Automatica 134 (2021) 109953

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Distributed control under compromised measurements: Resilient estimation, attack detection, and vehicle platooning[☆]

Xingkang He[a,*], Ehsan Hashemi[b], Karl H. Johansson[a]

[a] Division of Decision and Control Systems, School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, and Digital Futures, Sweden
[b] Mechanical Engineering Department, University of Alberta, Canada

ARTICLE INFO

Article history:
Received 19 October 2020
Received in revised form 24 March 2021
Accepted 24 July 2021
Available online 11 October 2021

Keywords:
Resilient estimation
Attack detection
Distributed control
Compromised measurements
Sensor attacks

ABSTRACT

We study how to design a secure observer-based distributed controller such that a group of vehicles can achieve accurate state estimates and formation control even if the measurements of a subset of vehicle sensors are compromised by a malicious attacker. We propose an architecture consisting of a resilient observer, an attack detector, and an observer-based distributed controller. The distributed detector is able to update three sets of vehicle sensors: the ones surely under attack, surely attack-free, and suspected to be under attack. The adaptive observer saturates the measurement innovation through a preset static or time-varying threshold, such that the potentially compromised measurements have limited influence on the estimation. Essential properties of the proposed architecture include: (1) The detector is fault-free, and the attacked and attack-free vehicle sensors can be identified in finite time; (2) The observer guarantees both real-time error bounds and asymptotic error bounds, with tighter bounds when more attacked or attack-free vehicle sensors are identified by the detector; (3) The distributed controller ensures closed-loop stability. The effectiveness of the proposed methods is evaluated through simulations by an application to vehicle platooning.

1. Introduction

Motivations and related work

Networked control systems (NCS) are ubiquitous. The performance of NCS significantly depends on widely deployed sensors which might be compromised due to the presence of malicious attackers (Baras & Liu, 2019; Shoukry et al., 2018). The attackers can strategically manipulate the sensor measurements in order to affect stability and performance of NCS. Attack detection, state estimation, and system control are three major components in the design of secure NCS in malicious environments.

To detect whether systems are under attack and identify attacked components, quite a few detection methods are proposed. Attack detection and identification for linear descriptor systems are studied in Pasqualetti, Dörfler, and Bullo (2013). Methods of attack detection and correction for noise-free linear systems are proposed in Tang, Kuijper, Chong, Mareels, and Leckie (2019). To detect the Byzantine adversaries with quantized false alarm rates, a trust-aware consensus algorithm is proposed in Baras and Liu (2019). In Gallo, Turan, Boem, Parisini, and Ferrari-Trecate (2020), Ge, Han, Zhong, and Zhang (2019), distributed detectors are designed for false data injection (FDI) attacks in communications. Detection and mitigation methods are proposed in Deghat, Ugrinovskii, Shames, and Langbort (2019) for distributed observers under a class of bias injection attacks. A joint detection and estimation problem is investigated in Forti et al. (2018) with the knowledge of some attack statistics. There are some methods for multi-observer based detector design (Chowdhury, Belikov, Baimel, & Levron, 2020; Kim, Lee, Shim, Eun, & Seo, 2018; Yang, Murguia, Kuijper, & Nešić, 2020). However, the computational complexity of these methods substantially increases as the number of sensors is increasing. Thus, designing single-observer based detectors without relying on the knowledge of attack signals needs more investigations. Moreover, most existing methods focus on detecting the attacked sensors, but few results are given for the identification of attack-free sensors.

There are two major approaches in the literature for handling state estimation under sensor attacks. The first approach is based on solving optimization problems (Fawzi, Tabuada, & Diggavi,

[☆] This work was supported by the Knut & Alice Wallenberg Foundation, the Swedish Research Council, the Swedish Foundation for Strategic Research, and also by the Natural Science and Engineering Research Council of Canada (NSERC), Discovery Grants. The material in this paper was not presented at any conference. This paper was recommended for publication in revised form by Associate Editor Mohammed Chadli under the direction of Editor Torsten Söderström.

[*] Corresponding author.
E-mail addresses: xingkang@kth.se (X. He), ehashemi@ualberta.ca (E. Hashemi), kallej@kth.se (K.H. Johansson).

https://doi.org/10.1016/j.automatica.2021.109953

2014; Gao, Sun, Liu, Shi, & Wu, 2020; Lu & Yang, 2019; Pajic, Lee, & Pappas, 2017; Shinohara, Namerikawa, & Qu, 2019; Shoukry et al., 2018, 2017). This approach needs a large number of computational resources in enumerating all sensor combinations in order to find the attacked sensor set. Thus, it is not suitable to large-scale sensor networks if the resources are constrained. The second approach is to use robust techniques in handling potentially compromised data, such as discarding a few largest and smallest elements (Mitra, Richards, Bagchi, & Sundaram, 2019; Mitra & Sundaram, 2019; Ren, Mo, Chen, & Johansson, 2020; Su & Shahrampour, 2020), using the signum information of measurement innovations (Lee, Kim, & Shim, 2020), and saturating the innovation which reaches a threshold (Chen, Kar, & Moura, 2019; He, Ren, Sandberg, & Johansson, 2021). This approach is more suitable in online estimation since it needs very less computational resources than the first approach. However, there are few results in this direction, especially for dynamical systems under FDI sensor attacks.

Some resilient distributed control strategies have been proposed to achieve formation control of a group of vehicles or robots in malicious environments. There are strategies on how to handle different attacks, such as replay attack on control commands (Zhu & Martínez, 2013), denial-of-service (DoS) attack on measurement and control channels (Zhu & Zheng, 2020), FDI attack in the transmission from controller to actuator (Zhao, Wang, Wei, & Han, 2020), attack on network topology of multi-agent systems (Feng, Wen, & Hu, 2017), and stealthy integrity attacks (Weerakkody, Liu, Son, & Sinopoli, 2016). However, there is no unified architecture integrating resilient estimation, attack detection and distributed control.

Contributions

In this paper, we propose an architecture comprising of a resilient observer, an online attack detector, and a distributed controller, such that a group of vehicles can achieve accurate state estimates and formation control even if the measurements of a subset of the vehicle sensors are compromised by a malicious attacker. The main contributions of this paper are summarized as follows:

(i) We propose an adaptive resilient observer, designed by saturating the measurement innovation through a preset static or time-varying threshold, such that the potentially compromised measurements have limited influence to the estimation (Algorithm 1). Some essential properties are found: (i) The observer is able to provide an upper bound of the estimation error at each time (Proposition 1); (ii) If the observer threshold is static and satisfied with some explicit design principle (Proposition 2), the estimation error is asymptotically upper bounded (Theorem 1); and (iii) If the observer threshold is time-varying and computed adaptively, the estimation error is also asymptotically upper bounded (Theorem 2).

(ii) We develop an online distributed attack detector with the potentially compromised sensor measurements and the observer's estimates. The designed detector is able to update three sets of vehicle sensors: the ones surely under attack, surely attack-free, and suspected to be under attack (Algorithm 2). Some properties are found: (i) The detector is fault-free (Lemma 1), which differs from the existing results with false alarms (e.g., Baras and Liu (2019)); and (ii) If some condition holds, all attacked and attack-free vehicle sensors are identified in finite time (Theorem 3);

(iii) We design a distributed controller (Algorithm 3) to achieve the formation control of the vehicles. We find that if the controller parameters satisfy some graph-related conditions, the overall performance function is asymptotically upper bounded in the presence of noise and tending to zero in the absence of noise (Theorem 4 and Corollary 1), which ensures the closed-loop stability of the proposed architecture.

The results of this paper are substantially different from the literature. Compared to the results of secure distributed estimation for estimating an overall system state (Deghat et al., 2019; Forti et al., 2018; Mitra & Sundaram, 2019), it should be noted that we estimate the local vehicle state under compromised sensor measurements. The developed sensor attack detector is based on one observer, which requires less computational resources than detectors proposed in the literature based on multiple observers, such as (Chowdhury et al., 2020) for attacked linear systems and Kim et al. (2018) and Yang et al. (2020) for nonlinear systems.

Outline

The remainder of the paper is organized as follows: Section 2 is on the problem formulation, followed by an overview of the proposed distributed observer-based control architecture in Section 3. Section 4 designs a resilient observer for each vehicle, based on which Section 5 studies the attack detection problem. In Section 6, a distributed controller is proposed to close the loop. After simulations of vehicle platooning in Section 7, the paper is concluded in Section 8. The main proofs are given in Appendix.

Notations: $\mathbb{R}^{n\times m}$ denotes the set of real-valued matrices with n rows and m columns, and $\mathbb{R}^n$ the set of n-dimensional real-valued vectors. Without specific explanation, the scalars and matrices in this paper are real-valued. Denote $\mathbb{Z}$ the set of integers, $\mathbb{N}^+$ the set of positive integers, and $\mathbb{N} = \mathbb{N}^+ \cup 0$. The matrix I_n stands for the n-dimensional square identity matrix. The superscript "T" represents the transpose. The operator diag$\{\cdot\}$ represents the diagonalization. We denote the Kronecker product of A and B by $A \otimes B$. The vector norm $\|x\|$ is the 2-norm of a vector x. The matrix norm $\|A\|$ is the induced 2-norm, i.e., $\|A\| = \sup_{x\neq 0} \|Ax\| / \|x\|$. The notations $\sigma_{\min}(A)$ and $\sigma_{\max}(A)$ are the minimum and maximum eigenvalues of a real-valued symmetric matrix A, respectively. The notation $a = (a_i)_{i=1,2,\ldots,n}$ is a column vector consisting of elements $a_1, \ldots, a_n$. Let $\mathbb{I}_{i\in\mathcal{C}}$ be an indicator function, which equals 1 if $i \in \mathcal{C}$; otherwise, it is 0. Let $\lceil x \rceil = \min\{n \in \mathbb{Z} | x \leq n\}$.

2. Problem formulation

In this section, we first motivate the problem through a vehicle platooning example, and then formulate the problem.

2.1. Motivating example

Consider the five-vehicle platooning in Fig. 1. The aim is to control the speed of all vehicles to a desired value while maintaining a safe distance between any two adjacent vehicles. Each vehicle is able to obtain its position and velocity measurements through a GPS receiver or a similar sensor, and the relative position and velocity measurements to its front vehicle through a sensor like a camera or radar. All vehicles collaborate in the platoon by using their local measurements, and vehicle-to-vehicle communication.

Suppose there is a malicious attacker, which aims to affect the platoon by compromising the position and velocity measurements of vehicle 1. Such attack could be a spoofing attack on a

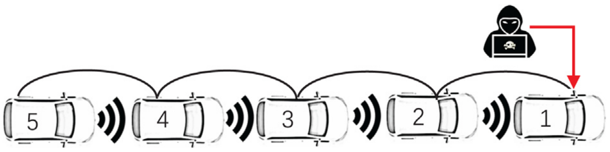


Fig. 1. Platoon of five vehicles, where the position and velocity measurements of vehicle 1 are compromised by a malicious attacker. Each vehicle is able to exchange messages with other vehicles nearby through wireless communication.

GPS receiver. By using the compromised measurements, vehicle 1 is unable to control its velocity to the desired value. Consequently, the platoon is not able to maintain a proper formation. The data redundancy resulting from the absolute and relative measurements of the follower vehicles, however, provides an opportunity for designing resilient estimation and control algorithms. The algorithms are expected to mitigate such sensor attacks in order to achieve vehicle platooning.

2.2. System model

Consider $N \geq 3$ vehicles, which are labeled from the leader to the tail by $1, 2, \ldots, N$. We study the second-order vehicle model: for $i = 1, 2, \ldots, N$,

$$\begin{aligned} x_i(t+1) &= Ax_i(t) + Bu_i(t) + d_i(t) \\ &= \begin{pmatrix} 1 & T \\ 0 & 1 \end{pmatrix} x_i(t) + \begin{pmatrix} 0 \\ T \end{pmatrix} u_i(t) + d_i(t), \end{aligned} \tag{1}$$

where $x_i(t) = (s_i(t), v_i(t))^{\mathsf{T}} \in \mathbb{R}^2$ is the state of vehicle i consisting of position $s_i(t)$ and velocity $v_i(t)$, $u_i(t) \in \mathbb{R}$ the control input, $d_i(t) \in \mathbb{R}^2$ the process noise, all at time $t \in \mathbb{N}$, and $T > 0$ the sampling time. Vehicle i is able to obtain its absolute measurements of position and velocity through sensor i, which is a potentially attacked sensor (e.g., a GPS receiver under spoofing attack):

$$y_{i,i}(t) = x_i(t) + a_i(t) + n_{i,i}(t) \tag{2}$$

where $y_{i,i}(t) \in \mathbb{R}^2$ and $n_{i,i}(t) \in \mathbb{R}^2$ are the measurement and measurement noise, and the vector $a_i(t) \in \mathbb{R}^2$ represents an attack signal injected by a malicious attacker. Moreover, we assume each vehicle $j \in \{2, 3, \ldots, N\}$ has a secured sensor (e.g., an onboard radar or camera) to measure the relative state between itself and its front vehicle (i.e., vehicle $j-1$):

$$y_{j-1,j}(t) = x_j(t) - x_{j-1}(t) + n_{j-1,j}(t), \tag{3}$$

where $y_{j-1,j}(t) \in \mathbb{R}^2$ and $n_{j-1,j}(t) \in \mathbb{R}^2$ are the measurement and measurement noise.

Although the relative state measurements $\{y_{j-1,j}(t)\}$ are secured, it is not possible to accurately estimate the absolute state $x_j(t)$ simply with these measurements. In the rest of the paper, we say that sensor i is under attack if the unsecured sensor of vehicle i is under attack.

2.3. Attack model

The attack model is provided in the following assumption.

Assumption 1. There is an unknown and time-invariant attack set $\mathcal{S}^a \subset \{1, 2, \ldots, N\}$ with at most $b \geq 1$ elements, such that the corresponding attack signals $a_i(t) \in \mathbb{R}^2$, $i \in \mathcal{S}^a$, $t \in \mathbb{N}$, are arbitrary, and the maximum number of attacked sensors b is known to each vehicle. For the set of attack-free vehicle sensors $\mathcal{S} := \{1, 2, \ldots, N\} \setminus \mathcal{S}^a$, it holds that $a_i(t) \equiv 0$, $i \in \mathcal{S}$, $t \in \mathbb{N}$.

Even if Assumption 1 restricts the attack set to be time-invariant, the attacker can compromise the sensors in the set in an arbitrarily and possibly time-varying way. A subset $\mathcal{S}^a$ of the vehicle sensor measurements (2) can thus be manipulated, but we do not know which ones. Assumption 1 does not impose any specific distribution or form of $a_i(t)$, and covers several relevant sensor attacks, including random attack, DoS attack, bias injection attack, and replay attack (Teixeira, Shames, Sandberg, & Johansson, 2015). The assumption is common in the literature (Fawzi et al., 2014; Lu & Yang, 2019; Pajic et al., 2017; Shinohara et al., 2019; Shoukry et al., 2018, 2017).

The upper bound b of the number of attacked vehicle sensors is used in the observer and detector designs. The assumption on the knowledge of b can be relaxed, but will result in worse performance for the same number of attacked sensors.

2.4. Problem

In order to achieve vehicle formation control (e.g., vehicle platooning) in a malicious environment, it is important to estimate the states of all vehicles simultaneously. For example, when a group of vehicles are required to achieve a platoon with a desired speed, it is necessary to estimate the state of the leader vehicle for controller design. However, its absolute measurements are potentially compromised as in (2). In order to have data redundancy for the state estimation of the leader vehicle, the secured relative measurements and accurate estimates of the follower vehicles are necessary.

To measure the overall estimation and control performance for system (1)–(3), we introduce the performance function $\varphi(t)$:

$$\varphi(t) = \frac{1}{N} \sum_{i=1}^{N} \left\| \hat{x}_i(t) - x_i(t) \right\| + \left\| x_i(t) - x_i^*(t) \right\|, \tag{4}$$

where $\hat{x}_i(t)$ is the estimate of $x_i(t)$ from the observer to be designed, and $x_i^*(t)$ is the desired vehicle state of the formation satisfying

$$x_i^*(t) = \begin{cases} x_0(t), & \text{if } i = 1 \\ x_{i-1}^*(t) - \Delta x_{i-1,i}(t), & \text{if } i \in \{2, 3, \ldots, N\}, \end{cases}$$

where $x_0(t)$ is the reference state of the leader vehicle, subject to $x_0(t+1) = Ax_0(t)$, and $\Delta x_{i-1,i}(t)$ is the desired relative state between vehicles $i-1$ and i, subject to $\Delta x_{i-1,i}(t+1) = A\Delta x_{i-1,i}(t)$, $i = 2, 3, \ldots, N$. For convenience, we let $\Delta x_{0,1}(t) \equiv [0, 0]^{\mathsf{T}}$.

If $\Delta x_{i-1,i}(t) \equiv [0, 0]^{\mathsf{T}}$, $i = 1, \ldots, N$, it means all vehicles aim to reach the reference state x_0; if $\Delta x_{i-1,i}(t) \equiv [s_0, 0]^{\mathsf{T}}$, where s_0 is a positive scalar, it means all vehicles are expected to have the same speed, and two nearest neighbor vehicles keep the distance s_0, which is a typical scenario in vehicle platooning.

Assumption 2. The noise in (1)–(3), and the initial estimation error satisfy: $\forall i \in \{1, \ldots, N\}$ and $\forall j \in \{2, \ldots, N\}$,

$$\|\hat{x}_i(0) - x_i(0)\| \leq q, \quad \sup_{t \geq 0} \|d_i(t)\| \leq \epsilon,$$

$$\sup_{t \geq 0} \max\{\|n_{i,i}(t)\|, \|n_{j-1,j}(t)\|\} \leq \mu,$$

where the scalars $q > 0$, $\epsilon \geq 0$, and $\mu \geq 0$ are known to each vehicle.

The bounded noise can be used to model sensor bias, output disturbances, unknown bounded inputs, unmodeled dynamics, and model errors from system linearization and discretization. The upper bounds q, ϵ, μ are used in the observer and detector designs. The assumption on the knowledge of q, ϵ, and μ can be relaxed, but will result in worse performance for the same noise and initial estimation error.

The following control design problem is solved in this paper and explicit estimates of the performance bound c_0 are derived.

Problem: Design an observer-based distributed controller for system (1)–(3) under Assumptions 1–2, and find conditions such that:

(i) In the presence of noise, there is a scalar $c_0 > 0$, such that $\limsup_{t\to\infty} \varphi(t) < c_0$;

(ii) In the absence of noise, $\lim_{t\to\infty} \varphi(t) = 0$.

3. Observer-based distributed control architecture

In this section, we first introduce the communication structure of the vehicle network, and then propose an architecture consisting of a resilient observer, an attack detector, and a distributed controller. Moreover, the measurements of each vehicle are reconstructed based on vehicle-to-vehicle communication.

3.1. Communication structure of vehicle network

We model the vehicle communication topology by an undirected graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, which consists of the set of nodes $\mathcal{V} = \{1, 2, \dots, N\}$ and the set of edges $\mathcal{E}$. If there is an edge $(i, j) \in \mathcal{E}$, node i can exchange information with node j. In this case, node j is called a neighbor of node i, and vice versa. Denote the neighbor set of node $i \in \mathcal{V}$ by $\mathcal{N}_i := \{j \in \mathcal{V} | (i, j) \in \mathcal{E}\}$, which in this paper is assumed to be

$$\mathcal{N}_i = \begin{cases} \{i-L, \dots, i-1, i+1, \dots, i+L\}, & \text{if } i \in \mathcal{V}_1 \\ \{1, \dots, i-1, i+1, \dots, i+L\}, & \text{if } i \in \mathcal{V}_{2,1} \\ \{i-L, \dots, i-1, i+1, \dots, N\}, & \text{if } i \in \mathcal{V}_2 \setminus \mathcal{V}_{2,1}, \end{cases}$$

where $L \in \mathbb{N}^+$ is a parameter indicating the neighbor range, $\mathcal{V}_{2,1} = \{1, 2, \dots, L\}$, and

$$\mathcal{V}_1 = \{L+1, L+2, \dots, N-L\}, \quad \mathcal{V}_2 = \mathcal{V} \setminus \mathcal{V}_1. \tag{5}$$

As seen, each vehicle $i \in \mathcal{V}_1$ has $2L$ neighbors, and each vehicle $j \in \mathcal{V}_2$ has less than $2L$ neighbors. In Section 3.3, we show that each vehicle i uses the messages received from its neighbors $j \in \mathcal{N}_i$ to reconstruct measurements for observer design. The communication topologies of five vehicle control systems (VCSs) for $L = 1$ and $L = 2$ are illustrated in Figs. 1 and 2, respectively. In the following, we use the term 'vehicle' to represent a VCS for convenience. Each vehicle $i \in \mathcal{V}$ is able to send its neighbor vehicle $j \in \mathcal{N}_i$ a message at time $t \in \mathbb{N}^+$, denoted by $\mathcal{M}_i(t)$ (omitting the time index t in the following notation):

$$\mathcal{M}_i = \begin{cases} \{y_{1,1}, \bar{x}_1, \hat{\mathcal{S}}_1, \hat{\mathcal{S}}_1^a, \hat{\mathcal{S}}_1^s, \alpha_1\} & \text{if } i = 1 \\ \{y_{i-1,i}, y_{i,i}, \bar{x}_i, \hat{\mathcal{S}}_i, \hat{\mathcal{S}}_i^a, \hat{\mathcal{S}}_i^s, \alpha_i\} & \text{otherwise}, \end{cases} \tag{6}$$

where $\bar{x}_i(t+1) = A\hat{x}_i(t) + [0, Tu_i(t)]^\top$ is the predicted value of $x_i(t+1)$ from the observer to be designed, $\alpha_i(t)$ denotes the estimation error bound to be specified in Proposition 1, and

- $\hat{\mathcal{S}}_i(t)$: the set of attack-free vehicle sensors estimated by vehicle i at time t, i.e., the estimate of $\mathcal{S}$
- $\hat{\mathcal{S}}_i^a(t)$: the set of attacked vehicle sensors estimated by vehicle i at time t, i.e., the estimate of $\mathcal{S}^a$
- $\hat{\mathcal{S}}_i^s(t)$: the set of vehicle sensors, which are suspected to be under attack, estimated by vehicle i.

When we find an anomaly via a detector using measurements of two vehicle sensors, it is safe to conclude that at least one sensor is under attack and two sensors are suspicious. Then the two sensors will be included in $\hat{\mathcal{S}}_i^s(t)$. The role of $\hat{\mathcal{S}}_i^s(t)$ is to update $\hat{\mathcal{S}}_i(t)$, as shown in line 22 of Algorithm 2 in Section 5. Note that $\hat{\mathcal{S}}_i^s(t) \subseteq \mathcal{V}$ is not necessarily a subset of $\mathcal{S}^a$, since $\hat{\mathcal{S}}_i^s(t)$ may include some attack-free vehicle sensors. The three sets $\{\hat{\mathcal{S}}_i(t), \hat{\mathcal{S}}_i^a(t), \hat{\mathcal{S}}_i^s(t)\}$ are shared between vehicles through the vehicle-to-vehicle network $\mathcal{G}$ and updated in a distributed manner described in Section 5. We assume there is no prior information on sensor identities, thus the sets are initialized as empty sets, i.e., $\hat{\mathcal{S}}_i(0) = \hat{\mathcal{S}}_i^a(0) = \hat{\mathcal{S}}_i^s(0) = \emptyset$, $i \in \mathcal{V}$. Otherwise, the sensors with known identities would be included in these initial sets respectively.

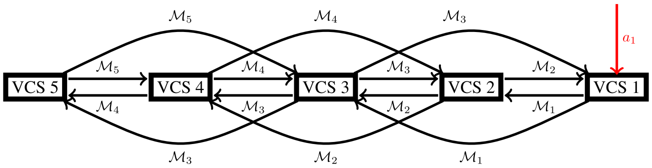


Fig. 2. Communication topology of the undirected graph $\mathcal{G}$ with five vehicle control systems (VCSs) for $L = 2$, where VCS 1 is under attack and $\mathcal{M}_j$, defined in (6), is the message sent out by VCS j to its neighbors, $j = 1, 2, \dots, 5$, and a_1 is the attack signal.

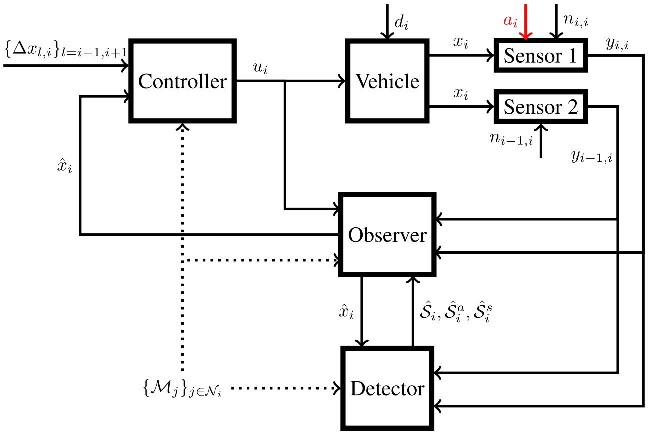

$\{\Delta x_{l,i}\}_{l=i-1,i+1}$
Controller
u_i
d_i
Vehicle
x_i
a_i
$n_{i,i}$
Sensor 1
$y_{i,i}$
x_i
Sensor 2
$n_{i-1,i}$
$y_{i-1,i}$
$\hat{x}_i$
Observer
$\hat{x}_i$
$\hat{\mathcal{S}}_i, \hat{\mathcal{S}}_i^a, \hat{\mathcal{S}}_i^s$
$\{\mathcal{M}_j\}_{j\in\mathcal{N}_i}$
Detector

Fig. 3. Vehicle control system architecture for vehicle i: The control signal for vehicle i utilizes information from other vehicles as indicated by the dashed arrows: $\mathcal{M}_j$ is defined in (6), $j \in \mathcal{N}_i$. The observer, detector, and controller are designed in Sections 4, 5, and 6, respectively.

3.2. Resilient observer-based distributed control architecture

We design an architecture for the VCS of each vehicle i in Fig. 3. The architecture integrates the resilient observer in Section 4, the attack detector in Section 5, and the distributed controller in Section 6. The observer leverages the measurements of vehicle i and neighbor vehicles. Then, the estimate $\hat{x}_i(t)$ from the observer is sent to the controller, which employs $\hat{x}_i(t)$ as well as the estimates of neighbor vehicles to generate control signal $u_i(t)$. If the observer is inefficient, the observer-based controller would not work well. Therefore, the key point for the observer is how to use the potentially attacked measurements and the measurements from neighbor vehicles efficiently. In Section 4, a resilient observer is proposed by leveraging a new saturation approach. The designed detector is able to update the three sets $\{\hat{\mathcal{S}}_i, \hat{\mathcal{S}}_i^a, \hat{\mathcal{S}}_i^s\}$, and send them to the observer. Then, in order to improve the estimation performance, the observer will discard the measurements of the untrustworthy vehicles henceforth, and fully utilize the measurements of the trustworthy vehicles. Note that the detector in Section 5 ensures consistency of the three sets in the sense that they will not conflict. In other scenarios, if an inconsistent case occurs due to some reasons (e.g., the detection data is manipulated), the architecture in Fig. 3 can be employed by abandoning the inconsistent subsets.

3.3. Measurement reconstruction via vehicle communication

Based on whether each vehicle has $2L$ neighbors, we split the vehicle set $\mathcal{V}$ into two subsets $\mathcal{V}_1$ and $\mathcal{V}_2$ as shown in (5). In the following, we first reconstruct the measurement equation of vehicle $i \in \mathcal{V}_1$ by employing the local measurements (2)–(3) and the messages from neighbor vehicles. It follows from (2) and (3) that

$$y_{i|j}(t) = x_i(t) + a_j(t) + n_{i|j}(t), \tag{7}$$

where

$$y_{i|j}(t) = \begin{cases} y_{j,j}(t) + \sum_{m=j+1}^{i} y_{m-1,m}(t), & \text{if } i > j \\ y_{i,i}(t), & \text{if } i = j \\ y_{j,j}(t) - \sum_{m=i+1}^{j} y_{m-1,m}(t), & \text{if } i < j \end{cases}$$

$$n_{i|j}(t) = \begin{cases} n_{j,j}(t) + \sum_{m=j+1}^{i} n_{m-1,m}(t), & \text{if } i > j \\ n_{i,i}(t), & \text{if } i = j \\ n_{j,j}(t) - \sum_{m=i+1}^{j} n_{m-1,m}(t), & \text{if } i < j. \end{cases}$$

In the view of vehicle j, it can measure the state of vehicle i with an artificial sensor as (7). Under Assumption 2, it holds that for any $j \in \mathcal{N}_i$,

$$\left\| n_{i|j}(t) \right\| \leq (L+1)\mu =: \bar{\mu}. \tag{8}$$

Through the graph $\mathcal{G}$, vehicle $i \in \mathcal{V}_1$ is able to receive the absolute measurements (i.e., $\{y_{j,j}(t)\}$, $j \in \mathcal{N}_i$) and relative measurements (i.e., $\{y_{j-1,j}(t)\}$), and then calculate the measurements $\{y_{i|j}(t)\}_{j \in \mathcal{N}_i \bigcup \{i\}}$. Hence, it is feasible to reconstruct the measurement equation of vehicle $i \in \mathcal{V}_1$:

$$z_i(t) = Cx_i(t) + \boldsymbol{a}_i(t) + \boldsymbol{n}_i(t), \tag{9}$$

where $C = \begin{pmatrix} I_2 & I_2 & \cdots & I_2 \end{pmatrix}^{\mathsf{T}} \in \mathbb{R}^{(4L+2)\times 2}$, and

$$z_i(t) = (y_{i|i-L}^{\mathsf{T}}(t), y_{i|i-L+1}^{\mathsf{T}}(t), \ldots, y_{i|i+L}^{\mathsf{T}}(t))^{\mathsf{T}} \in \mathbb{R}^{4L+2},$$
$$\boldsymbol{a}_i(t) = (a_{i-L}^{\mathsf{T}}(t), a_{i-L+1}^{\mathsf{T}}(t), \ldots, a_{i+L}^{\mathsf{T}}(t))^{\mathsf{T}} \in \mathbb{R}^{4L+2},$$
$$\boldsymbol{n}_i(t) = (n_{i|i-L}^{\mathsf{T}}(t), n_{i|i-L+1}^{\mathsf{T}}(t), \ldots, n_{i|i+L}^{\mathsf{T}}(t))^{\mathsf{T}} \in \mathbb{R}^{4L+2}.$$

Remark 1. It follows from Assumption 1 that the attack signal $\boldsymbol{a}_i(t)$ has at most $2b$ non-zero elements, which means at least $4L + 2 - 2b$ elements of $z_i(t)$ are not compromised. If $L \geq b$, according to the sparse observability (Shoukry & Tabuada, 2016), the measurement redundancy in (9) enables us to design an effective resilient observer for vehicle $i \in \mathcal{V}_1$.

Next, we reconstruct the measurement equation of vehicle $i \in \mathcal{V}_2$ by using the messages from neighbor vehicles:

$$\hat{y}_{i|j} = x_i + \hat{n}_{i|j}, \quad j \in \mathcal{N}_i \bigcap \mathcal{V}_1 =: \hat{\mathcal{N}}_i, \tag{10}$$

where the reconstructed measurement $\hat{y}_{i|j}$ satisfies

$$\hat{y}_{i|j} = \begin{cases} \bar{x}_j - \sum_{m=i+1}^{j} y_{m-1,m} & \text{if } j > i \\ \bar{x}_j + \sum_{m=j+1}^{i} y_{m-1,m} & \text{if } j < i, \end{cases}$$

and the noise $\hat{n}_{i|j}$ is subject to

$$\hat{n}_{i|j} = \begin{cases} \bar{x}_j - x_j - \sum_{m=i+1}^{j} n_{m-1,m} & \text{if } j > i \\ \bar{x}_j - x_j + \sum_{m=i+1}^{j} n_{m-1,m} & \text{if } j < i. \end{cases} \tag{11}$$

As seen, vehicle $i \in \mathcal{V}_2$ uses the estimate $\bar{x}_j$ from neighbor vehicle j and the relative measurements $\{y_{m-1,m}\}$ from neighbor vehicle m, where $j \in \hat{\mathcal{N}}_i$ and $m \in \mathcal{N}_i$. In the next section, we will design a resilient observer for vehicles $i \in \mathcal{V}_1$ and $i \in \mathcal{V}_2$ with the reconstructed measurements in (9) and (10), respectively.

4. Observer design

In this section, we design an observer algorithm and analyze an asymptotic upper bound of the estimation error with a static observer threshold and an adaptive observer threshold, respectively. Since the observer algorithm to be designed uses the detection results, we need the following assumption in this section.

Assumption 3. The sets $\hat{\mathcal{S}}_i(t)$ and $\hat{\mathcal{S}}_i^a(t)$ introduced in (6) satisfy the following two properties:

(i) monotonically non-decreasing, i.e., $\hat{\mathcal{S}}_i^a(t_1) \subseteq \hat{\mathcal{S}}_i^a(t_2)$, and $\hat{\mathcal{S}}_i(t_1) \subseteq \hat{\mathcal{S}}_i(t_2)$, if $t_1 \leq t_2$;
(ii) no false alarm at each time, i.e., $\hat{\mathcal{S}}_i(t)$ and $\hat{\mathcal{S}}_i^a(t)$ are fault-free, $t = 1, 2, \ldots$.

This assumption is removed after we introduce the detector in Section 5. In other words, the integrated observer and detector in this paper satisfy Assumption 3 (see Lemma 1).

4.1. Observer algorithm

From the reconstructed measurement equation (9), we denote the innovation of vehicle $i \in \mathcal{V}_1$ by $z_i(t) - C\bar{x}_i(t) = \eta_i(t) = (\eta_{i,m_s}(t))_{s=1,2,\ldots,2L+1}$, where $m_s \in \mathcal{N}_i \cup \{i\}$, $\eta_{i,m_s}(t) \in \mathbb{R}^2$, and $\eta_i(t) \in \mathbb{R}^{4L+2}$. For example, when $L = 1$ and $i \in \{2, \ldots, N-1\}$, we have $m_1 = i-1$, $m_2 = i$, $m_3 = i+1$. For each vehicle $i \in \mathcal{V}$, given the sets $\{\hat{\mathcal{S}}_i(t), \hat{\mathcal{S}}_i^a(t)\}$ from the detector, we design the following observer by employing the measurements from (2), (9), and (10):

$$\hat{x}_i(t) = \begin{cases} \bar{x}_i(t) + \frac{1}{2L} C^{\mathsf{T}} K_i(t) \eta_i(t), & \text{if } i \in \mathcal{V}_1 \\ \bar{x}_i(t) + \frac{1}{\varpi}(y_{i,i}(t) - \bar{x}_i(t)), & \text{if } i \in \mathcal{V}_2 \bigcap \hat{\mathcal{S}}_i(t), \\ \bar{x}_i(t) + \frac{1}{\varpi}(\hat{y}_{i|j_i(t)}(t) - \bar{x}_i(t)), & \text{if } i \in \mathcal{V}_2 \setminus \hat{\mathcal{S}}_i(t), \end{cases} \tag{12}$$

where

$$\begin{aligned} &\varpi \in \left(1, \frac{\|A\|}{\|A\| - 1}\right) \\ &j_i(t) = \arg \min_{j \in \hat{\mathcal{N}}_i \cup \hat{\mathcal{S}}_i(t)} |j - i| \\ &K_i(t) = \operatorname{diag}\{k_{i,m_s}(t) I_2\}_{s=1,2,\ldots,2L+1}, \end{aligned} \tag{13}$$

where $\hat{\mathcal{N}}_i$ is introduced in (10), and $k_{i,m_s}(t)$ is designed by leveraging the following saturation method with a threshold $\beta_i(t) > 0$ (designed in Sections 4.2 and 4.3):

$$k_{i,m_s}(t) = \begin{cases} 0, & \text{if } m_s \in \hat{\mathcal{S}}_i^a(t) \\ 1, & \text{if } m_s \in \hat{\mathcal{S}}_i(t) \\ \min\left\{1, \frac{\beta_i(t)}{\|\eta_{i,m_s}(t)\|}\right\}, & \text{otherwise.} \end{cases} \tag{14}$$

Remark 2. The observer (12) shows: (i) For one sensor in the set $\mathcal{V}_1$, if it is attacked, i.e., $m_s \in \hat{\mathcal{S}}_i^a(t)$, its measurements are no longer employed, i.e., $k_{i,m_s}(t) = 0$; If it is attack-free, i.e., $m_s \in \hat{\mathcal{S}}_i(t)$, its measurements are fully trusted, i.e., $k_{i,m_s}(t) = 1$. Otherwise, the saturation method with the threshold $\beta_i(t)$ can reduce the influence of the potentially compromised measurements. (ii) For each vehicle $i \in \mathcal{V}_2$, if it is attack-free (i.e., $i \in \mathcal{V}_2 \bigcap \hat{\mathcal{S}}_i(t)$), it uses its own local measurements with full trust to update the state estimate, otherwise, it uses the estimate of vehicle $j_i(t)$ which is either in the set $\mathcal{V}_1$ with redundant measurements or in the set of attack-free vehicle sensors $\mathcal{V}_2 \bigcap \hat{\mathcal{S}}_i(t)$.

For each vehicle $i \in \mathcal{V}$, based on (9)–(10) and (12)–(14), we propose a resilient observer in Algorithm 1.

Algorithm 1 Resilient Observer

```
Initialization: Initial estimate $\hat{x}_i(0)$, observer parameter $\varpi$, saturation parameter $\{\beta_i(t)\}$, and vehicle communication parameter $L$
Output: State estimate $\hat{x}_i(t)$
for $t \geq 1$ do
    Communications between neighboring vehicles: Vehicle $i$ sends out $\mathcal{M}_i$ defined in (6);
    Time update: For each vehicle $i \in \mathcal{V}$;
        $\bar{x}_i(t) = A\hat{x}_i(t-1) + [0, Tu_i(t-1)]^{\mathrm{T}},$        (15)
    where $u_i(t-1)$ is specifically designed by vehicle $i$;
    Measurement update: See (12).
end for
```

Next, we study a real-time upper bound of the estimation error of Algorithm 1. In the following (a)–(c) items, we define three sequences, namely, $\rho_i(t)$, $\lambda_i(t)$, and $\tau_i(t)$, which are proved in Proposition 1 to be the upper bounds of the estimation errors of the three updates in (12).

(a) For vehicle $i \in \mathcal{V}_1$, we denote $\hat{\mathcal{S}}_{i,1}(t)$ the estimate of the set of attack-free vehicle sensors in the $2L$-neighborhood of vehicle sensor i, i.e.,

$$\hat{\mathcal{S}}_{i,1}(t) = \hat{\mathcal{S}}_i(t) \bigcap \left(\mathcal{N}_i \bigcup \{i\}\right). \tag{16}$$

Then, for $i \in \mathcal{V}_1$, we define a sequence $\{\rho_i(t)\}$ with $\rho_i(0) = q$ in the following

$$\rho_i(t) = \bar{m}_i(t)\,\|A\|\,\rho_i(t-1) + \bar{Q}_i(t), \tag{17}$$

where

$$\begin{aligned} \bar{m}_i(t) =& 1 - \frac{|\hat{\mathcal{S}}_{i,1}(t)| + (2L+1-b-|\hat{\mathcal{S}}_{i,1}(t)|)\bar{k}_i(t)}{2L}, \\ \bar{k}_i(t) =& \min\left\{1, \frac{\beta_i(t)}{\|A\rho_i(t-1)+\epsilon+\bar{\mu}\|}\right\}, \\ \bar{Q}_i(t) =& \frac{(\epsilon+\bar{\mu})(2L+1-b)+(b-|\hat{\mathcal{S}}^a_i(t)|)\beta_i(t)}{2L}. \end{aligned}$$

(b) For vehicle $i \in \mathcal{V}_2 \bigcap \hat{\mathcal{S}}_i(t)$, we define a sequence $\{\lambda_i(t)\}$, as follows

$$\lambda_i(t) = \frac{(\varpi-1)\,\|A\|}{\varpi}\lambda_i(t-1) + \frac{\epsilon(\varpi-1)+\mu}{\varpi}, \tag{18}$$

where the parameter ϖ is introduced in (13), $\lambda_i(T_i) = \tau_i(T_i)$, the sequence $\{\tau_i(t)\}$ is to be defined in (19), and T_i is the time after which vehicle sensor i is attack-free by detection, i.e., $T_i = \min \bar{t}$, s.t., $i \in \hat{\mathcal{S}}_i(\bar{t}+1)$.

(c) For vehicle $i \in \mathcal{V}_2 \setminus \hat{\mathcal{S}}_i(t)$, we define a sequence $\{\tau_i(t)\}$, as follows

$$\begin{aligned} \tau_i(t) =& \frac{(\varpi-1)\,\|A\|}{\varpi}\tau_i(t-1) \\ &+ \frac{\epsilon\varpi + \mu|j_i(t)-i| + \|A\|\,s_i(t-1)}{\varpi}, \end{aligned} \tag{19}$$

where $\tau_i(0) = q$, $j_i(t)$ is given in (13), and $s_i(t-1) = \rho_{j_i}(t-1)$, if $j_i(t) \in \mathcal{V}_1$, otherwise $s_i(t-1) = \lambda_{j_i}(t-1)$, where $\rho_{j_i}(t)$ and $\lambda_{j_i}(t)$ are given in (17) and (18), respectively.

Remark 3. Although the constructions of the two sequences $\{\lambda_i(t)\}$ and $\tau_i(t)$ need each other, they are both well defined because $\tau_i(t)$ starts at time $t = 0$, which does not require $\lambda_i(t)$, and $\lambda_i(t)$ starts at $t = T_i$.

Proposition 1. *Consider Algorithm 1 for system (1)–(3) satisfying Assumptions 1–3. The estimation error of each vehicle $i \in \mathcal{V}$ is subject to*

$$\left\|\hat{x}_i(t) - x(t)\right\| \leq \alpha_i(t) := \begin{cases} \rho_i(t), & \text{if } i \in \mathcal{V}_1, \\ \lambda_i(t), & \text{if } i \in \mathcal{V}_2 \bigcap \hat{\mathcal{S}}_i(t), \\ \tau_i(t), & \text{if } i \in \mathcal{V}_2 \setminus \hat{\mathcal{S}}_i(t), \end{cases}$$

where $\rho_i(t)$, $\lambda_i(t)$ $\tau_i(t)$ are given in (17), (18), and (19), respectively.

Proof. See Appendix A.

Remark 4. Based on local information and the vehicle-to-vehicle network $\mathcal{G}$, vehicle $i \in \mathcal{V}$ is able to compute the sequence $\{\alpha_i(t)\}$. It enables evaluation of the error bounds offline by setting $\hat{\mathcal{S}}^a_i(t) \equiv \hat{\mathcal{S}}_i(t) \equiv \emptyset$, which reduces to the case without detection.

Since the observer threshold $\beta_j(t), j \in \mathcal{V}_1$, in (14) is essential, we study the properties of Algorithm 1 by designing $\beta_j(t)$ in a static way and in an adaptive way respectively in the following two subsections.

4.2. Observer property with static threshold

In this subsection, we design the observer threshold $\beta_j(t) \equiv \beta_j$, for all $j \in \mathcal{V}_1$. Given a scalar $\omega \in (0, 1)$, denote

$$\begin{aligned} \beta_0 &= \|A\|q + \epsilon + \bar{\mu} \\ \bar{\beta}_1(\omega) &= \frac{2L}{2L+1-b}\,\frac{(\omega+\|A\|-1)\,\beta_0}{\|A\|} \\ \bar{\beta}_2(\omega) &= \min\left\{\beta_0, \frac{2L}{b}\left(\omega q - \frac{(\epsilon+\bar{\mu})\,(2L+1-b)}{2L}\right)\right\}, \end{aligned} \tag{20}$$

where $\bar{\mu}$ is defined in (8). In the following theorem, we study boundedness of the estimation error of the observer in Algorithm 1 with a static observer threshold $\beta_j, j \in \mathcal{V}_1$ introduced in (14).

Theorem 1. *Consider the observer in Algorithm 1 for system (1)–(3) satisfying Assumptions 1–3 and sets $\hat{\mathcal{S}}_i(T_i)$ and $\hat{\mathcal{S}}^a_i(T_i)$ for any $i \in \mathcal{V}$. If there is a scalar $\omega \in (0, 1)$, such that $0 < \bar{\beta}_1(\omega) < \bar{\beta}_2(\omega)$, then for any $\beta_j \in (\bar{\beta}_1(\omega), \bar{\beta}_2(\omega))$ with $j \in \mathcal{V}_1$, the estimation error of vehicle i is asymptotically upper bounded, i.e.,*

$$\limsup_{t\to\infty} \left\|\hat{x}_i(t) - x_i(t)\right\| \leq \begin{cases} \tilde{\alpha}_1, & \text{if } i \in \mathcal{V}_1, \\ \tilde{\alpha}_2, & \text{if } i \in \mathcal{V}_2 \bigcap \hat{\mathcal{S}}_i(T_i), \\ \tilde{\alpha}_3, & \text{if } i \in \mathcal{V}_2 \setminus \hat{\mathcal{S}}_i(T_i), \end{cases}$$

where $\bar{\beta}_1(\omega)$ and $\bar{\beta}_2(\omega)$ are defined in (20), and

$$\begin{aligned} \tilde{\alpha}_1 &= \frac{\tilde{Q}_i}{1-\tilde{m}_i\,\|A\|} \\ \tilde{\alpha}_2 &= \frac{\epsilon(\varpi-1)+\mu}{\varpi-(\varpi-1)\,\|A\|} \\ \tilde{\alpha}_3 &= \frac{\epsilon\varpi + \mu|j^*_i - i| + \|A\|\max\{\tilde{\alpha}_1, \tilde{\alpha}_2\}}{\varpi-(\varpi-1)\,\|A\|}, \end{aligned} \tag{21}$$

in which

$$\begin{aligned} j^*_i =& \arg\min_{j\in\hat{\mathcal{N}}_i\cup\hat{\mathcal{S}}_i(T_i)} |j-i|, \\ \tilde{Q}_i =& \frac{(\epsilon+\bar{\mu})(2L+1-b)+(b-|\hat{\mathcal{S}}^a_i(T_i)|)\beta}{2L}, \\ \tilde{m}_i =& 1 - \frac{|\hat{\mathcal{S}}_{i,1}(T_i)| + (2L+1-b-|\hat{\mathcal{S}}_{i,1}(T_i)|)k_i{}^*}{2L}, \\ k^*_i =& \frac{\beta}{\|A\|q+\epsilon+\bar{\mu}}, \\ \hat{\mathcal{S}}_{i,1}(T_i) =& \hat{\mathcal{S}}_i(T_i) \cap (\mathcal{N}_i \cup \{i\}). \end{aligned} \tag{22}$$

Proof. See Appendix B.

Theorem 1 is based on the available information at $T_i \geq 0$. If $T_i = 0$, $\hat{\mathcal{S}}_i(T_i) = \hat{\mathcal{S}}_i^a(T_i) = 0$, the corresponding bound is the worst bound which can be offline obtained. With the increase of T_i, $|\hat{\mathcal{S}}_i(T_i)|$ and $|\hat{\mathcal{S}}_i^a(T_i)|$ are non-decreasing. As a result, the error bound is non-increasing. Thus, it motivates us to design effective detector to enlarge the sets $\hat{\mathcal{S}}_i(T_i)$ and $\hat{\mathcal{S}}_i^a(T_i)$.

In the following proposition, we study the feasibility of the condition on ω in Theorem 1.

Proposition 2. *The condition on ω in Theorem 1 holds*

- *only if $b \leq L$;*
- *if*

$$\begin{aligned} &\frac{2L+1-b}{b} > \frac{\omega q + f_2}{\omega q - f_1} > 0 \\ &\frac{2L+1-b}{2L} > \frac{\omega + \|A\| - 1}{\|A\|} \end{aligned} \tag{23}$$

where $f_1 = \frac{(\epsilon+\bar{\mu})(2L+1-b)}{2L}$ and $f_2 = \frac{\omega(\epsilon+\bar{\mu})+(\|A\|-1)\beta_0}{\|A\|}$.

Proof. See Appendix C.

Remark 5. Proposition 2 does not provide a sufficient and necessary condition. Note that the sufficient condition in Proposition 2 holds if q in Assumption 2 is much larger than noise bounds ϵ and μ and the maximum number of attacked sensors b satisfies

$$b < \sup_{w\in(0,1)} \min\{f_3(w), f_4(w)\}, \tag{24}$$

where $f_3(w) = (L + \frac{1}{2})\left(1 - \frac{\|A\|-1}{2w+\|A\|-1}\right)$ and $f_4(w) = 1 + \frac{2L(1-w)}{\|A\|}$. Since b is an integer, it follows that $b \leq L$. Hence, the maximum number of attacked vehicle sensors that can be tolerated is $b = L = \lceil N/2 \rceil - 1$. This agrees with the sparse observability in Shoukry and Tabuada (2016), which shows that if half or more than half of the sensors are attacked, it is infeasible to recover the states of all vehicles.

4.3. Observer property with adaptive threshold

In this subsection, we design the observer threshold $\beta_j(t)$ in the following way: for $t \geq 1$,

$$\beta_j(t) = k_{j,0}\left(\|A\|\,\rho_j(t-1) + \epsilon + \bar{\mu}\right), \quad j \in \mathcal{V}_1, \tag{25}$$

where $\rho_j(\cdot)$ is introduced in (17), $\bar{\mu}$ is in (8), and $k_{j,0} = \frac{\beta_{j,0}}{\|A\|q+\epsilon+\bar{\mu}}$, in which $\beta_{j,0}$ is a positive scalar designed in the following theorem.

Theorem 2. *Consider the observer in Algorithm 1 for system (1)–(3) satisfying Assumptions 1–3 and sets $\hat{\mathcal{S}}_i(T_i)$ and $\hat{\mathcal{S}}_i^a(T_i)$ for any $i \in \mathcal{V}$. If there is a scalar $\omega \in (0, 1)$, such that $0 < \bar{\beta}_1(\omega) < \bar{\beta}_2(\omega)$, then the design of $\beta_j(t)$ in (25) with $\beta_{j,0} \in (\bar{\beta}_1(\omega), \bar{\beta}_2(\omega))$ for $j \in \mathcal{V}_1$ ensures that the estimation error of vehicle i is asymptotically upper bounded, i.e.,*

$$\limsup_{t\to\infty} \left\|\hat{x}_i(t) - x_i(t)\right\| \leq \begin{cases} \bar{\alpha}_1, & \text{if } i \in \mathcal{V}_1, \\ \bar{\alpha}_2, & \text{if } i \in \mathcal{V}_2 \bigcap \hat{\mathcal{S}}_i(T_i), \\ \bar{\alpha}_3, & \text{if } i \in \mathcal{V}_2 \setminus \hat{\mathcal{S}}_i(T_i), \end{cases}$$

where $\bar{\beta}_1(\omega)$ and $\bar{\beta}_2(\omega)$ are defined in (20), and

$$\begin{aligned} \bar{\alpha}_1 &= \frac{h_{i,2}(T_i)}{1 - h_{i,1}(T_i)\,\|A\|} \\ \bar{\alpha}_2 &= \tilde{\alpha}_2 \\ \bar{\alpha}_3 &= \frac{\epsilon\varpi + \mu|j_i^* - i| + \|A\| \max\{\bar{\alpha}_1, \bar{\alpha}_2\}}{\varpi - (\varpi - 1)\,\|A\|}. \end{aligned} \tag{26}$$

in which

$$h_{i,1}(T_i) = 1 - \frac{|\hat{\mathcal{S}}_{i,1}(T_i)| + (\bar{L} - b + |\hat{\mathcal{S}}_i^a(T_i)| - |\hat{\mathcal{S}}_{i,1}(T_i)|)k_{i,0}}{2L},$$

$$h_{i,2}(T_i) = \frac{\bar{L} + (b - |\hat{\mathcal{S}}_i^a(T_i)|)k_{i,0}}{2L}(\epsilon + \bar{\mu}),$$

$$k_{i,0} = \frac{\beta_{i,0}}{\|A\|q + \epsilon + \bar{\mu}},$$

where $\bar{L} = 2L + 1 - b$, the scalar j_i^* and the set $\hat{\mathcal{S}}_{i,1}(T_i)$ are the same as in (22), and the scalar $\tilde{\alpha}_2$ is in (21).

Proof. See Appendix D.

Remark 6. To ensure the required steady performance, from Theorems 1–2, β_j cannot be very large, but $\beta_{j,0}$ can, where $j \in \mathcal{V}_1$. Since a larger threshold makes the estimation become steady faster (see (17)), the adaptive threshold $\beta_j(t)$ enables the system to have better dynamic and steady performance than the static threshold β_j.

5. Detector design

In this section, we design an attack detector algorithm and then study when all attacked and attack-free vehicle sensors can be identified by the detector in finite time.

5.1. Detector algorithm

We propose an online distributed attack detector in Algorithm 2 based on the observer in Algorithm 1 and the detection conditions (27)–(29) as follows. For $i \in \mathcal{V}$ (and $i \geq 2$ for (27))

$$\|y_{i-1,i}(t) + y_{i-1,i-1}(t) - y_{i,i}(t)\| > 3\mu, \tag{27}$$

$$\|y_{i,i}(t) - A\hat{x}_i(t-1)\| > g_i(t), \tag{28}$$

$$\sum_{j=1}^{l_i} \lceil |\overline{\hat{\mathcal{S}}^s}_{i,j}(t)|/3 \rceil = b. \tag{29}$$

where $g_i(t) = \epsilon + \mu + \|A\|\,\rho_i(t-1)$ if $i \in \mathcal{V}_1$, otherwise, $g_i(t) = \epsilon + \mu + \|A\|\,\tau_i(t-1)$, in which $\rho_i(t-1)$ and $\tau_i(t-1)$ are generated through (17) and (19), respectively. Here, $\{\overline{\hat{\mathcal{S}}^s}_{i,j}(t)\}_{j=1}^{l_i}$ are disjoint subsets of $\overline{\hat{\mathcal{S}}^s}_i(t) := \hat{\mathcal{S}}_i^s(t) \bigcup \hat{\mathcal{S}}_i^a(t)$, such that each subset has successive sensor labels and the union of all subsets is equal to $\overline{\hat{\mathcal{S}}^s}_i(t)$.

5.2. Detector properties

Lemma 1. *Consider system (1)–(3) and Algorithms 1–2 under Assumptions 1–2. Then the sets $\hat{\mathcal{S}}_i(t)$ and $\hat{\mathcal{S}}_i^a(t)$ of Algorithm 2 satisfies Assumption 3.*

Algorithm 2 Online Attack Detector

```
Initialization: Initial estimate for attacked vehicle sensor set $\hat{\mathcal{S}}_i^a(0) = \emptyset$, initial estimate for suspicious vehicle set $\hat{\mathcal{S}}_i^s(0) = \emptyset$, and initial estimate for attack-free vehicle set $\hat{\mathcal{S}}_i(0) = \emptyset$, $i \in \mathcal{V}$.
Output: Sets $\hat{\mathcal{S}}_i^a(t)$, $\hat{\mathcal{S}}_i^s(t)$, and $\hat{\mathcal{S}}_i(t)$
for $t \geq 1$ do
    Communications between neighboring vehicles: Vehicle $i$ sends out $\mathcal{M}_i$ defined in (6)  Each vehicle $i$ fuses the sets from its neighbors: $\hat{\mathcal{S}}_i^a(t) = \cup_{j \in \mathcal{N}_i} \hat{\mathcal{S}}_j^a(t-1) \cup \hat{\mathcal{S}}_i^a(t-1)$, $\hat{\mathcal{S}}_i^s(t) = \cup_{j \in \mathcal{N}_i} \hat{\mathcal{S}}_j^s(t-1) \cup \hat{\mathcal{S}}_i^s(t-1)$, $\hat{\mathcal{S}}_i(t) = \cup_{j \in \mathcal{N}_i} \hat{\mathcal{S}}_j(t-1) \cup \hat{\mathcal{S}}_i(t-1)$
    if $i \geq 2$, and $i \notin \hat{\mathcal{S}}_i^a(t)$, and $i-1 \notin \hat{\mathcal{S}}_i^a(t)$ then
        if (27) holds then
            if $i \in \hat{\mathcal{S}}_i(t)$ then
                let $\hat{\mathcal{S}}_i^a(t) = \hat{\mathcal{S}}_i^a(t) \cup \{i-1\}$
            else if $i-1 \in \hat{\mathcal{S}}_i(t)$ then
                let $\hat{\mathcal{S}}_i^a(t) = \hat{\mathcal{S}}_i^a(t) \cup \{i\}$
            else
                let $\hat{\mathcal{S}}_i^s(t) = \hat{\mathcal{S}}_i^s(t) \cup \{i-1, i\}$
            end if
        end if
    end if
    if $i \notin \hat{\mathcal{S}}_i^a(t)$ and $i \notin \hat{\mathcal{S}}_i(t)$ then
        if (28) holds then
            let $\hat{\mathcal{S}}_i^a(t) = \hat{\mathcal{S}}_i^a(t) \cup \{i\}$
        end if
    end if
    if (29) holds then
        $\hat{\mathcal{S}}_i(t) = \hat{\mathcal{S}}_i(t) \cup \big(\mathcal{V} - \hat{\mathcal{S}}_i^s(t) - \hat{\mathcal{S}}_i^a(t)\big)$
    end if
    if $|\hat{\mathcal{S}}_i^a(t)| = b$ then
        $\hat{\mathcal{S}}_i(t) = \mathcal{V} - \hat{\mathcal{S}}_i^a(t)$
    end if
end for
```

Proof. See Appendix E.

Remark 7. Condition (29) is to infer whether all attacked sensors have been included in the detected sensor sets. The basic idea is that one attacked sensor can yield at most three suspicious sensors including itself and its two neighbors. For example, suppose $\hat{\mathcal{S}}_i^s(t) = \{1, 3, 9, 10, 11, 12\}$ and $\hat{\mathcal{S}}_i^a(t) = \{2, 6, 15\}$, then $\overline{\hat{\mathcal{S}}^s}_i(t) = \{1, 2, 3, 6, 9, 10, 11, 12, 15\}$. By splitting $\overline{\hat{\mathcal{S}}^s}_i(t)$, we have $\overline{\hat{\mathcal{S}}^s}_{i,1}(t) = \{1, 2, 3\}$, $\overline{\hat{\mathcal{S}}^s}_{i,2}(t) = \{6\}$, $\overline{\hat{\mathcal{S}}^s}_{i,3}(t) = \{9, 10, 11, 12\}$, and $\overline{\hat{\mathcal{S}}^s}_{i,4}(t) = \{15\}$. Then there are at least five attacked sensors in the set $\overline{\hat{\mathcal{S}}^s}_i(t)$. Because $\overline{\hat{\mathcal{S}}^s}_{i,1}(t)$ has at least one, $\overline{\hat{\mathcal{S}}^s}_{i,2}(t)$ has one, $\overline{\hat{\mathcal{S}}^s}_{i,3}(t)$ has at least two, and $\overline{\hat{\mathcal{S}}^s}_{i,4}(t)$ has one. If $b = 5$, then by (29), we conclude that the sensors not belonging to these sets are attack-free.

Lemma 1 states that the two sets $\hat{\mathcal{S}}_i(t)$ and $\hat{\mathcal{S}}_i^a(t)$ are fault-free, which differs from the existing results of false alarms (e.g., Baras and Liu (2019)). The following theorem studies the finite-time convergence of the detection sets $\hat{\mathcal{S}}_i^a(t)$ and $\hat{\mathcal{S}}_i(t)$.

Theorem 3. *Consider the observer in Algorithm 1 and the detector in Algorithm 2 for system (1)–(3) under Assumptions 1–2. If there is a time T_j and a vehicle $j \in \mathcal{V}$, such that the number of the attacked vehicle sensors estimated by vehicle j is equal to its upper bound in Assumption 1, i.e., $|\hat{\mathcal{S}}_j^a(T_j)| = b$, then there exists a time T_*, such that for $t \geq T_*$, the sets of attacked and attack-free vehicle sensors estimated by each vehicle $i \in \mathcal{V}$ equal the true sets, i.e.,*

$$\hat{\mathcal{S}}_i^a(t) = \mathcal{S}^a, \quad \hat{\mathcal{S}}_i(t) = \mathcal{S}.$$

Proof. By Algorithm 2, when there is a time T_j and a vehicle $j \in \mathcal{V}$, such that $|\hat{\mathcal{S}}_j^a(T_j)| = b$, then $\hat{\mathcal{S}}_j^a(T_j) = \mathcal{S}^a$ and $\hat{\mathcal{S}}_j(T_j) = \mathcal{S}$. Since both $|\hat{\mathcal{S}}_i^a(t)|$ and $|\hat{\mathcal{S}}_i(t)|$ are non-decreasing and the vehicle network is finite, there is a time at which all vehicles update their set estimates to the true sets.

Algorithm 3 Distributed Controller

```
Initialization: Initial estimates $\hat{x}_i(0)$ and $\bar{x}_i(0)$, control parameter $g_s$ and $g_v$, desired relative position and velocity between vehicles $i-1$ and $i$, i.e., $\{\Delta x_{i-1,i}^s(t)\}$ and $\{\Delta x_{i-1,i}^v(t)\}$, $i = 1, 2, \ldots, N$
Output: Control input $u_i(t)$
for $t \geq 0$ do
    Communications between neighboring vehicles: Vehicle $i$ sends out $\mathcal{M}_i$ defined in (6)
    Distributed controller
        $$u_i(t) = \sum_{j \in \bar{\mathcal{N}}_i} \big(g_s(\bar{s}_j(t) - \hat{s}_i(t) + \Delta x_{j,i}^s(t))$$
        $$+ g_v(\bar{v}_j(t) - \hat{v}_i(t) + \Delta x_{j,i}^v(t))\big),$$
    where $[\bar{s}_0(t), \bar{v}_0(t)]^{\top} =: x_0(t)$.
end for
```

Theorem 3 holds under the condition that the attacker compromises b sensors with aggressive attack signals, which is possible when the attacker has no knowledge of the detector. Otherwise, the attacker can inject stealthy signals making the attacked sensors undetectable.

6. Controller design

In this section, we design an observer-based distributed controller algorithm, and then analyze boundedness of the overall performance function of the architecture consisting of the observer in Algorithm 1, the detector in Algorithm 2, and the distributed controller.

6.1. Controller algorithm

Denote $\bar{\mathcal{N}}_i$ the set of vehicle(s) nearest to vehicle i, $i = 0, 1, \ldots, N$, i.e.,

$$\bar{\mathcal{N}}_i = \begin{cases} \{1\}, & \text{if } i = 0 \\ \{i-1, i+1\}, & \text{if } i \in \{1, 2, \ldots, N-1\} \\ \{N-1\}, & \text{if } i = N, \end{cases} \qquad (30)$$

where vehicle 0, which is virtual and introduced for convenience, stands for the reference state of vehicle 1. Assume $\hat{s}_i(t)$ and $\bar{s}_i(t)$ are the estimate and predicted value of $s_i(t)$, and $\hat{v}_i(t)$ and $\bar{v}_i(t)$ are the estimate and predicted value of $v_i(t)$. Then, we propose a distributed observer-based controller in Algorithm 3, where $\Delta x_{i-1,i}^s(t)$ and $\Delta x_{i-1,i}^v(t)$ are the desired relative position and velocity between vehicles $i-1$ and i, and $g_s > 0$, $g_v > 0$ are parameters to be determined.

Remark 8. The relative state measurements in (3) are not directly used in the controller but the estimates, because: (i) The relative measurements are noisy. (ii) There is no sensor of the leader vehicle to measure the relative state to the reference state (i.e., $x_1(t) - x_0(t)$).

6.2. Closed-loop property

The following lemma, proved in He, Hashemi, and Johansson (2020), is useful in the following analysis.

Lemma 2. *Consider the linear system $x(t+1) = Fx(t) + G(t)$, where $F \in \mathbb{R}^{n\times n}$ is a Schur stable matrix. If $\limsup_{t\to\infty} \|G(t)\| \le \varsigma$, then $\limsup_{t\to\infty} \|x(t)\| \le \sqrt{\frac{2\theta\varsigma^2\sigma_{\max}(P)}{\sigma_{\min}(P)}}$, where $P \succ 0$ is the unique solution to $F^{\mathsf{T}}PF - P = -I_n$ and $\theta = \|P\| + 2\,\|PF\|^2$.*

Let $\mathcal{L} \in \mathbb{R}^{(N+1)\times(N+1)}$ be the graph Laplacian matrix (Xie & Wang, 2012) corresponding to the neighbor sets in (30). Denote $\mathcal{L}_g \in \mathbb{R}^{N\times N}$ the grounded graph Laplacian matrix with respect to the nodes $\{1, 2, 3, \dots, N\}$, which is obtained by removing the first row and first column of Laplacian matrix $\mathcal{L}$.

Assumption 4. The parameters g_s and g_v of the controller in Algorithm 3 are subject to $g_v > Tg_s > 0$ and $T^2 g_s - 2Tg_v > -4/\sigma_{\max}(\mathcal{L}_g)$.

Assumption 4 is satisfied for any positive g_s and g_v provided that the sampling time $T > 0$ is sufficiently small. In the following theorem, the closed-loop performance function $\varphi(t)$ in (4) is studied.

Theorem 4. *Consider the observer in Algorithm 1, the detector in Algorithm 2, and the controller in Algorithm 3 satisfying Assumption 4 for system (1)–(3). Then the following properties hold:*

(i) If the observer threshold is static and the conditions in Theorem 1 are satisfied, the performance function $\varphi(t)$ in (4) is asymptotically upper bounded, i.e.,

$$\limsup_{t\to\infty} \varphi(t) \le \hat{\alpha} + \eta\xi;$$

(ii) If the observer threshold is adaptive and the conditions in Theorem 2 are satisfied, $\varphi(t)$ is asymptotically upper bounded, i.e.,

$$\limsup_{t\to\infty} \varphi(t) \le \bar{\hat{\alpha}} + \bar{\eta}\xi;$$

where

$$\begin{aligned}
&\xi = \sqrt{\frac{2\kappa\sigma_{\max}(M)}{\sigma_{\min}(M)}},\ M = \sum_{i=0}^{\infty}(P_0^i)^{\mathsf{T}}P_0^i,\ F_0 = \left(\begin{smallmatrix} 0 & 0 \\ Tg_s & Tg_v \end{smallmatrix}\right),\\
&P_0 = I_N \otimes A - \mathcal{L}_g \otimes F_0, \quad \kappa = \|M\| + 2\|MP_0\|^2,\\
&\eta = 2\sqrt{N}T\hat{\alpha}\,(g_s(\|A\|+1) + 2g_v) + \sqrt{N}\epsilon,\\
&\bar{\eta} = 2\sqrt{N}T\bar{\hat{\alpha}}\,(g_s(\|A\|+1) + 2g_v) + \sqrt{N}\epsilon,\\
&\hat{\alpha} = \max\{\tilde{\alpha}_1, \tilde{\alpha}_2, \tilde{\alpha}_3\}, \quad \bar{\hat{\alpha}} = \max\{\bar{\alpha}_1, \bar{\alpha}_2, \bar{\alpha}_3\},
\end{aligned} \tag{31}$$

in which $\tilde{\alpha}_i$ and $\bar{\alpha}_i$, for $i = 1, 2, 3$, are introduced in Theorems 1 and 2, respectively.

Proof. See Appendix F.

Theorem 4 and the following corollary provide the solution to the problem in Section 2.4.

Corollary 1. *Consider the observer in Algorithm 1, the detector in Algorithm 2, and the controller in Algorithm 3 satisfying Assumption 4 for system (1)–(3). Then the performance function $\varphi(t)$ tends to zero, i.e.,*

$$\lim_{t\to\infty} \varphi(t) = 0,$$

if the system is noise-free, i.e., $\mu = \epsilon = 0$, and one of the following two conditions is satisfied:

(i) the observer threshold is static, the conditions in Theorem 1 hold, and there is a vehicle sensor i at some $T_i < \infty$, such that $|\hat{\mathcal{S}}_i^a(T_i)| = b$;

(ii) the observer threshold is adaptive, and the conditions in Theorem 2 hold.

Proof. The proof follows from Theorems 1–4.

Remark 9. Corollary 1 shows the improvement of performance achieved in the noise-free case in comparison to the noisy case Theorem 4. Note that the first conclusion of Corollary 1 means that there is one vehicle that has detected the maximal number of attacked sensors. This makes it possible to conclude that there can be no other attacked sensors, so the mitigation mechanism of the observer can fully compensate for the attack. The second conclusion of Corollary 1 means that whatever the detection results, the observer with the adaptive threshold makes the space of stealthy attacks diminish to an empty set asymptotically.

7. Simulations

In this section, the effectiveness of the proposed methods is evaluated through simulations by an application to vehicle platooning.

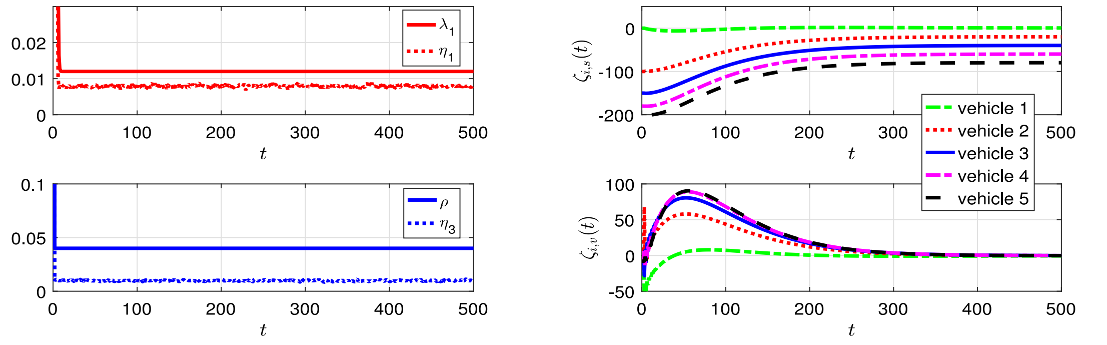
Suppose there are five vehicles, i.e., $N = 5$, with sampling time $T = 0.01$ and time range $t = 0, 1, \dots, 500$. All elements of the process noise $d_i(t)$ and measurement noise $n_{i,j}(t)$, $j \in \mathcal{N}_i \cup \{i\}$, $i = 1, \dots, 5$, follow the uniform distribution between $(0, \mu_0/\sqrt{2})$, where $\mu_0 = 0.1$. The bounds in Assumption 2 are assumed to be $\mu = \epsilon = \mu_0$ and $q = 300$. The initial state is $x_1(0) = (200, 10)^{\mathsf{T}}$, $x_2(0) = (100, 8)^{\mathsf{T}}$, $x_3(0) = (50, 6)^{\mathsf{T}}$, $x_4(0) = (20, 4)^{\mathsf{T}}$, $x_5(0) = (0, 2)^{\mathsf{T}}$, whose observer estimates are all $0^{2\times 1}$. The required position distance between vehicles i and $i+1$ is $|\Delta_{i,i+1}| = 20$, $i = 1, 2, 3, 4$. The control gains in Algorithm 3 are $g_s = g_v = 50$, and the communication range $L = 2$. Suppose the reference position and the reference velocity of the leader vehicle are $s_0(t+1) = s_0(t) + v_0 T$ and $v_0 = 10$, where $s_0(0) = 200$. We assume all vehicles share the same observer threshold $\beta(\cdot)$. In the following, the time-varying observer gain $\beta(t)$ is designed as in (25) with $\beta(0) = 200$.

We conduct a Monte Carlo experiment with 100 runs. Denote the maximum estimation error by $\eta_i(t) := \max_{j=1}^{100} \left\| e_i^j(t) \right\|$, where $e_i^j(t)$ is the state estimation error of vehicle i at time t in the jth run. Define the relative position and velocity between vehicle $i = 1, \dots, 5$ and the leader vehicle 0 by

$$\zeta_{i,s}(t) = \frac{\sum_{j=1}^{100}(s_i^j(t) - s_0(t))}{100}, \quad \zeta_{i,v}(t) = \frac{\sum_{j=1}^{100}(v_i^j(t) - v_0)}{100},$$

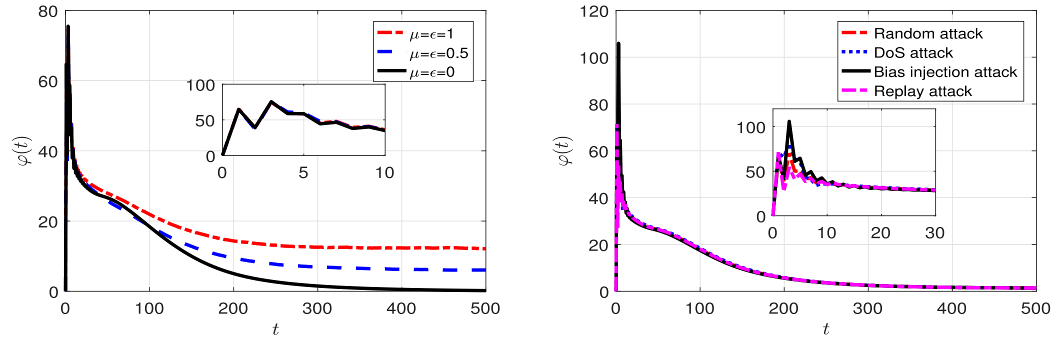
where $s_i^j(t)$ and $v_i^j(t)$ are the position and velocity of vehicle i, respectively, at time t in the jth run.

First, we study the performance of Algorithms 1–3 with the adaptive observer parameter $\beta(t)$. For vehicle i under FDI sensor attacks, assume that the measurements would be compromised by the random attack signal $a_i(t) = w_i(t)x_i(t)$, where $w_i(t)$ is drawn from the standard normal distribution. For the case of the attacked vehicle sensor set $\mathcal{S}^a = \{3\}$, the state estimation errors and error bounds for vehicles 1 and 3, and vehicle platooning errors are provided in Fig. 4. Fig. 4(a) shows that the estimation errors and their upper bounds are convergent to small neighborhoods of zero rapidly. By Algorithm 2, vehicle 1 is known to be attack-free, thus it uses its secure measurements to update the estimate, achieving better performance than attacked vehicle 3 as in Fig. 4(a). Fig. 4(b) shows that the velocities of all vehicles converge to the reference velocity, and the relative positions between two neighbor vehicles tend to the desired one, i.e., 20. Then, we study the performance function $\varphi(t)$ (averaged over 100 runs) of Algorithms 1–3 with $\mathcal{S}^a = \{2, 3\}$ under different noise magnitudes (i.e., ϵ and μ) and under different attacks in (a) and (b) of Fig. 5, respectively. Fig. 5(a) shows that $\varphi(t)$ decreases as the noise magnitudes decrease. In Fig. 5(b), we study four typical attack types, including random attack introduced previously, DoS



(a) Estimation errors and error bounds for vehicles 1 and 3 (b) Relative states between reference state and the vehicles

Fig. 4. Estimation and platooning errors of Algorithms 1–3.



(a) The error function $\varphi(t)$ with different noise magnitudes (b) The error function $\varphi(t)$ under different attacks

Fig. 5. The influence of some essential variables to the performance of Algorithms 1–3.

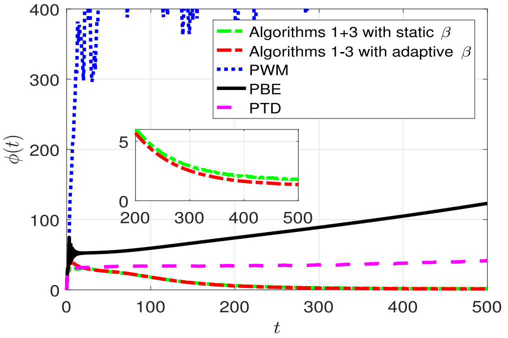


Fig. 6. Comparison of five algorithms in platooning error.

attack $a_i(t) = -\tilde{y}_{i,i}(t)$, bias injection attack $a_i(t) = 100$, and replay attack $a_i(t) = \tilde{y}_{i,i}(t-10) - \tilde{y}_{i,i}(t)$ for $t \geq 11$, where $\tilde{y}_{i,i}(t) = x_i(t) + n_{i,i}(t)$ is the true measurement. It shows that Algorithms 1–3 with adaptive observer parameter are able to deal with these attacks.

Moreover, under the same setting as Fig. 4, we compare the proposed methods, i.e., Algorithms 1+3 (1 and 3) with a static observer parameter β, Algorithms 1–3 with an adaptive observer parameter $\beta(t)$, with PWM, which is obtained from Algorithm 3 by replacing the estimates by measurements, and with PBE, which is obtained from Algorithm 3 by using the estimates following Byzantine strategy (Mitra & Sundaram, 2019), as well as PTD (Lin & Jia, 2009). Choose $\beta = 3$ for Algorithms 1+3, and $\beta(t)$ for Algorithms 1–3. To evaluate the platooning error of each algorithm, we use the performance function $\phi(t)$: $\phi(t) = \frac{1}{500}\sum_{j=1}^{100}\sum_{i=1}^{5}\left\|x_i^j(t) - x_i^*(t)\right\|$, where $x_i^j(t)$ is the state of vehicle i at time t in the jth run. The algorithm comparison result is provided in Fig. 6, which shows that our algorithms outperform the other three algorithms, and Algorithms 1–3 achieve best platooning performance among the five algorithms. In Fig. 6, PWM is divergent since the compromised measurements directly affect the platooning.

8. Conclusion and future work

This paper studied how to design a secure observer-based distributed controller such that a group of vehicles can achieve accurate state estimates and formation control under the case that the measurements of a subset of vehicle sensors are compromised by a malicious attacker. We proposed an architecture consisting of a resilient observer, an online attack detector, and a distributed controller. Some important properties of the observer, detector, and controller were analyzed. An application of the proposed architecture to vehicle platooning was investigated in numerical simulations.

There are some directions of future work. One is to extend the architecture to the attack detection on actuators of vehicles in platoon. Another is to study more general models of vehicles and sensors. It is also promising to extend the methods from the string vehicle topology to more complex vehicle topologies with higher dimensions and more leaders.

Appendix A. Proof of Proposition 1

Denote the estimation error by $e_i(t) = \hat{x}_i(t) - x_i(t)$, the prediction error by $\bar{e}_i(t) = \bar{x}_i(t) - x_i(t)$, $i \in \mathcal{V}$. For notational convenience, we let $\lambda_i(t) = \tau_i(t)$, $t \leq T_i$, where T_i is the time after which vehicle i is attack-free by detection, i.e., $i \in \hat{\mathcal{S}}_i(t)$, $t \geq T_i + 1$. We use an inductive method for proof. At the initial time, due to $\rho_i(0) = \lambda_i(0) = \tau_i(0) = q$, according to Assumption 2, the conclusion holds. Assume at time $t-1 \geq 0$, the conclusion holds. In the following, we consider the case at time $t \geq 1$.

First, we consider each vehicle sensor $i \in \mathcal{V}_1$, which has at least $2L+1-b$ attack-free vehicle sensors as neighbors. Suppose $\mathcal{J}$ is the set of these $2L+1-b$ sensors, i.e., $\mathcal{J} \subseteq \mathcal{S}$ with $|\mathcal{J}| = 2L+1-b$, which is unknown to vehicles but useful for the following analysis. Let $\mathcal{J}^a = \mathcal{N}_i \cup \{i\} - \mathcal{J}$. It holds that $|\mathcal{J}^a| = b$ and the sensors in the set $\hat{\mathcal{S}}_i^a(t) \subseteq \mathcal{J}^a$ are surely attacked under Assumption 3. Denote $\bar{K}_{i,\mathcal{J}}(t) = \operatorname{diag}\left\{k_{i,m_s}(t)\mathbb{I}_{m_s\in\mathcal{J}}I_2\right\}_{s=1}^{2L+1} \in \mathbb{R}^{(4L+2)\times(4L+2)}$ where $k_{i,m_s}(t)$ is introduced in (14). Let $\bar{K}_i^{[j]}(t)$ be the jth diagonal element of $\bar{K}_{i,\mathcal{J}}(t)$, $j = 1, \ldots, 4L+2$, $\boldsymbol{n}_i^{[j]}(t)$ be the jth element of $\boldsymbol{n}_i(t)$ in (9), and

$$\hat{K}_i(t) = \operatorname{diag}\left\{\sum_{j=1,3,\ldots,4L+1} \bar{K}_i^{[j]}(t), \sum_{j=2,4,\ldots,4L+2} \bar{K}_i^{[j]}(t)\right\},$$

$$W_i(t) = \sum_{j=1,3,\ldots,4L+1} \begin{pmatrix} \bar{K}_i^{[j]}(t)\boldsymbol{n}_i^{[j]}(t) \\ \bar{K}_i^{[j+1]}(t)\boldsymbol{n}_i^{[j+1]}(t) \end{pmatrix},$$

through which we have $\hat{K}_i(t) \in \mathbb{R}^{2\times 2}$ and $W_i(t) \in \mathbb{R}^2$. By Algorithm 1, we have

$$\begin{aligned} e_i(t) =& (I_2 - \frac{1}{2L}\hat{K}_i(t))Ae_i(t-1) + \frac{1}{2L}\hat{K}_i(t)d_i(t-1) \\ &+ \frac{1}{2L}W_i(t) + \frac{1}{2L}C^{\top}\bar{K}_{i,\mathcal{J}^a}(t)(z_i(t) - C\bar{x}_i(t)), \end{aligned}$$

where $\bar{K}_{i,\mathcal{J}^a}(t) = K_i(t) - \bar{K}_{i,\mathcal{J}}(t)$. According to (14), the measurement update of sensor i at time t will be affected by at most $b - |\hat{\mathcal{S}}_i^a(t)|$ attacked vehicle sensors, which remain stealthy till time t. The measurements of these vehicles will be used at time t. According to the noise bound in (8) and the saturation operation in (14), taking 2-norm of $e_i(t)$ yields

$$\begin{aligned} \|e_i(t)\| \leq& \|(I_2 - \frac{1}{2L}\hat{K}_i(t))A\| \, \|e_i(t-1)\| \\ &+ |\mathcal{J}|\frac{\epsilon + \bar{\mu}}{2L} + (b - |\hat{\mathcal{S}}_i^a(t)|)\frac{\beta_i(t)}{2L} \leq \rho_i(t), \end{aligned}$$

where the last inequality is obtained because: (1) In the set $\mathcal{J}$, there are $|\hat{\mathcal{S}}_{i,1}(t)|$ attack-free vehicles whose measurements have been fully utilized in the update at time t (i.e., without saturation), where $\hat{\mathcal{S}}_{i,1}(t)$ is defined in (16); (2) There are $2L+1-b-|\hat{\mathcal{S}}_{i,1}(t)|$ attack-free vehicles, whose measurement innovations are saturated with the corresponding gain satisfying $\hat{K}_i^{[j]}(t) \geq \bar{k}_i(t) = \min\{1, \frac{\beta_i(t)}{\|A\|\rho_i(t-1)+\epsilon+\bar{\mu}}\}$.

Second, for vehicle $i \in \mathcal{V}_2 \bigcap \hat{\mathcal{S}}(t)$, according to (12) and Assumption 2, it is straightforward to prove that the estimation error is upper bounded by $\lambda_i(t)$. Third, for vehicle $i \in \mathcal{V}_2 - \hat{\mathcal{S}}(t)$, by Algorithm 1, we have

$$e_i(t) = \frac{(\varpi - 1)A}{\varpi}e_i(t-1) - \frac{(\varpi-1)d_i(t-1)}{\varpi} + \frac{\hat{n}_{i|j_i(t)}(t)}{\varpi}.$$

Regarding $\hat{n}_{i|j_i(t)}(t)$ in (11), according to Assumption 2, the definition $j_i(t) = \arg\min_{j\in\hat{\mathcal{N}}_i\cup\hat{\mathcal{S}}_i(t)} |j-i|$, and $\left\|\bar{e}_{j_i(t)}(t)\right\| \leq \|A\| \, s_i(t-1) + \epsilon$, we have $\left\|\hat{n}_{i|j_i(t)}(t)\right\| \leq \mu|j_i(t) - i| + \|A\| \, s_i(t-1) + \epsilon$, where $s_i(t-1) = \rho_{j_i}(t-1)$, if $j_i(t) \in \mathcal{V}_1$, otherwise $s_i(t-1) = \lambda_{j_i}(t-1)$. Taking 2-norm of both sides of $e_i(t)$, we have $\|e_i(t)\| \leq \tau_i(t)$.

Appendix B. Proof of Theorem 1

At time $T_i \geq 0$, the estimate of the attacked vehicle sensor set is $\hat{\mathcal{S}}_i^a(T_i)$ and the estimate of the attack-free vehicle set is $\hat{\mathcal{S}}(T_i)$. By Assumption 3, both $|\hat{\mathcal{S}}_i^a(t)|$ and $|\hat{\mathcal{S}}_i(t)|$ are non-decreasing, thus $|\hat{\mathcal{S}}_i^a(t)| \geq |\hat{\mathcal{S}}_i^a(T_i)|$ and $|\hat{\mathcal{S}}_i(t)| \geq |\hat{\mathcal{S}}_i(T_i)|$, for any $t \geq T_i$. Instead of proving the upper boundedness of the estimation error, in the following we prove the upper boundedness of $\rho_i(t)$, $\lambda_i(t)$, and $\tau_i(t)$, which are upper bounds of the estimation error according to Proposition 1.

First, we consider the case for $i \in \mathcal{V}_1$. By choosing $\forall \beta_i \in (\bar{\beta}_1(\omega), \bar{\beta}_2(\omega))$, where $\bar{\beta}_1(\omega)$ and $\bar{\beta}_2(\omega)$ are in (20), we directly have

$$\beta_i < \beta_0 \tag{B.1}$$

$$\beta_i < \frac{2L}{b}\left(\omega q - \frac{(\epsilon+\bar{\mu})(2L+1-b)}{2L}\right) \tag{B.2}$$

$$\beta_i > \frac{2L}{2L+1-b}\frac{(\omega + \|A\| - 1)\,\beta_0}{\|A\|}. \tag{B.3}$$

It follows from (B.1) that $k_i^* := \frac{\beta_i}{\|A\|q+\epsilon+\bar{\mu}} < 1$. Then according to (B.3), it is derived that $(1 - L_0k_i^*)\,\|A\|\,q < (1-\omega)q$, where $L_0 = \frac{2L+1-b}{2L}$. Since the inequality in (B.2) is equivalent to $\frac{(\epsilon+\bar{\mu})(2L+1-b)+b\beta_i}{2L} < \omega q$, we have

$$(1 - L_0k_i^*)\,\|A\|\,q + \frac{(\epsilon+\bar{\mu})(2L+1-b)+b\beta_i}{2L} < q. \tag{B.4}$$

From (B.4) and Proposition 1, by using an inductive method, we are able to obtain that $\rho_i(t) < q$, for $t \geq 1$, which, together with (17), ensures that

$$\rho_i(t+1) \leq \tilde{m}_i\,\|A\|\,\rho_i(t) + \tilde{Q}_i, \quad t \geq T_i, \tag{B.5}$$

where $\tilde{m}_i$ and $\tilde{Q}_i$ are given in (22). According to (B.4), we have $(1 - L_0k_i^*)\,\|A\| < 1$, which, together with $0 < \tilde{m}_i \leq 1 - L_0k_i^*$, leads to $\tilde{m}_i\,\|A\| \in (0, 1)$. Thus, it follows from (B.5) that $\limsup_{t\to\infty} \rho_i(t) \leq \tilde{\alpha}_1$, where $\tilde{\alpha}_1$ is in (21).

Second, for vehicle $i \in \mathcal{V}_2 \bigcap \hat{\mathcal{S}}_i(T_i)$, according to (18) and $\frac{(\varpi-1)\|A\|}{\varpi} \in (0, 1)$, we have $\limsup_{t\to\infty} \lambda_i(t) \leq \tilde{\alpha}_2$, where $\tilde{\alpha}_2$ is in (21).

Third, for vehicle $i \in \mathcal{V}_2 - \hat{\mathcal{S}}_i(T_i)$, since $\hat{\mathcal{S}}_i(t)$ is non-decreasing, we have $|j_i(t) - i| \leq |j_i^* - i|$, where j_i^* is in (22), and $j_i(t) = \arg\min_{j\in\hat{\mathcal{N}}_i\cup\hat{\mathcal{S}}_i(t)} |j-i|, \quad t \geq T_i$. From (19) and $\limsup_{t\to\infty} s_i(t) \leq \max\{\tilde{\alpha}_1, \tilde{\alpha}_2\}$, we obtain $\limsup_{t\to\infty} \tau_i(t) \leq \tilde{\alpha}_3$, where $\tilde{\alpha}_3$ is in (21).

Appendix C. Proof of Proposition 2

Necessity: We assume $b > L$ for the proof by contradiction. Then $2L+1-b \leq b$, which leads to $\frac{2L}{2L+1-b} \geq \frac{2L}{b}$. It is known from $\bar{\beta}_1(\omega) > 0$ that $2L+1 > b$. Given $\omega \in (0, 1)$, due to $\|A\| > 1$, we have $\frac{(\omega+\|A\|-1)\beta_0}{\|A\|} > \left(\omega q - \frac{(\epsilon+\bar{\mu})(2L+1-b)}{2L}\right)$, where $\beta_0 = \|A\|\,q + \epsilon + \bar{\mu}$. Thus, $\bar{\beta}_1(\omega) > \bar{\beta}_2(\omega)$. The assumption $b > L$ does not hold.

Sufficiency: We will prove that if the inequalities in (23) are satisfied, the scalar ω is such that $0 < \bar{\beta}_1(\omega) < \bar{\beta}_2(\omega)$. According to (20) and the first inequality in (23), $\bar{\beta}_1(\omega) < \frac{2L}{b}\left(\omega q - \frac{(\epsilon+\bar{\mu})(2L+1-b)}{2L}\right)$. If the second inequality in (23) holds, then $\frac{2L}{2L+1-b}\frac{(\omega+\|A\|-1)}{\|A\|} < 1$. Multiplying both sides of this inequality by β_0 in (20) leads to $\bar{\beta}_1(\omega) < \beta_0$. Therefore, $\bar{\beta}_1(\omega) < \bar{\beta}_2(\omega)$. Due to $\|A\| > 1$ and $2L+1-b > 0$, we have $\bar{\beta}_1(\omega) > 0$.

Appendix D. Proof of Theorem 2

According to Proposition 1, we prove boundedness of the three sequences $\rho_i(t)$, $\lambda_i(t)$ $\tau_i(t)$ for the case that $\beta_i(t)$ is designed as in (25). Denote $\bar{L} = 2L+1-b$.

First, we consider the case for vehicle $i \in \mathcal{V}_1$. Since $\beta_{i,0}$ satisfies the same condition as β_i in Theorem 1, according to the proof of Theorem 1, we have $k_{i,0} := \frac{\beta_{i,0}}{\|A\|q+\epsilon+\bar{\mu}} < 1$ and

$$(1 - \frac{\bar{L}}{2L}k_{i,0})\,\|A\|\,q + \frac{(\epsilon+\bar{\mu})\bar{L} + b\beta_{i,0}}{2L} < q, \tag{D.1}$$

which corresponds to (B.4). From (D.1) and $\beta_{i,0} = k_{i,0}(\|A\|\, q + \epsilon + \bar{\mu})$, we are able to obtain

$$\left(1 - \frac{\bar{L} - b}{2L} k_{i,0}\right) \|A\| < 1. \tag{D.2}$$

Submitting $\beta_i(t)$ in (25) into (17) yields

$$\rho_i(t) = h_{i,1}(t)\, \|A\|\, \rho_i(t-1) + h_{i,2}(t), \tag{D.3}$$

where

$$h_{i,1}(t) = 1 - \frac{|\hat{\mathcal{S}}_{i,1}(t)| + (\bar{L} - b + |\hat{\mathcal{S}}^a_i(t)| - |\hat{\mathcal{S}}_{i,1}(t)|)k_{i,0}}{2L},$$
$$h_{i,2}(t) = \frac{\bar{L} + (b - |\hat{\mathcal{S}}^a_i(t)|)k_{i,0}}{2L}(\epsilon + \bar{\mu}),$$

By Assumption 3, both $|\hat{\mathcal{S}}^a_i(t)|$ and $|\hat{\mathcal{S}}_i(t)|$ are non-decreasing, thus $|\hat{\mathcal{S}}^a_i(t)| \geq |\hat{\mathcal{S}}^a_i(T_i)|$ and $|\hat{\mathcal{S}}_i(t)| \geq |\hat{\mathcal{S}}_i(T_i)|$, for any $t \geq T_i$. Due to $k_{i,0} < 1$, we have $\sup_{t \geq T_i} h_{i,1}(t) \leq h_{i,1}(T_i) \leq 1 - \frac{\bar{L}-b}{2L}k_{i,0}$ and $\sup_{t \geq T_i} h_{i,2}(t) \leq h_{i,2}(T_i)$, which, together with (D.2)–(D.3), leads to $\limsup_{t\to\infty} \rho_i(t) \leq \frac{h_{i,2}(T_i)}{1 - h_{i,1}(T_i)\|A\|}$.

The proofs for vehicle $i \in \mathcal{V}_2 \bigcap \hat{\mathcal{S}}_i(T_i)$ and for vehicle $i \in \mathcal{V}_2 - \hat{\mathcal{S}}_i(T_i)$ are similar to the proofs in Theorem 1.

Appendix E. Proof of Lemma 1

We use induction to prove the result. At the initial time, Assumption 3 holds trivially. Assume at time $t-1$, Assumption 3 is satisfied. Then, we consider the case at time t. First, we aim to prove the following conclusions corresponding to lines 6, 17, and 21 of Algorithm 2 under the preconditions in lines 5 and 16:

(i) If the detection condition (27) is satisfied, either sensor i or sensor $i-1$ is attacked.

(ii) If the detection condition (28) is satisfied, sensor i is attacked.

(iii) If the detection condition (29) is satisfied, the sensors in the set $\mathcal{V} \setminus (\hat{\mathcal{S}}^s_i(t) \cup \hat{\mathcal{S}}^a_i(t))$ are attack-free.

If (i)–(iii) hold, Algorithms 1–2 ensure that the sets $\hat{\mathcal{S}}^a_i(t)$, $\hat{\mathcal{S}}^s_i(t)$, and $\hat{\mathcal{S}}_i(t)$ are all fault-free. The updates of the three sets in Algorithm 2 ensure that $\hat{\mathcal{S}}_i(t)$ and $\hat{\mathcal{S}}^a_i(t)$ are monotonically non-decreasing. Therefore, Assumption 3 is satisfied at time t. In the following, we prove (i)–(iii).

Proof of (i): By (2), for two attack-free sensors $i-1$ and i, due to $a_i = a_{i-1} = 0$, it holds that $y_{i,i}(t) - y_{i-1,i-1}(t) = x_i(t) - x_{i-1}(t) + n_{i,i}(t) - n_{i-1,i-1}(t)$, which, together with (3), leads to $y_{i-1,i}(t) + y_{i-1,i-1}(t) - y_{i,i}(t) = n_{i-1,i}(t) + n_{i-1,i-1}(t) - n_{i,i}(t)$. Under Assumption 2, taking 2-norm of its both sides yields the conclusion. The conclusion (ii) is satisfied according to Proposition 1 by noting that $i \notin \hat{\mathcal{S}}_i(t)$. Proof of (iii): Since $\bigcup_{j=1}^{j_i} \overline{\hat{\mathcal{S}}^s}_{i,j}(t) = \overline{\hat{\mathcal{S}}^s}_i(t)$ and each set $\overline{\hat{\mathcal{S}}^s}_{i,j}(t)$ contains successive sensor labels, the minimum number of the attacked sensors is no smaller than the sum of the minimum attacked sensor number in each $\overline{\hat{\mathcal{S}}^s}_{i,j}(t)$. One attacked sensor can lead to at most three suspicious sensors comprising of itself and its two neighbor sensors, hence, each $\overline{\hat{\mathcal{S}}^s}_{i,j}(t)$ contains $\lceil |\overline{\hat{\mathcal{S}}^s}_{i,j}(t)|/3 \rceil$ attacked sensors at least. Given the detection condition (29), the conclusion of (iii) is obtained by noting that the set $\overline{\hat{\mathcal{S}}^s}_i(t) = \hat{\mathcal{S}}^s_i(t) \bigcup \hat{\mathcal{S}}^a_i(t)$ contains all attacked sensors.

Appendix F. Proof of Theorem 4

Recall from (4) that $x^*_i(t) = [s^*_i(t), v^*_i(t)]^\mathsf{T}$ is the desired state of vehicle i, $0 \leq i \leq N$, which is subject to $s^*_i(t) = s^*_j(t) + \Delta x^s_{j,i}(t)$ and $v^*_i(t) = v^*_j(t) + \Delta x^v_{j,i}(t)$, $j \in \bar{\mathcal{N}}_i$. Denote $\tilde{e}_i(t) = x_i(t) - x^*_i(t) = [\tilde{s}_i(t), \tilde{v}_i(t)]^\mathsf{T}$ the tracking error of vehicle i. Since the virtual reference vehicle 0 is in its desired state, then $\tilde{s}_0(t) = \tilde{v}_0(t) = 0$. For $1 \leq i \leq N$, it holds that

$$\begin{aligned} \tilde{e}_i(t+1) &= A\tilde{e}_i(t) + [0, T\tilde{u}_i(t)]^\mathsf{T} + \delta_i(t), \\ \delta_i(t) &= [0, T\hat{u}_i(t)]^\mathsf{T} + d_i(t), \end{aligned} \tag{F.1}$$

where

$$\begin{aligned} \tilde{u}_i(t) = &\sum_{j \in \bar{\mathcal{N}}_i} \Big(g_s(\tilde{s}_j(t) - \tilde{s}_i(t)) \\ &+ g_v(\tilde{v}_j(t) - \tilde{v}_i(t)) \Big), 0 \leq i, j \leq N, \\ \hat{u}_i(t) = &\sum_{j \in \bar{\mathcal{N}}_i} \Big(g_s((\bar{s}_j(t) - s_j(t)) - (\hat{s}_i(t) - s_i(t))) \\ &+ g_v((\bar{v}_j(t) - v_j(t)) - (\hat{v}_i(t) - v_i(t))) \Big). \end{aligned} \tag{F.2}$$

From (F.1) and (F.2), we have

$$\tilde{E}(t+1) = P_0 \tilde{E}(t) + \delta(t), \tag{F.3}$$

where P_0 is in (31), $\tilde{E}(t) = [\tilde{e}_1(t)^\mathsf{T}, \ldots, \tilde{e}_N(t)^\mathsf{T}]^\mathsf{T}$, and $\delta(t) = [\delta_1(t)^\mathsf{T}, \ldots, \delta_N(t)^\mathsf{T}]^\mathsf{T}$. By Theorem 1, $\sup_{t \geq 0} \|\delta(t)\| < \infty$. Based on the BIBO stability principle, the asymptotic stability of $\tilde{E}(t)$ in (F.3) is determined by the eigenvalues of P_0. According to Hao, Barooah, and Veerman (2010), the spectrum of P_0 is $\sigma(P_0) = \bigcup_{\sigma_l \in \sigma(\mathcal{L}_g)} \sigma\{A - \sigma_l F_0\} = \bigcup_{\sigma_l \in \sigma(\mathcal{L}_g)} \sigma\{Q_l\}$, where $\mathcal{L}_g$ is introduced before Assumption 4, $\sigma(\cdot)$ is the set of distinct eigenvalues, and $Q_l = \left(\begin{smallmatrix} 1 & T \\ -\sigma_l T g_s & 1 - \sigma_l T g_v \end{smallmatrix}\right)$, $l = 1, 2, \ldots, N$. From Hao et al. (2010), all eigenvalues of $\mathcal{L}_g$ are real-valued and positive, i.e., $\sigma_l > 0$. Denote the eigenvalues of Q_l by s, which are the roots of $\phi(s) = 0$, where $\phi(s) = s^2 + (\sigma_l T g_v - 2)s + \sigma_l T^2 g_s - \sigma_l T g_v + 1$. To prove the Schur stability of P_0, in the following, we aim to prove for each σ_l, $l = 1, 2, \ldots, N$, s falls into the open unit disk, i.e., $|s| < 1$. By applying bilinear transformation to $\phi(s)$, we can transfer the Schur stability of $\phi(s)$ into the Hurwitz stability of a continuous-time system. Then we are able to prove that s falls into the open unit disk, i.e., $|s| < 1$, if and only if $g_v > T g_s > 0$ and $T^2 g_s - 2T g_v > -4/\sigma_l$. We refer to Xie and Wang (2012) for a similar proof. Thus, when (g_s, g_v) are chosen as in Assumption 4, P_0 is Schur stable. From Theorem 1, (F.1), and (F.2), we have $\limsup_{t\to\infty} \|\delta(t)\| \leq \eta$, where η is given in (31). Since P_0 is Schur stable, we use Lemma 2 with respect to (F.3). Due to $\left\|\tilde{e}_i(t)\right\| \leq \left\|\tilde{E}(t)\right\|$, from the definition of the overall function $\varphi(t)$ in (4) and Theorem 1, the conclusion in (1) is obtained. The proof of (ii) is the same as the proof of (ii) but using Theorem 2 in the evaluation of the estimation error instead of using Theorem 1.

References

Baras, J. S., & Liu, X. (2019). Trust is the cure to distributed consensus with adversaries. In *Mediterranean conference on control and automation* (pp. 195–202).

Chen, Y., Kar, S., & Moura, J. M. (2019). Resilient distributed estimation: Sensor attacks. *IEEE Transactions on Automatic Control*, *64*(9), 3772–3779.

Chowdhury, N. R., Belikov, J., Baimel, D., & Levron, Y. (2020). Observer-based detection and identification of sensor attacks in networked CPSs. *Automatica*, *121*, Article 109166.

Deghat, M., Ugrinovskii, V., Shames, I., & Langbort, C. (2019). Detection and mitigation of biasing attacks on distributed estimation networks. *Automatica*, *99*, 369–381.

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

Feng, Z., Wen, G., & Hu, G. (2017). Distributed secure coordinated control for multiagent systems under strategic attacks. *IEEE Transactions on Cybernetics*, *47*(5), 1273–1284.

Forti, N., Battistelli, G., Chisci, L., Li, S., Wang, B., & Sinopoli, B. (2018). Distributed joint attack detection and secure state estimation. *IEEE Transactions on Signal and Information Processing over Networks*, *4*(1), 96–110.

Gallo, A. J., Turan, M. S., Boem, F., Parisini, T., & Ferrari-Trecate, G. (2020). A distributed cyber-attack detection scheme with application to DC microgrids. *IEEE Transactions on Automatic Control*, *65*(9), 3800–3815.

Gao, Y. B., Sun, G. H., Liu, J. X., Shi, Y., & Wu, L. G. (2020). State estimation and self-triggered control of CPSs against joint sensor and actuator attacks. *Automatica*, *113*.

Ge, X., Han, Q.-L., Zhong, M., & Zhang, X.-M. (2019). Distributed Krein space-based attack detection over sensor networks under deception attacks. *Automatica*, *109*, Article 108557.

Hao, H., Barooah, P., & Veerman, J. (2010). Effect of network structure on the stability margin of large vehicle formation with distributed control. In *IEEE conference on decision and control* (pp. 4783–4788).

He, X., Hashemi, E., & Johansson, K. H. (2020). Secure platooning of autonomous vehicles under attacked GPS data. arXiv preprint arXiv:2003.12975.

He, X., Ren, X., Sandberg, H., & Johansson, K. H. (2021). How to secure distributed filters under sensor attacks. *IEEE Transactions on Automatic Control*.

Kim, J., Lee, C., Shim, H., Eun, Y., & Seo, J. H. (2018). Detection of sensor attack and resilient state estimation for uniformly observable nonlinear systems having redundant sensors. *IEEE Transactions on Automatic Control*, *64*(3), 1162–1169.

Lee, J. G., Kim, J., & Shim, H. (2020). Fully distributed resilient state estimation based on distributed median solver. *IEEE Transactions on Automatic Control*, *65*(9), 3935–3942.

Lin, P., & Jia, Y. (2009). Consensus of second-order discrete-time multi-agent systems with nonuniform time-delays and dynamically changing topologies. *Automatica*, *45*(9), 2154–2158.

Lu, A. Y., & Yang, G. H. (2019). Secure switched observers for cyber-physical systems under sparse sensor attacks: A set cover approach. *IEEE Transactions on Automatic Control*, *64*(9), 3949–3955.

Mitra, A., Richards, J. A., Bagchi, S., & Sundaram, S. (2019). Resilient distributed state estimation with mobile agents: overcoming Byzantine adversaries, communication losses, and intermittent measurements. *Autonomous Robots*, *43*(3), 743–768.

Mitra, A., & Sundaram, S. (2019). Byzantine-resilient distributed observers for LTI systems. *Automatica*, *108*, Article 108487.

Pajic, M., Lee, I., & Pappas, G. J. (2017). Attack-resilient state estimation for noisy dynamical systems. *IEEE Transactions on Control of Network Systems*, *4*(1), 82–92.

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

Ren, X., Mo, Y., Chen, J., & Johansson, K. H. (2020). Secure state estimation with Byzantine sensors: A probabilistic approach. *IEEE Transactions on Automatic Control*, *65*(9), 3742–3757.

Shinohara, T., Namerikawa, T., & Qu, Z. H. (2019). Resilient reinforcement in secure state estimation against sensor attacks with a priori information. *IEEE Transactions on Automatic Control*, *64*(12), 5024–5038.

Shoukry, Y., Chong, M., Wakaiki, M., Nuzzo, P., Sangiovanni-Vincentelli, A., Seshia, S. A., et al. (2018). SMT-based observer design for cyber-physical systems under sensor attacks. *ACM Transactions on Cyber-Physical Systems*, *2*(1), 1–27.

Shoukry, Y., Nuzzo, P., Puggelli, A., Sangiovanni-Vincentelli, A. L., Seshia, S. A., & Tabuada, P. (2017). Secure state estimation for cyber-physical systems under sensor attacks: A satisfiability modulo theory approach. *IEEE Transactions on Automatic Control*, *62*(10), 4917–4932.

Shoukry, Y., & Tabuada, P. (2016). Event-triggered state observers for sparse sensor noise/attacks. *IEEE Transactions on Automatic Control*, *61*(8), 2079–2091.

Su, L., & Shahrampour, S. (2020). Finite-time guarantees for Byzantine-resilient distributed state estimation with noisy measurements. *IEEE Transactions on Automatic Control*, *65*(9), 3758–3771.

Tang, Z. H., Kuijper, M., Chong, M. S., Mareels, I., & Leckie, C. (2019). Linear system security-detection and correction of adversarial sensor attacks in the noise-free case. *Automatica*, *101*, 53–59.

Teixeira, A., Shames, I., Sandberg, H., & Johansson, K. H. (2015). A secure control framework for resource-limited adversaries. *Automatica*, *51*, 135–148.

Weerakkody, S., Liu, X., Son, S. H., & Sinopoli, B. (2016). A graph-theoretic characterization of perfect attackability for secure design of distributed control systems. *IEEE Transactions on Control of Network Systems*, *4*(1), 60–70.

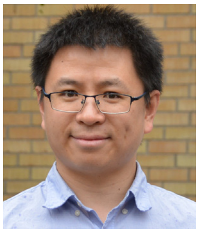
Xie, D., & Wang, S. (2012). Consensus of second-order discrete-time multi-agent systems with fixed topology. *Journal of Mathematical Analysis and Applications*, *387*(1), 8–16.

Yang, T., Murguia, C., Kuijper, M., & Nešić, D. (2020). A multi-observer based estimation framework for nonlinear systems under sensor attacks. *Automatica*, *119*, Article 109043.

Zhao, D., Wang, Z. D., Wei, G. L., & Han, Q. L. (2020). A dynamic event-triggered approach to observer-based PID security control subject to deception attacks. *Automatica*, *120*.

Zhu, M., & Martínez, S. (2013). On distributed constrained formation control in operator–vehicle adversarial networks. *Automatica*, *49*(12), 3571–3582.

Zhu, Y. Z., & Zheng, W. X. (2020). Observer-based control for cyber-physical systems with periodic DoS attacks via a cyclic switching strategy. *IEEE Transactions on Automatic Control*, *65*(8), 3714–3721.



Xingkang He is a postdoctoral researcher at the Division of Decision and Control Systems, KTH Royal Institute of Technology, Stockholm, Sweden. He received the B.S. degree in School of Mathematics from Hefei University of Technology in 2013, and the Ph.D. degree in Academy of Mathematics and Systems Science, Chinese Academy of Sciences at Beijing in 2018. His research interests include security of cyber–physical systems, estimation and control of networked systems, filtering theory.

He is a recipient of Best Paper Award in 2018 IEEE Data Driven Control and Learning Systems Conference.

Ehsan Hashemi received his Ph.D. in Mechanical and Mechatronics Engineering in 2017 from University of Waterloo, ON, Canada; M.Sc. in Mechanical Engineering in 2005 from Amirkabir University of Technology (Tehran Polytechnic). He is currently an Assistant Professor at the Department of Mechanical Engineering, University of Alberta. His research interests are robotics, control theory, human–machine interaction, and fault-tolerance.

Karl H. Johansson is Professor with the School of Electrical Engineering and Computer Science at KTH Royal Institute of Technology in Sweden and Director of Digital Futures. He received M.Sc. and Ph.D. degrees from Lund University. He has held visiting positions at UC Berkeley, Caltech, NTU, HKUST Institute of Advanced Studies, and NTNU. His research interests are in networked control systems and cyber–physical systems with applications in transportation, energy, and automation networks. He is a member of the Swedish Research Council's Scientific Council for Natural Sciences and Engineering Sciences. He has served on the IEEE Control Systems Society Board of Governors, the IFAC Executive Board, and is currently Vice-President of the European Control Association. He has received several best paper awards and other distinctions from IEEE, IFAC, and ACM. He has been awarded Distinguished Professor with the Swedish Research Council and Wallenberg Scholar with the Knut and Alice Wallenberg Foundation. He has received the Future Research Leader Award from the Swedish Foundation for Strategic Research and the triennial Young Author Prize from IFAC. He is Fellow of the IEEE and the Royal Swedish Academy of Engineering Sciences, and he is IEEE Control Systems Society Distinguished Lecturer.