Stealthy Attacks in Cloud-Connected Linear Impulsive Systems

Alessandra Duz, Sean Phillips, Adriano Fagiolini, Ricardo G. Sanfelice, and Fabio Pasqualetti

***Abstract*— This paper studies a security problem for a class cloud-connected multi-agent systems, where autonomous agents coordinate via a combination of short-range ad-hoc communication links and long-range cloud services. We consider a simplified model for the dynamics of a cloud-connected multi-agent system and attacks, where the states evolve according to linear time-invariant impulsive dynamics, and attacks are modeled as exogenous inputs designed by an omniscent attacker that alters the continuous and impulsive updates. We propose a definition of attack detectability, characterize the existence of stealthy attacks as a function of the system parameters and attack properties, and design a family of undetectable attacks. We illustrate our results on a cloud-based surveillance example.**

I. Introduction

Distributed systems and networks are the building blocks of smart and citizen-centric services in modern urban environments. Due to their very nature, these cyber-physical systems offer open and physically accessible interfaces on both their cyber side (e.g., network interfaces and control algorithms) and their physical side (e.g., sensors and actuators), which can be exploited by capable adversaries to deny control, disable alarms, manipulate sensors, and initiate actions to cause physical damage. Examples are unfortunately abundant, e.g., see [1], [2], [3], [4], showing that it is likely unfeasible to secure all attack surfaces completely, and that the emphasis of defense must be on careful design, timely detection and localization, and reactive controls.

Security is even a more imminent threat for the class of cyber-physical systems arising from the integration of autonomous units with cloud-based technologies. Examples include HVAC control [5], industrial automation [6], [7], [8], assistive robotics [9], [10], and intelligent transportation systems, where cloud computing enables, for instance, detection and prevention of incidents, localization, and fast rerouting [11], [12], [13], [14]. Novel security theories and tools are needed for these systems. In fact, typical security methods that rely on purely cyber mechanisms, such as data protection and authentication, or on anomaly detection techniques based on simple representations of the physical dynamics, are likely unable to predict and prevent coordinated attacks leveraging complex interactions among cyber and physical components.

In this work we consider a class of cloud-connected systems with linear-impulsive dynamics, and model attacks as exogenous inputs altering both the continuous and impulsive dynamics. In addition to allowing us to formally analyze attacks, our modeling framework is also capable of representing the effect of several attack strategies against cloud-connected systems, including man-in-the-middle attacks, malware injection, and authentication attacks. We propose a notion of attack detectability, and use tools from geometric control theory [15], [16], [17] to characterize the existence and engineer a family of stealthy attacks.

Related work With security emerging as a major concern for cyber-physical systems, different modeling frameworks and protection schemes have been proposed for a variety of systems and attacks. While early works focus on static representations [18], [19], game-theoretic [20], [21], information theoretic [22], [23], and control-theoretic methods [24], [25], [26], [27] have been developed for dynamic models and attacks. These approaches represent a step toward addressing dynamic security features, and the threshold for the new fundamental approach proposed here. To the best of our knowledge, most papers study detection, identification, and resilience for systems with linear dynamics and attacks compromising integrity or availability of resources [28]. Yet, as systems evolve and become more complex, security methods based on simple dynamic models will likely be inapplicable or ineffective in practical scenarios. New security methods are needed for systems with coupled cyber and physical dynamics, and constraints on the utilization of resources and timing. In particular, despite general theoretical developments [29] and recent results [30], security for systems featuring hybrid dynamics remains a largely unexplored area.

Contributions of the paper The main contributions of this paper are as follows. First, we propose a modeling framework for a class of cloud-connected multi-agent systems under attack, where the states evolve according to linear-impulsive dynamics and attacks are modeled as exogenous inputs to the continuous and impulsive dynamics. Although we restrict our analysis to linear dynamics, the proposed framework captures a broad class of coordination algorithms, and different attacks enabled by cloud communication or physical interaction. Second, we introduce a notion of attack detectability for linear-impulsive systems, and characterize the existence of undetectable attacks as a function of the system and attack parameters. Third and finally, we design a family of undetectable attacks and validate its effectiveness on a system describing a cloud-based surveillance scenario.

This material is based upon work supported in part by NSF CAREER ECS-1450484, NSF ECS-1710621, NSF CNS-1544396, AFOSR FA9550-16-1-0015, and AFOSR FA9453-16-1-0053, in part by CITRIS, the Banatao Institute at the University of California, ARO 71603NSYIP, and NSF ECCS-1405330, and in part by CORI2016 from the University of Palermo. Alessandra Duz and Fabio Pasqualetti are with the Mechanical Engineering Department, University of California at Riverside, {alessd, fabiopas}@engr.ucr.edu. Adriano Fagiolini is with the Department of Energy, Computer Science, and Mathematical Models, University of Palermo, Italy, fagiolini@unipa.it. Sean Phillips and Ricardo G. Sanfelice are with the Computer Engineering Department, University of California at Santa Cruz, {seaphill, ricardo}@ucsc.edu.

Our approach relies on tools from geometric control theory.

Paper organization The rest of the paper is organized as follows. Section II contains our setup and preliminary notions. Section III contains our characterization and design of undetectable attacks. Finally, Section IV contains our numerical examples, and Section V concludes the paper.

II. Problem setup and preliminary notions

We consider a cyber-physical system with the following linear-impulsive dynamics:

$$\begin{cases} \dot{x}(t) = A_{\mathrm{c}}\, x(t) + B_{\mathrm{c}}\, u(t)\,, \\ y(t) = C_{\mathrm{c}}\, x(t)\,, \end{cases} \quad \text{for } t \in \mathbb{R}_{\geq 0} \setminus \mathcal{T}\,, \quad (1)$$

$$\begin{cases} x(t) = A_{\mathrm{i}}\, x(t^-) + B_{\mathrm{i}}\, u(t)\,, \\ y(t) = C_{\mathrm{i}}\, x(t^-)\,, \end{cases} \quad \text{for } t \in \mathcal{T}\,, \quad (2)$$

where $x : \mathbb{R} \to \mathbb{R}^n$ is the state vector, $u : \mathbb{R} \to \mathbb{R}^m$ is the attack input (see below), and $\mathcal{T} = \{\tau_1, \tau_2, \dots\}$ is the set of *jump* times, which satisfy $\tau_1 \geq \tau_{\min}$ and $\tau_{k+1} - \tau_k \geq \tau_{\min}$ where $\tau_0 = 0$. That is, $\tau_{\min} > 0$ is the minimum time between any two consecutive jumps. The notation $x(t^-)$ stands for $x(t^-) = \lim_{\varepsilon \to 0^+} x(t-\varepsilon)$. We are interested in characterizing the existence of *stealthy* attacks for the system (1) – (2); namely, attacks that cannot be detected by any monitor through the available measurements. To this aim, we consider an omniscient attacker that (i) knows the system matrices A_{c}, C_{c}, A_{i}, and C_{i}, and (ii) has infinite computational power. Further, we assume that (iii) the attacker knows the values of $\tau_{\min}$ and recognizes jump times when they happen, but does not know the set $\mathcal{T}$ a priori. While assumptions (i) and (ii) are motivated by our worst-case perspective, (iii) reflects the fact that jump times $\mathcal{T}$ are determined by the interaction between the physical system and the cloud. These instants can be measured, but they are not known in advance. This assumption limits the ability of the attacker to implement feed-forward policies to compensate for the impulsive updates.

The matrices B_{c}, B_{i} and the input u are an abstract representation of the attack strategy, which is convenient for an analytical study of attack detectability, but they are general enough to represent a large class of attacks against cloud-connected cyber-physical systems. A similar framework has already proven useful to study attacks in non-impulsive systems [31].

Remark 1: ***(Attacks against cloud-connected cyber-physical systems modeled by our framework)*** As we show in Section IV for the case of cloud-connected agents for urban surveillance, several cloud-connected systems can be modeled by the equations (1) – (2). For these systems, attacks represented by additive inputs include malware injection, authentication, and man-in-the-middle attacks. Injection of malware into the cloud results in an alteration of the cyber services, and thus in a modification of the impulsive update that can be modeled as an appropriate input as in (2). In an authentication attack, the attacker possesses authorized credentials to access cloud services, and can manipulate the information processed by the cloud to induce incorrect computation. Authentication attacks can be modeled by inputs to the state and output equations of the impulsive dynamics. Finally, in a man-in-the-middle attack the attacker tampers with the information exchanged between the physical agents and the cloud, or between different physical agents. Thus, man-in-the-middle attacks can be modeled as exogenous inputs to the continuous and impulsive dynamics. Attacks affecting the output equations are not considered here, although the analysis can be easily extended to include this case. Other attacks in cyber-physical systems modeled as unknown inputs are described in [31]. □

In this paper, we characterize detectability of attacks against systems with dynamics (1) - (2). To reveal fundamental detectability limitations that are independent of the choice of monitor, we assume that the monitor is any algorithm that can be constructed with knowledge of the system matrices A_{c}, C_{c}, A_{i}, and C_{i}, and of the measurements y at all times. Further, we assume that the monitor recognizes the jump times. We adopt the following general definition of detectability of attacks, where $y(x_0, B_{\mathrm{c}}, B_{\mathrm{i}}, u, \mathcal{T}, t)$ denotes the output signal at time t of the system (1) – (2) with initial state x_0, jump times $\mathcal{T}$, input matrices B_{c}, B_{i}, and input u.

Definition 1: ***(Detectability of attacks)*** The attack $(B_{\mathrm{c}}, B_{\mathrm{i}}, u)$ against the system (1) – (2) with initial condition x_0 is undetectable if

$$y(\bar{x}_0, 0, 0, 0, \mathcal{T}, t) = y(x_0, B_{\mathrm{c}}, B_{\mathrm{i}}, u, \mathcal{T}, t) \quad (3)$$

at all times $t \in \mathbb{R}_{\geq 0}$, for some initial state $\bar{x}_0$ □.

In other words, an attack is undetectable if the continuous and impulsive evolutions generate measurements that are equal to those generated by the autonomous system with (possibly) a different initial condition and without attack. Accordingly, we consider an attack to be detectable when (3) is violated. In such case, a possible algorithm to detect the attack is described in [30].

We conclude this section by recalling some concepts from geometric control theory that will be used throughout the paper. We refer the interested reader to [15], [32], [33] for a comprehensive treatment of this subject. For a linear time-invariant system with matrices (A, B, C) (continuous or discrete time), the subspace $\mathcal{V}$ is called controlled invariant if the state trajectory can be maintained in $\mathcal{V}$ by a suitable control signal or, equivalently, if there exists a matrix F satisfying $(A + BF)\mathcal{V} \subseteq \mathcal{V}$. The largest output-nulling reachable subspace is the largest subspace of the state space that can be reached from the origin with state trajectories belonging to the null space of the output matrix C. Finally, we use $\mathrm{Im}(M)$ to denote the image of the matrix M, $\mathrm{Ker}(M)$ denotes the null space of M, $M^{-1}\mathcal{V}$ denotes the pre-image of the subspace $\mathcal{V}$ through M, and $\mathrm{Basis}(\mathcal{V})$ is the basis of $\mathcal{V}$.

III. Undetectable attacks in linear-impulsive cyber-physical systems

In this section, we present necessary and sufficient conditions for the existence of undetectable attacks for the system (1) – (2), and characterize a class of undetectable attacks. The following result creates a link between undetectable attacks

and output-nulling inputs [15] of the system (1) – (2), and extends the result in [25] to linear-impulsive systems.

Lemma 3.1: ***(Undetectable attacks and output-nulling inputs)*** The attack $(B_{\rm c}, B_{\rm i}, u)$ against the system (1)-(2) is undetectable if and only if

$$y(\tilde{x}_0, B_{\rm c}, B_{\rm i}, u, \mathcal{T}, t) = 0 \tag{4}$$

at all times $t \in \mathbb{R}_{\geq 0}$, for some initial state $\tilde{x}_0$.

Proof: Let x_0 be the initial state of the system and define $\bar{x}_0 = x_0 - \tilde{x}_0$. Recall that $\mathcal{T} = \{\tau_1, \tau_2, \dots\}$ is the set of jump times. Because (1)-(2) is linear in the interval $[0, \tau_1)$, we have

$$y(x_0, B_{\rm c}, B_{\rm i}, u, \mathcal{T}, t) = y(\bar{x}_0, 0, 0, 0, \mathcal{T}, t) + \underbrace{+ C_{\rm c} e^{A_{\rm c} t} \tilde{x}_0 + C_{\rm c} \int_0^t e^{A_{\rm c}(t-\tau)} B_{\rm c} u(\tau)\, d\tau}_{y(\tilde{x}_0, B_{\rm c}, B_{\rm i}, u, \mathcal{T}, t)},$$

and for $t = \tau_1$ we have

$$y(x_0, B_{\rm c}, B_{\rm i}, u, \mathcal{T}, \tau_1) = y(\bar{x}_0, 0, 0, 0, \mathcal{T}, \tau_1) + \underbrace{+ C_{\rm i} e^{A_{\rm c} \tau_1^-} \tilde{x}_0 + C_{\rm i} \int_0^{\tau_1^-} e^{A_{\rm c}(\tau_1^- - \tau)} B_{\rm c}\, u(\tau)\, d\tau}_{y(\tilde{x}_0, B_{\rm c}, B_{\rm i}, u, \mathcal{T}, \tau_1)}.$$

Thus, if (3) holds, then

$$y(\tilde{x}_0, B_{\rm c}, B_{\rm i}, u, \mathcal{T}, t) = 0$$

which provides the sufficiency of the statement for $t \in [0, \tau_1]$. Conversely, if (4) holds, then

$$y(\bar{x}_0, 0, 0, 0, \mathcal{T}, t) = y(x_0, B_{\rm c}, B_{\rm i}, u, \mathcal{T}, t)$$

which proves the necessity of the statement for $t \in [0, \tau_1]$.

Similarly, it can be shown that the state after the first impulsive update is $\tilde{x}(\tau_1) = x(\tau_1) - \bar{x}(\tau_1)$, where $\bar{x}(\tau_1)$ is the state of (1)-(2) in the absence of attack when the initial state is $\bar{x}_0$, while $\tilde{x}(\tau_1)$ and $x(\tau_1)$ are the state of (1)-(2) with attack $(B_{\rm c}, B_{\rm i}, u)$ and initial state $\tilde{x}_0$ and x_0 respectively. The claimed statement follows by repeating the arguments for every interval of time $[\tau_{k-1}, \tau_k]$ defined by two subsequent jump times. ■

From Lemma 3.1, undetectable attacks correspond to those inputs that render the output of the system identically zero over time, for some initial condition of the system.

Remark 2: ***(Comparison with existing results for linear-impulsive systems)*** Output-nulling inputs for linear-impulsive systems have been studied and characterized in [17], [34]. Yet, the setup considered in this paper is different, and it leads to a different analysis and results. Compared to [17], we let the input affect both the continuous and impulsive dynamics, and we consider two different output matrices for the continuous and impulsive dynamics. Reference [34], instead, builds a general geometric theory for the control of linear-impulsive systems, which leads to the characterization of output-nulling inputs under the assumption that the jump times are known. Instead, motivated by our applications, when designing stealthy attacks we only assume knowledge of the minimum time between consecutive jumps. □

We next characterize the existence of undetectable attacks.

Theorem 3.2: ***(Existence of undetectable attacks)*** There exist undetectable attacks $(B_{\rm c}, B_{\rm i}, u)$ for the system (1)-(2) if and only if there exists a subspace $\mathcal{V} \subseteq \mathbb{R}^n$ satisfying

$$\mathcal{V} \subseteq \mathrm{Ker}(C_{\rm c}) \cap \mathrm{Ker}(C_{\rm i}), \tag{5}$$

$$A_{\rm c}\mathcal{V} \subseteq \mathcal{V} + \mathrm{Im}(B_{\rm c}), \tag{6}$$

$$A_{\rm i}\mathcal{V} \subseteq \mathcal{V} + \mathrm{Im}(B_{\rm i}) + \mathcal{R}_{\rm c}, \tag{7}$$

where $\mathcal{R}_{\rm c}$ is the largest output-nulling reachable subspace of the continuous-time system with matrices $(A_{\rm c}, B_{\rm c}, C_{\rm c})$.

Proof: Sufficiency of the conditions above is guaranteed by Theorem 3.4, which provides an undetectable attack starting from conditions (5)-(7). We now prove that conditions (5)-(7) are necessary to generate undetectable attacks.

Let $(B_{\rm c}, B_{\rm i}, u)$ be an undetectable attack and consider the evolution of the state $\tilde{x}(t)$ according to (1)-(2). Consider the time interval $[\tau_{\min}, \tau_1)$, where the system evolves according to the linear continuous-time dynamics. Notice that $y(\tilde{x}_0, B_{\rm c}, B_{\rm i}, u, \mathcal{T}, t) = C_{\rm c}\tilde{x}(t)$ for $t \in [\tau_{\min}, \tau_1)$, and $y(\tilde{x}_0, B_{\rm c}, B_{\rm i}, u, \mathcal{T}, t) = C_{\rm i}\tilde{x}(t^-)$ for $t = \tau_1$. To satisfy the undetectability condition (4), because the jump time τ_1 is not known a priori by the attacker, the state trajectory needs to verify $\tilde{x}(t) \in \mathrm{Ker}(C_{\rm c}) \cap \mathrm{Ker}(C_{\rm i})$ for all times $t \in [\tau_{\min}, \tau_1)$.[1] This condition implies the existence of a controlled-invariant subspace $\mathcal{V}$ for $(A_{\rm c}, B_{\rm c})$ satisfying conditions (5)-(6); see [15]. Following the above reasoning, the state trajectory needs to satisfy $\tilde{x}(t) \in \mathcal{V} \subseteq \mathrm{Ker}(C_{\rm c}) \cap \mathrm{Ker}(C_{\rm i})$ for $t \in [\tau_1 + \tau_{\min}, \tau_2)$. Consequently, there needs to be an input for the continuous-time dynamics that brings the state from $\tilde{x}(\tau_1)$ to some state $\tilde{x}(\tau_1 + \tau_{\min}) \in \mathcal{V}$, in a way that $y(\tilde{x}_0, B_{\rm c}, B_{\rm i}, u, \mathcal{T}, t) = 0$ for all times $t \in [\tau_1, \tau_1 + \tau_{\min}]$. For this to be possible, the state $\tilde{x}(\tau_1)$ must be written as $\tilde{x}(\tau_1) = \tilde{x}_{\mathcal{V}} + \tilde{x}_{\mathcal{R}_{\rm c}}$, where $\tilde{x}_{\mathcal{V}} \in \mathcal{V}$ (a controlled invariant for the continuous-time dynamics) and $\tilde{x}_{\mathcal{R}_{\rm c}} \in \mathcal{R}_{\rm c}$, the largest output-nulling reachable subspace of the continuous-time dynamics $(A_{\rm c}, B_{\rm c}, C_{\rm c})$ [15]. This implies condition (7) and concludes the proof. ■

In loose words, the conditions in Theorem 3.2 state that undetectable attacks exist if and only if there is one evolution of the linear-impulsive system that can be maintained in the null space of both the output matrices during the intervals in which the impulsive update can suddenly occur, and in the null space of the continuous output matrix only for the remaining time. In particular, condition (5) ensures that the subspace $\mathcal{V}$, which will contain the state trajectory due to the attack for all times at which the impulsive update is possible, is included in the null space of both the output matrices. Condition (6) ensures that the subspace $\mathcal{V}$ is controlled-invariant for the continuous part of the dynamics, so that the trajectories of the system can be constrained to $\mathcal{V}$, and generate zero output for both the continuous and the impulsive updates. Finally, condition (7) states that the subspace

[1] If the attacker knew the jump times, then the trajectory $\tilde{x}(t)$ would have to satisfy $\tilde{x}(\tau_1^-) \in \mathrm{Ker}(C_{\rm i})$, and $\tilde{x}(t) \in \mathrm{Ker}(C_{\rm c})$ at all other times.

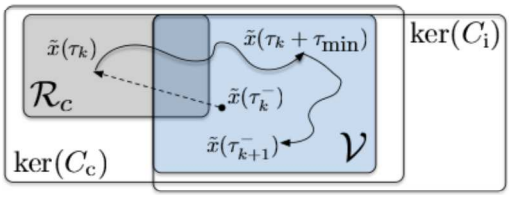

Fig. 1. Representative illustration of the evolution of the auxiliary state $\tilde{x}$ between two successive jumps times τ_k and τ_{k+1}.

$\mathcal{V}$ must be controlled-invariant for the impulsive part of the dynamics, except possibly for a nonzero projection of the state onto $\mathcal{R}_{\rm c}$. In fact, after an impulsive update, the part of the state constrained in $\mathcal{V}$ results in a zero output over time due to $\mathcal{V}$ being controlled-invariant for the discrete dynamics. Instead, the part of the state belonging to $\mathcal{R}_{\rm c}$ can be controlled to zero in time $\tau_{\min}$ with trajectories yielding a zero output, after which an impulsive update is again possible. We will further exploit this characterization in Theorem 3.4 to design an undetectable attack strategy.

The conditions in Theorem 3.2 extend the results for the existence of undetectable attacks in [25] to linear-impulsive systems. In fact, in the absence of impulsive updates, equivalently when $A_{\rm i}$ equals the identity matrix and $C_{\rm i} = 0$, the conditions in Theorem 3.2 reduce to $\mathcal{V} \subseteq \mathrm{Ker}(C_{\rm c})$ and $A_{\rm c}\mathcal{V} \subseteq \mathcal{V} + \mathrm{Im}(B_{\rm c})$, which guarantee that $\mathcal{V}$ is a controlled-invariant subspace contained in the null space of the output matrix, and are equivalent to the conditions described in [25].

Before describing a class of undetectable attacks, we adapt a result from [34] for the computation of the largest subspace that satisfies the conditions in Theorem 3.2.

Lemma 3.3: ***(Largest subspace satisfying Theorem 3.2)*** The largest subspace satisfying conditions (5), (6), and (7) coincides with the last term of the non-increasing sequence

$$\begin{aligned} \mathcal{V}_0 &= \ker(C_{\rm c}) \cap \ker(C_{\rm i}), \\ \mathcal{V}_k &= \mathcal{V}_{k-1} \cap A_{\rm c}^{-1}(\mathcal{V}_{k-1} + \mathrm{Im}(B_{\rm c})) \cap \\ &\quad \cap (A_{\rm i}^{-1}(\mathcal{V}_{k-1} + \mathrm{Im}(B_{\rm i}) + \mathcal{R}_{\rm c})), \quad k \geq 1, \end{aligned} \tag{8}$$

where $\mathcal{R}_{\rm c}$ is the largest output-nulling controllability subspace of $(A_{\rm c}, B_{\rm c}, C_{\rm c})$.

Proof: The proof follows from [34, Theorem 4.2], and the details are omitted here. ■

Lemma 3.3 allows us to assess the existence of undetectable attacks given the system matrices. It should be noticed that the sequence (8) converges in finite time and requires a finite number of linear algebra operations [34].

We conclude this section by presenting a class of undetectable attacks. Our undetectable attack strategy can intuitively be described as follows. Consider the interval $[\tau_k, \tau_{k+1})$, and let $\tilde{x}(\tau_k) \in \mathcal{V} \cup \mathcal{R}_{\rm c} \subseteq \mathrm{Ker}(C_{\rm c})$. The attacker's strategy consists of (i) driving continuously the system state along $\mathcal{V} \cup \mathcal{R}_{\rm c}$ – thus injecting an undetectable input – in a way that $\tilde{x}(\tau_k + \tau_{\min}) \in \mathcal{V}$, and (ii) implementing a feedback action to render $\mathcal{V}$ invariant for the continuous dynamics until the subsequent jump time τ_{k+1} – thus injecting an undetectable input even in the interval $[\tau_k + \tau_{k\min}, \tau_{k+1})$. Due to the properties of the subspace $\mathcal{V}$ (see Theorem 3.2), $\tilde{x}(\tau_{k+1}) \in \mathcal{V} \cup \mathcal{R}_{\rm c} \subseteq \mathrm{Ker}(C_{\rm c})$ and the attacker's strategy can be repeated in subsequent intervals. Our strategy is illustrated in Fig. 1 and formalized in the following theorem.

Theorem 3.4: ***(Undetectable attack strategy)*** Let the subspace $\mathcal{V}$ satisfy conditions (5)-(7) in Theorem 3.2. Let $\tilde{x}(t)$ be an auxiliary state variable evolving according to the dynamics (1)-(2) with initial condition $\tilde{x}_0 \in \mathcal{V} \cup \mathcal{R}_{\rm c}$ and piecewise continuous input u defined as:

$$u(t) = \begin{cases} F_{\rm c}\,\tilde{x}(t) + u_g(t) & \text{for } t \in (\tau_k, \tau_k + \tau_{\min}), \\ F_{\rm c}\,\tilde{x}(t) & \text{for } t \in [\tau_k + \tau_{\min}, \tau_{k+1}), \\ F_{\rm i}\,\tilde{x}(t^-) & \text{for } t = \tau_{k+1}, \end{cases} \tag{9}$$

where

$$u_g(t) = \tilde{B}_{\rm c}^{\mathsf{T}} e^{\tilde{A}_{\rm c}^{\mathsf{T}}(\tau_{\min} + \tau_k - t)}\, W_{\rm c}^{-1} w(\tau_k), \tag{10}$$

$$W_{\rm c} = \int_0^{\tau_{\min}} e^{\tilde{A}_{\rm c}(\tau_{\min} - t)} \tilde{B}_{\rm c} \tilde{B}_{\rm c}^{\mathsf{T}} e^{\tilde{A}_{\rm c}^{\mathsf{T}}(\tau_{\min} - t)} dt, \tag{11}$$

$$\tilde{A}_{\rm c} = A_{\rm c} + B_{\rm c} F_{\rm c}, \tag{12}$$

$$\tilde{B}_{\rm c} = \mathrm{Basis}(\mathrm{Im}(B_{\rm c}) \cap \mathcal{R}_{\rm c}), \tag{13}$$

$$w(\tau_k) = (-e^{\tilde{A}_{\rm c}\tau_{\min}} \tilde{x}(\tau_k) \perp (\mathcal{R}_{\rm c} \setminus \mathcal{V})) + x_c, \tag{14}$$

$$x_c \in \mathcal{R}_{\rm c} \cap \mathcal{V}, \tag{15}$$

and $F_{\rm c}$ and $F_{\rm i}$ are feedback matrices satisfying

$$\begin{aligned} (A_{\rm c} + B_{\rm c}F_{\rm c})(\mathcal{V} + \mathcal{R}_{\rm c}) &\subseteq \mathcal{V} + \mathcal{R}_{\rm c}, \\ (A_{\rm c} + B_{\rm c}F_{\rm c})\mathcal{V} &\subseteq \mathcal{V}, \\ (A_{\rm i} + B_{\rm i}F_{\rm i})\mathcal{V} &\subseteq \mathcal{V} + \mathcal{R}_{\rm c}. \end{aligned}$$

The input signal u defined in (9) is an undetectable attack against the system (1)-(2) for any initial state x_0.

Proof: Consider the evolution of the system from the initial condition $\tilde{x}_0$. Let $\mathcal{V}^* = \mathcal{V} \cup \mathcal{R}_{\rm c}$, and regard the interval $[0, \tau_{\min})$. Recall that $\tilde{x}_0 \in \mathcal{V}^*$, and let $\tilde{x}_0 = \tilde{x}_{\mathcal{V}} + \tilde{x}_{\mathcal{R}_{\rm c}}$ with $\tilde{x}_{\mathcal{V}} \in \mathcal{V}$ and $\tilde{x}_{\mathcal{R}_{\rm c}} \in \mathcal{V}^* \setminus \mathcal{V}$. Notice that $\mathcal{V}^*$ is a controlled-invariant subspace for the continuous dynamics $(A_{\rm c}, B_{\rm c})$ [15], and that $\mathcal{V}^* \subseteq \mathrm{Ker}(C_{\rm c})$. Thus, the feedback input $F_{\rm c}\tilde{x}$ guarantees that the system trajectories remain confined in $\mathcal{V}^*$ and the system output remains at zero. Further, the minimum-energy control input u_g drives the system to a point in $\mathcal{V}$ at time $\tau_{\min}$ along trajectories in $\mathcal{V}^*$, thus invisible from the system output. Consider now the interval $[\tau_{\min}, \tau_1)$, and notice that the feedback input $F_{\rm c}\tilde{x}$ ensures that the system trajectories remain confined in $\mathcal{V}$ until time τ_1, and hence invisible from the system output. Finally, at time τ_1, the discrete feedback $F_{\rm i}\tilde{x}$ guarantees that $\tilde{x}(\tau_1) \in \mathcal{V}^*$. Undetectability of the proposed attack follows by repeating the above arguments for every interval of time $[\tau_k, \tau_{k+1}]$ defined by two subsequent jump times. ■

The following comments on Theorem 3.4 are in order. First, the attack strategy described in Theorem 3.4 is not unique. For example, the minimum-energy control signal u_g could be replaced by any function steering the system to a state $\tilde{x}(\tau_{\min})$ that belongs to $\mathcal{V}$. Similarly, the values of the states $x_{\rm c}$, $\tilde{x}_0$ and of the feedback matrices $F_{\rm c}$ and $F_{\rm i}$ may not be uniquely defined. Second, the control input is computed based on the knowledge of $\tau_{\min}$ and is independent

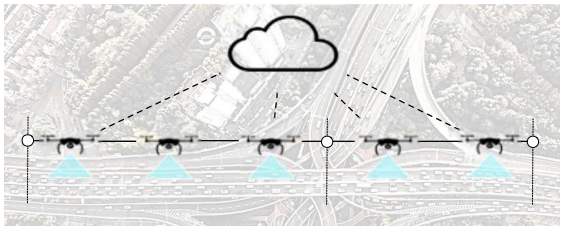

Fig. 2. Cloud-connected multi-agent network. Nearby agents communicate over short-range ad-hoc links and coordinate via a consensus-like surveillance protocol. The agents sporadically connect to the cloud, which allows the to interact over longer distances. Virtual mobile markers (white circles) are used to guide the agents and are impulsively updated through the cloud.

of the actual set $\mathcal{T}$. Thus, the attack strategy in Theorem 3.4 is undetectable for every set of jump times $\mathcal{T}$ where the minimum interval between any two consecutive jumps is lower bounded by $\tau_{\min}$. Third and finally, because the attacker does not know the jump times $\mathcal{T}$, the trajectory $\tilde{x}$, and hence the attack signal, needs to be computed at the same time of the system evolution. While this requirement is satisfied by our attack model, which assumes infinite computational power, it may be difficult to realize in practice. Different attack models are left as the subject of future investigation.

IV. Numerical analysis of attacks in cloud-connected multi-agent networks

To validate our study, consider the problem of surveying an urban environment with a fleet of autonomous agents. To coordinate, we assume that nearby agents can establish ad-hoc communication links, and that a subset of agents can connect to the cloud and interact over a longer distance; see Fig. 2. Cloud-cooperation in autonomous networks is a promising avenue [35], [36], because it not only improves the agents communication range, but it also increases their computational capabilities and contextual awareness. The state of the network consists of two vectors of variables: the first vector, x_{c}, is updated continuously based on the interaction of nearby agents; the second vector, x_{d}, is updated sporadically whenever cloud connections are established. For instance, as shown in Fig. 2, x_{c} may contain the physical positions of the agents, while x_{d} may represent virtual information that is extracted by the cloud and used to guide the agents. When the physical and cloud cooperation algorithms are linear, the nominal network dynamics read as

$$\begin{bmatrix} \dot{x}_{\mathrm{c}}(t) \\ \dot{x}_{\mathrm{d}}(t) \end{bmatrix} = \begin{bmatrix} A_{\mathrm{cc}} & A_{\mathrm{cd}} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_{\mathrm{c}}(t) \\ x_{\mathrm{d}}(t) \end{bmatrix}, \tag{16}$$

for all times $t \in \mathbb{R}_{\geq 0} \setminus \mathcal{T}$, and

$$\begin{bmatrix} x_{\mathrm{c}}(t) \\ x_{\mathrm{d}}(t) \end{bmatrix} = \begin{bmatrix} I & 0 \\ A_{\mathrm{dc}} & A_{\mathrm{dd}} \end{bmatrix} \begin{bmatrix} x_{\mathrm{c}}(t^-) \\ x_{\mathrm{d}}(t^-) \end{bmatrix}, \tag{17}$$

for all times $t \in \mathcal{T} = \{\tau_1, \tau_2, \dots\}$ when cloud communication takes place, with $\tau_{k+1} - \tau_k \geq \tau_{\min}$.

For our study we assume $\tau_{\min} = 5$ and we let the robots follow a consensus-like interaction protocol with matrices

$$A_{\mathrm{cc}} = \begin{bmatrix} 2 & -1 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & 0 & 0 \\ 0 & 0 & 0 & 2 & -1 \\ 0 & 0 & 0 & -1 & 2 \end{bmatrix}, \quad A_{\mathrm{cd}} = -\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

$$A_{\mathrm{dc}} = \frac{1}{6}\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad A_{\mathrm{dd}} = \frac{1}{6}\begin{bmatrix} 6 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 6 \end{bmatrix}.$$

These matrices realize a simple global surveillance strategy, where the two outer virtual beacons remain fixed at the area boundaries, the internal beacons are adjusted based on the positions of the nearest aircrafts, and the robots distribute within the virtual beacons. Notice that the system (16)-(17) is an instance of our general model (1)-(2).

Consider a monitor being able to measure the distances between adjacent aircrafts during the continuous dynamics, and the instantaneous positions of some agents at the jump instants. For our numerical study, the output matrices are

$$C_{\mathrm{c}} = \begin{bmatrix} 0 & 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \end{bmatrix},$$
$$C_{\mathrm{i}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Consider now a man-in-the-middle attack, where the attacker modifies the messages exchanged between the agents and the cloud, and a malware injection on the cloud servers. The attack matrices are

$$B_{\mathrm{c}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}^{\mathsf{T}},$$
$$B_{\mathrm{i}} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}^{\mathsf{T}},$$

which yield the subspaces (see Theorem 3.2)

$$\mathcal{V} = \mathrm{Im}\left(\begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}^{\mathsf{T}}\right),$$
$$\mathcal{R}_{\mathrm{c}} = \mathrm{Im}\left(\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}^{\mathsf{T}}\right).$$

Based on the above matrices, the attacker implements the attack strategy described in Theorem 3.4. To implement this attack, the attacker needs to hijack the information received by the first, third, and fifth aircraft, and also the computation of the cloud to update the second virtual boundary point. While the former action can be achieved by intercepting the data exchanged by the robots, the latter requires the injection of a malicious service into the cloud.

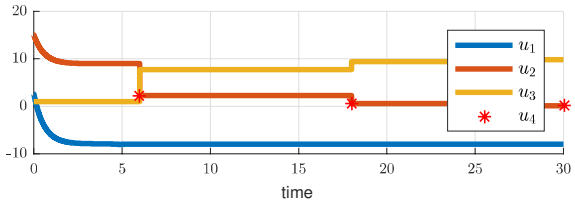

Fig. 3. Attack inputs for the example in Section IV. The signals u_1, u_2, u_3 affect the continuous dynamics, while u_4 alters the impulsive update.

From Theorem 3.4, possible attack matrices are

$$F_{\mathrm{c}} = \begin{bmatrix} 1 & -\frac{1}{4} & -\frac{1}{4} & 0 & 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{bmatrix},$$
$$F_{\mathrm{i}} = 10^{-2} \begin{bmatrix} 0 & 0 & 0 & -8 & -8 & 0 & -8 & 0 \end{bmatrix},$$

the auxiliary initial state is $\tilde{x}_0 = (6, -9, -9, 8, 8, 8, -9, -10)^{\mathsf{T}} \in \mathcal{V} \cup \mathcal{R}_{\mathrm{c}}$ and $x_{\mathrm{c}} = (0, 0, 0, 0, 0, 0, 0, 0)^{\mathsf{T}}$. Fig. 3 shows the attack signal, while Fig. 4(a) and Fig. 4(d) contain the state trajectory and the output associated with the auxiliary state $\tilde{x}$. The attack is applied to the system with initial state $x_0 = (8, -7, -7, 0, 0, 18, -13, -20)^{\mathsf{T}} \notin \mathcal{V} \cup \mathcal{R}_{\mathrm{c}}$, and the resulting trajectories (see Fig. 4(b)) are compared with the ones generated by the system without attack and with initial condition $\bar{x}_0 = x_0 - \tilde{x}_0$ (see Fig. 4(c)). It can be seen that the two state trajectories are different, but the output signals (see Fig. 4(e) and Fig. 4(f) for the attacked and the nominal outputs) are identical, which makes the attack undetectable.

V. Conclusion

In this work we study a security problem for a class of cloud-connected multi-agent systems. We model the system dynamics as a linear-impulsive system, and attacks as exogenous inputs affecting both the continuous and impulsive dynamics. We propose a notion of attack detectability, characterize the existence of undetectable attacks, and design a family of undetectable attacks. Finally, we illustrate our results on a model of cloud-based surveillance network.

References

[1] D. Kushner. The real story of stuxnet. *IEEE Spectrum*, 3(50):48–53, 2013.

[2] S. Karnouskos. Stuxnet worm impact on industrial cyber-physical system security. In *Annual Conference on IEEE Industrial Electronics Society*, pages 4490–4494. IEEE, 2011.

[3] N. Shachtman. Computer virus hits US drone fleet. *cnn.com*, 2011.

[4] K. Hartmann and C. Steup. The vulnerability of UAVs to cyber attacks-an approach to the risk assessment. In *International Conference on Cyber Conflict*, pages 1–23. IEEE, 2013.

[5] A. Javed, H. Larijani, A. Ahmadinia, and D. Gibson. Smart random neural network controller for HVAC using cloud computing technology. *IEEE Trans. on Industrial Informatics*, 13(1):351–360, Feb 2017.

[6] M. Wollschlaeger, T. Sauter, and J. Jasperneite. The future of industrial communication: Automation networks in the era of the internet of things and industry 4.0. *IEEE Industrial Electronics Magazine*, 11(1):17–27, March 2017.

[7] J. Weinman. The economics and strategy of manufacturing and the cloud. *IEEE Cloud Computing*, 3(4):6–11, July 2016.

[8] D. Georgakopoulos, P. P. Jayaraman, M. Fazia, M. Villari, and R. Ranjan. Internet of things and edge cloud computing roadmap for manufacturing. *IEEE Cloud Computing*, 3(4):66–73, July 2016.

[9] G. Mohanarajah, D. Hunziker, R. D'Andrea, and M. Waibel. Rapyuta: A cloud robotics platform. *IEEE Transactions on Automation Science and Engineering*, 12(2):481–493, April 2015.

[10] J. Salmerón-Garca, P. Íñigo-Blasco, F. Daz del Ro, and D. Cagigas-Muñiz. A tradeoff analysis of a cloud-based robot navigation assistant using stereo image processing. *IEEE Transactions on Automation Science and Engineering*, 12(2):444–454, April 2015.

[11] M. Whaiduzzaman, M. Sookhak, A. Gani, and R. Buyya. A survey on vehicular cloud computing. *Journal of Network and Computer Applications*, 40(1):325–344, 2014.

[12] M. Eltoweissy, S. Olariu, and M. Younis. *Towards Autonomous Vehicular Clouds*, pages 1–16. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.

[13] M. Sookhak, F. R. Yu, Y. He, H. Talebian, N. Sohrabi Safa, N. Zhao, M. K. Khan, and N. Kumar. Fog vehicular computing: Augmentation of fog computing using vehicular cloud computing. *IEEE Vehicular Technology Magazine*, 12(3):55–64, Sept 2017.

[14] C. Lochert, B. Scheuermann, M. Caliskan, and M. Mauve. The feasibility of information dissemination in vehicular ad-hoc networks. In *2007 Fourth Annual Conference on Wireless on Demand Network Systems and Services*, pages 92–99, Jan 2007.

[15] G. Basile and G. Marro. *Controlled and Conditioned Invariants in Linear System Theory*. Prentice Hall, 1991.

[16] E. A. Medina and D. A. Lawrence. Controlled and conditioned invariants for linear impulsive systems. In *Proceedings of the 45th IEEE Conf. on Decision and Control*, pages 2753–2758, Dec 2006.

[17] A. M. Perdon, E. Zattoni, and G. Conte. Disturbance decoupling in hybrid linear systems with state jumps. *IEEE Transactions on Automatic Control*, 62(12):6552–6559, 2017.

[18] Y. Liu, M. K. Reiter, and P. Ning. False data injection attacks against state estimation in electric power grids. In *ACM Conference on Computer and Communications Security*, pages 21–32, Chicago, IL, USA, November 2009.

[19] A. Teixeira, S. Amin, H. Sandberg, K. H. Johansson, and S.S. Sastry. Cyber security analysis of state estimators in electric power systems. In *IEEE Conf. on Decision and Control*, pages 5991–5998, Atlanta, GA, USA, December 2010.

[20] T. Bhattacharya, S.and Başar. Differential game-theoretic approach to a spatial jamming problem. In *Advances in Dynamic Games*, pages 245–268. Springer, 2013.

[21] S. Maharjan, Q. Zhu, Y. Zhang, S. Gjessing, and T. Başar. Dependable demand response management in the smart grid: A stackelberg game approach. *IEEE Transactions on Smart Grid*, 4(1):120–132, 2013.

[22] H. Fawzi, P. Tabuada, and S. Diggavi. Secure estimation and control for cyber-physical systems under adversarial attacks. *IEEE Transactions on Automatic Control*, 59(6):1454–1467, 2014.

[23] C.-Z. Bai, F. Pasqualetti, and V. Gupta. Security in stochastic control systems: Fundamental limitations and performance bounds. In *American Control Conference*, pages 195–200, Chicago, Il, July 2015.

[24] F. Pasqualetti, A. Bicchi, and F. Bullo. Consensus computation in unreliable networks: A system theoretic approach. *IEEE Transactions on Automatic Control*, 56(12), 2011.

[25] F. Pasqualetti, F. Dörfler, and F. Bullo. Attack detection and identification in cyber-physical systems. *IEEE Transactions on Automatic Control*, 58(11):2715–2729, 2013.

[26] S. Sundaram and C. Hadjicostis. Distributed function calculation via linear iterative strategies in the presence of malicious agents. *IEEE Transactions on Automatic Control*, 56(7):1495–1508, 2011.

[27] R. Smith. A decoupled feedback structure for covertly appropriating network control systems. In *IFAC World Congress*, pages 90–95, Milan, Italy, August 2011.

[28] A. A. Cárdenas, S. Amin, and S. S. Sastry. Research challenges for the security of control systems. In *Proceedings of the 3rd Conference on Hot Topics in Security*, pages 6:1–6:6, Berkeley, CA, USA, 2008.

[29] R. Goebel, R. G. Sanfelice, and A. R. Teel. *Hybrid Dynamical Systems: Modeling, Stability, and Robustness*. Princeton University Press, 2012.

[30] S. Phillips, A. Duz, F. Pasqualetti, , and R. G. Sanfelice. Hybrid attack monitor design to detect recurrent attacks in a class of cyber-physical systems. In *IEEE Conf. on Decision and Control*, pages 1368–1373, Melbourne, Australia, December 2017.

[31] F. Pasqualetti, F. Dörfler, and F. Bullo. Control-theoretic methods for cyberphysical security: Geometric principles for optimal cross-layer resilient control systems. *IEEE Control Systems Magazine*, 35(1):110–127, 2015.

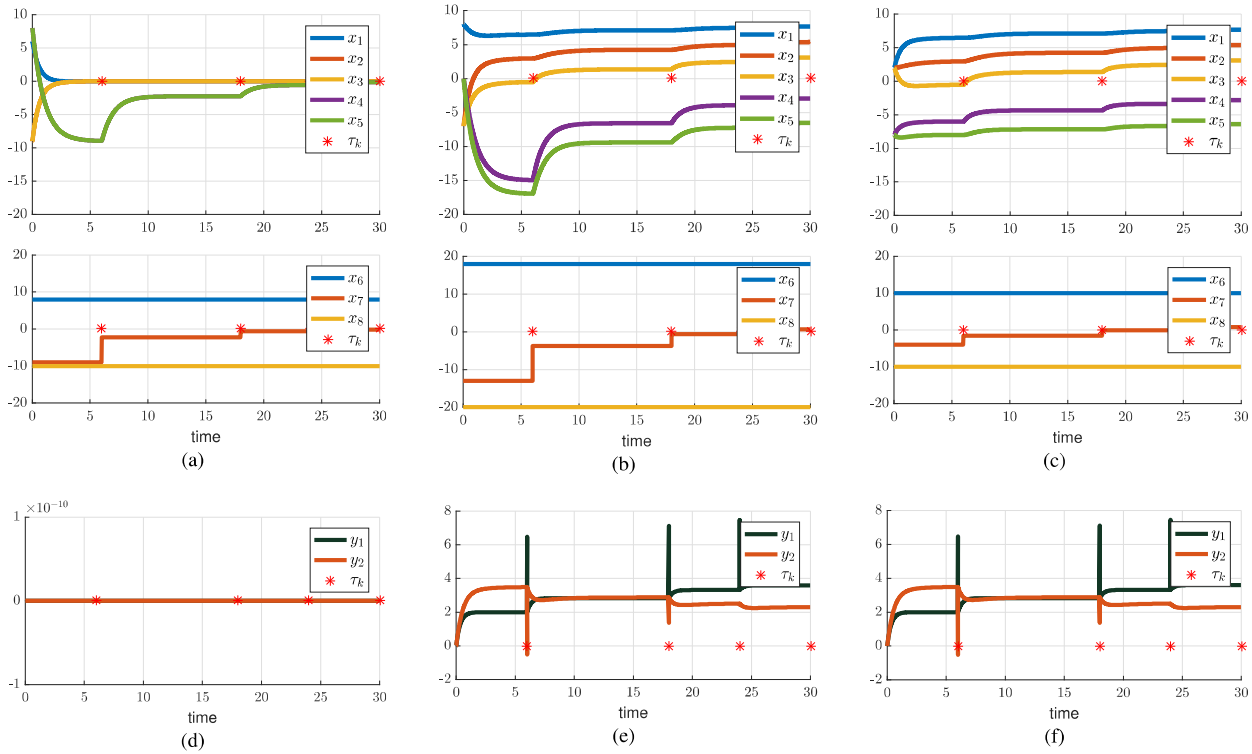

Fig. 4. State trajectory (4(a)) and output (4(d)) of the auxiliary system in Section IV under the attack in Fig. 3. Notice that the output is identically zero. State (4(b)) and output (4(e)) of the attacked system; state (4(c)) and output (4(f)) of the corresponding non-attacked system. See Section IV. Since the attacked and non-attacked systems exhibit the same outputs, no monitor can distinguish between the two scenarios without knowledge of the initial state.

[32] W. M. Wonham. *Linear Multivariable Control: A Geometric Approach*. Springer, 3 edition, 1985.

[33] H. L. Trentelman, A. Stoorvogel, and M. Hautus. *Control Theory for Linear Systems*. Springer, 2001.

[34] D. A. Lawrence. Controlled invariant subspaces for linear impulsive systems. In *2014 American Control Conference*, pages 2336–2341, June 2014.

[35] M. Gharibi, R. Boutaba, and S. L. Waslander. Internet of drones. *IEEE Access*, 4:1148–1162, 2016.

[36] S. Srinivasan, H. Latchman, J. Shea, T. Wong, and J. McNair. Airborne traffic surveillance systems: video surveillance of highway traffic. In *Workshop on Video Surveillance & Sensor Networks*, pages 131–135, 2004.