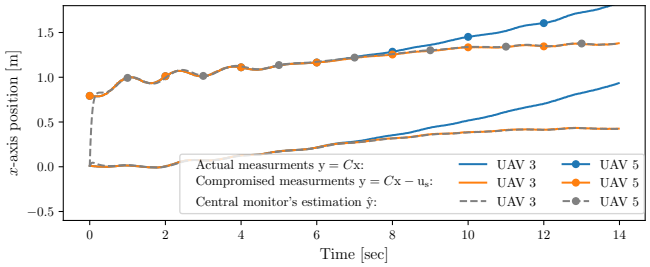
1) Stealthy zero-dynamics attack: We conducted two experiments evaluating the effectiveness of the central and local monitors in detection of stealthy zero-dynamics attack (ZDA). In the first experiment, UAVs 1, 4, and 5 are compromised such that their control channels are subject to the discretized version of ZDA signals in (16) as $\mathbf{u}_{\mathrm{a}} = \mathrm{col}\,(\boldsymbol{u}_{\mathrm{a}_i})_{i\in\mathcal{A}}$, $\mathcal{A} = \{1, 4, 5\}$ with $\boldsymbol{u}_{\mathrm{a}_i} = [\mathrm{u}_{\mathrm{a}_i}^x(0)e^{\lambda_o^x(kT_{\mathrm{s}})} \quad \mathrm{u}_{\mathrm{a}_i}^y(0)e^{\lambda_o^y(kT_{\mathrm{s}})}\ ]^\top$, $\lambda_o^x = \lambda_o^y = 0.5$, $k \in \mathbb{Z}_{\geq 0}$, $\mathrm{u}_{\mathrm{a}}^x(0) = [\mathrm{u}_{\mathrm{a}_1}^x(0) \ \mathrm{u}_{\mathrm{a}_4}^x(0) \ \mathrm{u}_{\mathrm{a}_5}^x(0)]^\top = [-2.34\mathrm{x}_4^{\mathrm{a}}(0) \quad 10.24\mathrm{x}_4^{\mathrm{a}}(0) \quad -2.34\mathrm{x}_4^{\mathrm{a}}(0)]^\top$, $\mathrm{u}_{\mathrm{a}}^y(0) = -0.7\mathrm{u}_{\mathrm{a}}^x(0)$, $\mathrm{x}_4^{\mathrm{a}}(0) = 0.0086$, and the starting time $t = (kT_{\mathrm{s}}) = 0$, $k = 0$. The network-level measurements (11) with $\mathcal{M}_{\mathrm{p}} = \{3, 5\}$ and $\mathcal{M}_{\mathrm{v}} = \emptyset$, on the other hand, are not subject to sensory attacks that is $\mathbf{u}_{\mathrm{s}} = \mathbf{0}$.

In the first experiment, as shown in Figure 4a, all the UAVs start from some initial positions and coordinate to achieve the desired formation while the stealthy ZDA steers UAV 4 away from its desired configuration that meets (6). This effect has been illustrated in Figure 4b showing the relative positions of the UAVs as well as their desired values in the y direction over time. It is necessary to note that UAV 4 hits the safety net enclosing the indoor flight area at $t \approx 9.8$ sec.

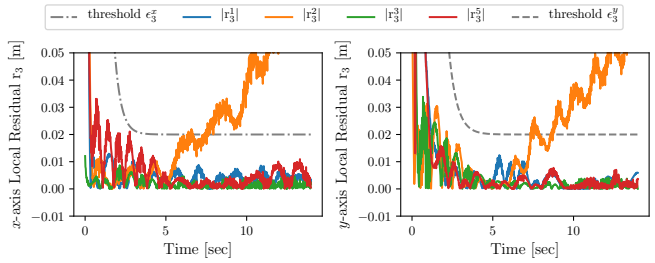
In terms of attack detection, Figures 5a and 5b show the residuals of local monitors $\Sigma_{\mathcal{O}}^i$'s, $i \in \{1, 3\}$ in (24) for UAVs 1 and 3, respectively. Also, Figure 5c shows the residuals of the central monitor $\Sigma_{\mathcal{O}}^{\mathcal{M}}$ in (19) available in the control center. One can verify that the local monitor of UAV 1, $\Sigma_{\mathcal{O}}^1$, running Algorithm 1, has detected the stealthy ZDA in a timely manner ($t = 3.22$ sec) that is before UAV 4 collides with the safety net of the flight area at $t \approx 9.8$ sec. However, the ZDA remains stealthy in the residuals of the UAV 3's local monitor, $\Sigma_{\mathcal{O}}^3$, and those of the central monitor running Algorithm 2, regardless of the switch in the inter-UAV's communication topology from mode 1 to mode 4 (see Figure 3) that is triggered by the local monitor $\Sigma_{\mathcal{O}}^1$ at $t = 3.22$ sec. This is due to the fact that switching from



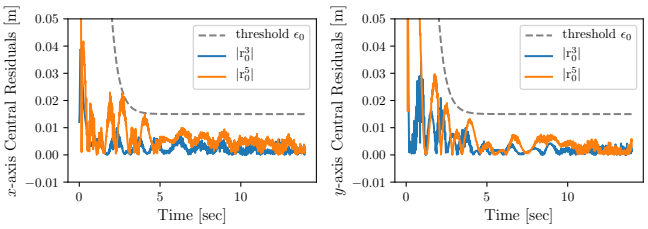
(a) UAVs' position in the $x-y$ plane.



(b) Realization of stealthiness condition (13) using $\mathbf{u}_\mathrm{s}$ in (17) with the starting time $t_\mathrm{a} = 5$ sec.



(c) Residuals of local monitor $\Sigma_\mathcal{O}^3$ run on UAV 1. The stealthy ZDA is detected at $t = 6.4$ sec using Algorithm 1.



(d) Residuals of central monitor $\Sigma_\mathcal{O}^\mathcal{M}$ run on the control center.

Fig. 7. Experiment 3: covert attack on UAV 2 and topology switching from mode 1 to 2, which is triggered by local monitor $\Sigma_\mathcal{O}^1$ at $t = 6.4$ sec. (a) UAVs' position in the $x-y$ plane with the same annotations as in Figure 4a, except the gray lines that visualize the V-shape formation achieved by the UAVs at $t_\mathrm{a} = 5$ sec, the starting time of the covert attack. (b) The effect of measurement alteration using sensory attack $\mathbf{u}_\mathrm{s}$ starting at $t_\mathrm{a} = 5$ sec. (c)-(d) The residuals of local monitor $\Sigma_\mathcal{O}^3$ and central monitor $\Sigma_\mathcal{O}^\mathcal{M}$ with the same annotations as in Figures 5a and 5c, respectively.

mode 1 to mode 4 does not meet the necessary conditions required for a topology switching to render stealhy attacks detectable for the central monitor in (19). The details of such conditions have been studied in [27, Th. III.3].

In the second experiment, with the results shown in Figure 6, the UAVs are under the same ZDA as in the first experiment expect $\mathrm{x}_4^\mathrm{a}(0) = 0.012$. The local monitor $\Sigma_\mathcal{O}^1$ successfully detects the ZDA at $t = 5.08$ sec (see Figure 6b) and then triggers a switch in the UAVs' communication from mode 1 to mode 3 (cf. Figure 3) that as opposed to Experiment 1, this topology switching results in detection of stealthy ZDA by the central monitor $\Sigma_\mathcal{O}^\mathcal{M}$ at $t = 5.6$ sec (see Figure 6c).

It is necessary to note that any topology switching in the inter-UAVs communications results in a discrepancy between the actual dynamics of networked UAVs and its nominal counterpart that is used by the attacker to design stealthy attacks. Yet, the model discrepancy in Experiment 1 did not interfere with the stealthiness of ZDA in the central monitor's residuals while it renders ZDA detectable in the central monitor's residuals in Experiment 2. These results indicate that not only zero-dynamics attacks (ZDA) can be implemented in real-time on networked UAVs with partial measurements, they also can remain stealthy regardless of switches in the inter-UAVs' communication topology. Theoretical results to detect stealthy ZDA through topology switching in networked systems with full-state measurements and with partial measurements can be found, respectively, in [9] and [27].

2) Covert attack: Similar to the ZDA case, we evaluated the detection of covert attack on networked UAVs subject to topology switching by using the local monitors $\Sigma_\mathcal{O}^1$ and $\Sigma_\mathcal{O}^3$, and the central monitor $\Sigma_\mathcal{O}^\mathcal{M}$. In this experiment, a covert attack, $\mathbf{u}_{a_i}$, $i \in \mathcal{A} = \{2\}$, in the form of a ramp signal with a slope of 3 Deg, as the roll and pitch angles' perturbation, and the starting time of $t_\mathrm{a} = 5$ sec is injected through the control channel of UAV 2. The covert attack's effect on the UAVs' formation is shown in Figure 7a. As illustrated, all of the UAVs have deviated from their desired formation configuration that meets (6). The effect of this deviation/perturbation on the measurements $\mathbf{y}$ in (11) is simultaneously canceled out by implementing the discretized version of the sensory attack $\mathbf{u}_\mathrm{s}$ given in (17). Figure 7b illustrates how the alteration of actual measurement $\mathbf{y}$ of the monitored UAV 3 using the sensory attack $\mathbf{u}_\mathrm{s}$ gives rise to a false state estimation by the central monitor $\Sigma_\mathcal{O}^\mathcal{M}$, rendering the injected attack covert in the central residuals. The local monitors, however, are not subject to such alterations and thus are capable of detecting the covert attack in a timely manner as shown in Figure 7c for the local monitor $\Sigma_\mathcal{O}^3$ of UAV 3. We note that the local monitor $\Sigma_\mathcal{O}^3$ triggers a topology switch from mode 1 to mode 2 (cf. Figure 3) at $t = 6.4$ sec to make the covert attack detectable in the residuals of the central monitor $\Sigma_\mathcal{O}^\mathcal{M}$. However, the attack remains stealthy in the central residuals, shown in Figure 7d, regardless of topology switching. The results, consistent with those in the ZDA case, shows the outperformance of the decentralized detection scheme (Algorithm 1) over the centralized detec-

tion scheme (Algorithm 2). It is worth mentioning that one can leverage a larger number of switching communication links on which the centralized monitor relies to improve the performance of the centralized detection scheme. However, this solution raises other challenges such as switching-induced unobservability as well as communication overhead. In the case of the decentralized detection scheme that relies on the network model of UAVs, scalability is a concern for larger teams of UAVs, for which clustering-based solutions such as the one in [27] can be applied.

V. Conclusions

In this paper, we studied the detection of sophisticated stealthy attacks including zero-dynamics and covert attacks on networked UAVs with switching communication topology in formation control settings. Centralized and decentralized attack detection schemes were developed to detect the stealthy attacks. We performed indoor flight tests using quadrotor UAVs to evaluate the effectiveness of our proposed methods. The experiments showed feasibility of stealthy attacks on UAV networks which was previously investigated only in theory or simulations. The experiments also validated our theoretical developments and the effectiveness of the attack detection schemes proposed in this paper. The experimental results demonstrated the importance of timely detection of stealthy attacks and trade-offs existing in centralized vs. decentralized attack detection. Future work will be devoted to extending the results to more complex cooperative settings where we will incorporate cooperative attack mitigation and mission adaptation.

Acknowledgment

This research is partially supported by the National Science Foundation (award no. 2137753).

References

[1] S.-J. Chung, A. A. Paranjape, P. Dames, S. Shen, and V. Kumar, "A survey on aerial swarm robotics," *IEEE Transactions on Robotics*, vol. 34, no. 4, pp. 837–855, 2018.

[2] K.-K. Oh, M.-C. Park, and H.-S. Ahn, "A survey of multi-agent formation control," *Automatica*, vol. 53, pp. 424–440, 2015.

[3] J. Cortes, S. Martinez, T. Karatas, and F. Bullo, "Coverage control for mobile sensing networks," *IEEE Transactions on robotics and Automation*, vol. 20, no. 2, pp. 243–255, 2004.

[4] K. Harikumar, J. Senthilnath, and S. Sundaram, "Multi-uav oxyrrhis marina-inspired search and dynamic formation control for forest firefighting," *IEEE Transactions on Automation Science and Engineering*, vol. 16, no. 2, pp. 863–873, 2018.

[5] P. Tokekar, J. Vander Hook, D. Mulla, and V. Isler, "Sensor planning for a symbiotic uav and ugv system for precision agriculture," *IEEE Transactions on Robotics*, vol. 32, no. 6, pp. 1498–1511, 2016.

[6] M. R. Manesh and N. Kaabouch, "Cyber-attacks on unmanned aerial system networks: Detection, countermeasure, and future research directions," *Computers & Security*, vol. 85, pp. 386–401, 2019.

[7] J.-P. Yaacoub, H. Noura, O. Salman, and A. Chehab, "Security analysis of drones systems: Attacks, limitations, and recommendations," *Internet of Things*, vol. 11, p. 100218, 2020.

[8] G. Choudhary, V. Sharma, I. You, K. Yim, R. Chen, and J.-H. Cho, "Intrusion detection systems for networked unmanned aerial vehicles: a survey," in *2018 14th International Wireless Communications & Mobile Computing Conference (IWCMC)*. IEEE, 2018, pp. 560–565.

[9] 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*, 2020.

[10] R. S. Smith, "Covert misappropriation of networked control systems: Presenting a feedback structure," *IEEE Control Systems Magazine*, vol. 35, no. 1, pp. 82–92, 2015.

[11] Y. Mo and B. Sinopoli, "Secure control against replay attacks," in *2009 47th annual Allerton conference on communication, control, and computing (Allerton)*. IEEE, 2009, pp. 911–918.

[12] C. Kwon, W. Liu, and I. Hwang, "Analysis and design of stealthy cyber attacks on unmanned aerial systems," *Journal of Aerospace Information Systems*, vol. 11, no. 8, pp. 525–539, 2014.

[13] P. Dash, M. Karimibiuki, and K. Pattabiraman, "Out of control: stealthy attacks against robotic vehicles protected by control-based techniques," in *Proceedings of the 35th Annual Computer Security Applications Conference*, 2019, pp. 660–672.

[14] F. Fei, Z. Tu, R. Yu, T. Kim, X. Zhang, D. Xu, and X. Deng, "Cross-layer retrofitting of uavs against cyber-physical attacks," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 550–557.

[15] Y.-C. Liu, G. Bianchin, and F. Pasqualetti, "Secure trajectory planning against undetectable spoofing attacks," *Automatica*, vol. 112, p. 108655, 2020.

[16] A. K. Bozkurt, Y. Wang, and M. Pajic, "Secure planning against stealthy attacks via model-free reinforcement learning," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 10 656–10 662.

[17] L. Liu, H. Wu, Z. Xi, and Y. Cui, "Fdi attack detection for formation control of quantized uav systems by coding sensor outputs," in *2020 International Conference on Unmanned Aircraft Systems (ICUAS)*. IEEE, 2020, pp. 1278–1285.

[18] C. Liang, M. Miao, J. Ma, H. Yan, Q. Zhang, and X. Li, "Detection of global positioning system spoofing attack on unmanned aerial vehicle system," *Concurrency and Computation: Practice and Experience*, p. e5925, 2020.

[19] L. Negash, S.-H. Kim, and H.-L. Choi, "Distributed observes for cyberattack detection and isolation in formation-flying unmanned aerial vehicles," *Journal of Aerospace Information Systems*, vol. 14, no. 10, pp. 551–565, 2017.

[20] E. M. Khanapuri, R. Sharma, and K. Brink, "Learning-based detection of stealthy false data injection attack applied to cooperative localization problem," in *AIAA SCITECH 2022 Forum*, 2022, p. 2543.

[21] H. Ahn, "Deep learning based anomaly detection for a vehicle in swarm drone system," in *2020 International Conference on Unmanned Aircraft Systems (ICUAS)*. IEEE, 2020, pp. 557–561.

[22] R. Mahony, V. Kumar, and P. Corke, "Multirotor aerial vehicles: Modeling, estimation, and control of quadrotor," *IEEE Robotics and Automation magazine*, vol. 19, no. 3, pp. 20–32, 2012.

[23] W. Ren and N. Sorensen, "Distributed coordination architecture for multi-robot formation control," *Robotics and Autonomous Systems*, vol. 56, no. 4, pp. 324–333, 2008.

[24] M. Turpin, N. Michael, and V. Kumar, "Decentralized formation control with variable shapes for aerial robots," in *2012 IEEE international conference on robotics and automation*. IEEE, 2012, pp. 23–30.

[25] W. Ren, "Consensus strategies for cooperative control of vehicle formations," *IET Control Theory & Applications*, vol. 1, no. 2, pp. 505–512, 2007.

[26] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson, "A secure control framework for resource-limited adversaries," *Automatica*, vol. 51, pp. 135–148, 2015.

[27] M. Bahrami and H. Jafarnejadsani, "Privacy-preserving stealthy attack detection in multi-agent control systems," in *2021 60th IEEE Conference on Decision and Control (CDC)*, 2021, pp. 4194–4199.

[28] J. Chen, Z. Feng, J.-Y. Wen, B. Liu, and L. Sha, "A container-based dos attack-resilient control framework for real-time uav systems," in *2019 Design, Automation & Test in Europe Conference & Exhibition (DATE)*. IEEE, 2019, pp. 1222–1227.