2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 25-29, 2020, Las Vegas, NV, USA (Virtual)

Output-Only Fault Detection and Mitigation of Networks of Autonomous Vehicles

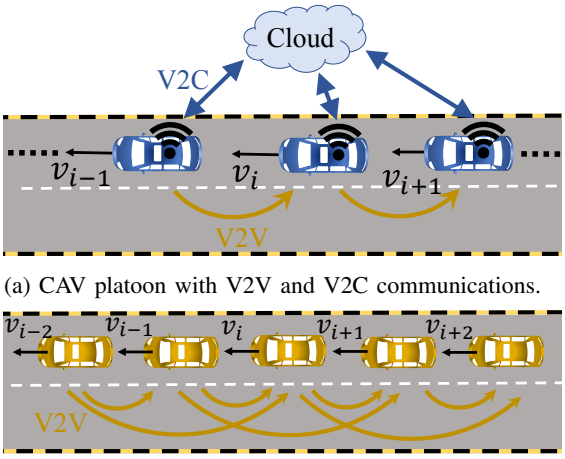
Abdelrahman Khalil, Mohammad Al Janaideh, Khaled F. Aljanaideh, and Deepa Kundur

***Abstract*— An autonomous vehicle platoon is a network of autonomous vehicles that communicate together to move in a desired way. One of the greatest threats to the operation of an autonomous vehicle platoon is the failure of either a physical component of a vehicle or a communication link between two vehicles. This failure affects the safety and stability of the autonomous vehicle platoon. Transmissibility-based health monitoring uses available sensor measurements for fault detection under unknown excitation and unknown dynamics of the network. After a fault is detected, a sliding mode controller is used to mitigate the fault. Different fault scenarios are considered including vehicle internal disturbances, cyber attacks, and communication delays. We apply the proposed approach to a bond graph model of the platoon and an experimental setup consisting of three autonomous robots.**

I. Introduction

Connected autonomous vehicles (CAV) platoons represent a new technology where a network of vehicles communicate together using wireless communication to achieve a desired speed and position of the vehicles in the network. This new technology represents an emerging cyber-physical system (networking, computation, and physical processes) with significant potential to enhance traffic safety, ease congestion, and positively impact the environment through autonomous platoon control, see for example [1]. The cyber component of such a system incorporates the vehicle-to-vehicle (V2V) and vehicle-to-cloud (V2C) communication networks [2], while the physical component includes physical vehicle dynamics and human-driver responses.

It is evident that as the technology and complexity of connected autonomous vehicles evolve, several grand research challenges need to be addressed. These include securing the connected autonomous vehicles from malicious cyber attacks that can affect the actuators and sensors in the CAV platoon, see for example [3]. Other sources of failures include cyber-physical attacks, faults in sensors and actuators, and unknown nonlinear dynamics in the CAV [4]. Several studies have considered detecting and mitigating techniques for a class of faults and cyber attacks to enhance traffic safety of CAV. For example, in [5], real-time observers designed using sliding mode and adaptive estimation theory were used to detect what is called a denial-of-service cyber attack. In [6], distance and velocity controllers were used to avoid collisions and to guarantee string stability under communication delay. In [7], an algorithm that is based on a distributed function calculation was used to detect faults in the platoon's V2V communications between the two preceding and two following vehicles.



(a) CAV platoon with V2V and V2C communications.

(b) CAV platoon with V2V communications.

Fig. 1: CAV platoon with different typologies: (a) each vehicle follows the preceding vehicle through V2V communication, and (b) each vehicle follows the average of the velocities of the two preceding vehicles.

Transmissibility operators, which are mathematical models that characterise relationships between sensors of the same system, were used to detect and localize different faults for networks of autonomous-vehicle platoons [8]. In this paper, we use transmissibility operators to detect faults in the connected autonomous-vehicles platoon. Then we design a sliding mode controller to mitigate these faults. Most fault detection techniques require knowledge of a model of the system and the excitation signal that acts on it, see for example [9]. However, transmissibility-based fault detection uses output-only measurements and does not require knowledge of a model of the system or the excitation signal that acts on it [10]. For further validation, we simulate different possible faults that are introduced in the literature such as physical faults in the vehicles as in [11], cyber-physical attacks [3], and time delay in autonomous vehicles networks [12]. Since transmissibilities are constructed from output-only measurements, they can be noncausal, unstable, and of unknown order [13]. Therefore, to identify transmissibilities we use noncausal FIR models, which can approximate systems with

A. Khalil and M. Al Janaideh are with the Department of Mechanical Engineering, Memorial University, St. John's, NL A1B 3X5 Canada, email: amkhalil@mun.ca, maljanaideh@mun.ca.

K. F. Aljanaideh is with the Department of Aeronautical Engineering, Jordan University of Science and Technology, Irbid, Jordan, email: kfaljanaideh@just.edu.jo.

D. Kundur is with the Edward S.Rogers Sr. Department of Electrical Computer Engineering, University of Toronto, Toronto, ON M5S 3G4, Canada, email: dkundur@ece.utoronto.ca.

these properties accurately [14].

The paper is organized as follows: In Section II, we use the bond graph approach to model CAV platoons. In Section III, we use least squares with noncausal FIR models to identify transmissibility operators. In Section IV, we model possible faults in autonomous vehicles platoons. Section V shows fault mitigation using a sliding mode controller. Section VI shows simulation results. Section VII shows experimental setup and experimental results. Conclusions are shown in Section VIII.

II. CAV Modeling Using Bond Graphs

Bond graphs simulate the power transfer between the system components, which conveys two physical quantities, namely, effort (e.g. force or electrical voltage) and flow (e.g. linear velocity or electrical current) [15]. The bond graph model of the CAV platoon is used to obtain measurements at several locations of the platoon. These measurements are used to identify transmissibilities, which in turn are used to detect and mitigate faults in the platoon as will be shown later.

A. CAV Modeling using Bond Graphs

Consider a platoon of n identical connected autonomous vehicles with linear motion as in Figure 1. For all $i = 1, \dots, n$, let v_i denote the velocity of vehicle i and v_i^* denote the desired velocity of vehicle i. We consider the architecture of both V2V and V2C communications in the platoon as shown in Figure 1a where for all $i = 2, \dots, n$, $v_i^* = v_{i-1}$. However, if the platoon is out of the cloud range, we consider the architecture of V2V communications only as in Figure 1b, where for all $i = 3, \dots, n$, $v_i^* = \frac{1}{2}(v_{i-1} + v_{i-2})$. Figure 2 shows a schematic diagram of the connected autonomous vehicles. Figure 3 shows the bond graph representation of the schematic diagram of the ith vehicle shown in Figure 2, where parameters description and their considered numerical values are shown in Table I.

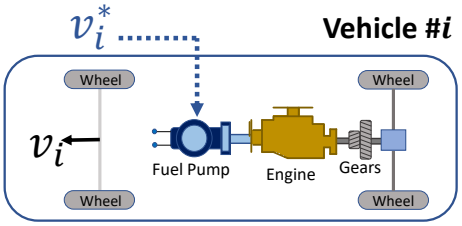


Fig. 2: Schematic representation of an autonomous vehicle, where the vehicle can be modeled as shown in Figure 3.

Consider the bond graph model of vehicle i shown in Figure 3. For all $i = 2, \dots, n$, the relationship between the velocity of vehicle i and its desired velocity v_i^* is given by

$$\ddot{v}_i(t) + \beta_i \dot{v}_i(t) + \gamma_i v_i(t) = \alpha_i \dot{v}_i^*(t) \tag{1}$$

where $\alpha_i = \frac{C_i P_i \rho_i}{Z_i G_i \lambda_i M_i}$, $\beta_i = \frac{\rho_i^2 P_i^2}{M_i \lambda_i^2 G_i^2 Z_i} + F_i$, $\gamma_i = \frac{R_i}{S_i}$, and the rest of the parameters are defined in Table I. Then, for

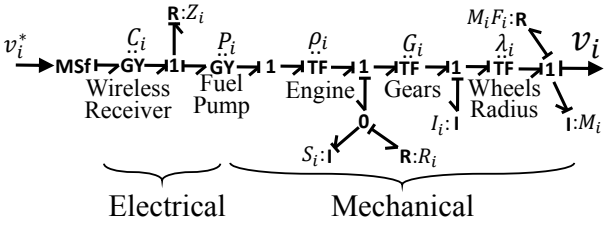


Fig. 3: Vehicle physical model using the bond graph representation, where the components of the vehicle are shown in Figure 2.

all $i = 1, \dots, n$, the state space representation for vehicle i is given by

$$\dot{x}_i(t) = A_i x_i(t) + B_i v_i^*(t), \tag{2}$$

where

$$A_i = \begin{bmatrix} -\beta_i & -\gamma_i \\ 1 & 0 \end{bmatrix} \in \mathbb{R}^{2\times 2}, B_i = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \in \mathbb{R}^{2\times 1}.$$

Next, consider the case with both V2V and V2C communications for all $t \geq 0$, let $u(t) = \dot{v}^*(t)$, $y(t) = [v_1(t) \dots v_n(t)]^{\mathrm{T}}$. Then the state space representation for a platoon of n vehicles can be written as

$$\dot{x}(t) = Ax(t) + Bu(t), \tag{3}$$
$$y(t) = Cx(t), \tag{4}$$

where

$$A = \begin{bmatrix} A_1 & & \dots & & 0 \\ \sigma_2 & A_2 & & & \\ \vdots & \sigma_3 & A_3 & & \vdots \\ & & \ddots & \ddots & \\ 0 & & \dots & \sigma_n & A_n \end{bmatrix} \in \mathbb{R}^{2n\times 2n},$$

$$B = \begin{bmatrix} \alpha_1 & 0 & \dots & 0 \end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^{2n\times 1},$$

$$C = \begin{bmatrix} o_1 & & \dots & 0 \\ & o_2 & & \\ \vdots & & \ddots & \\ 0 & \dots & & o_n \end{bmatrix} \in \mathbb{R}^{n\times 2n},$$

where

$$\sigma_i = \begin{bmatrix} \alpha_i & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{2\times 2}, o_i = \begin{bmatrix} 0 & 1 \end{bmatrix} \in \mathbb{R}^{1\times 2}.$$

It is important to mention that the structure of the matrices A, B, and C in (3) and (4) is valid if each vehicle follows its preceding vehicle. Note that, a different communication topology between the vehicles yields a different structure of A, B, and C in (3) and (4) [16].

B. V2V Communication

V2V wireless information communications consists of short-to-medium range wireless communications [17]. Wireless access communication in vehicular environments covers

up to 1 km in range with a rate of data transmission that is up to 27 Mbps, 5.9 GHz frequency, and a 75 MHz channel bandwidth.

C. V2C Communications

For V2C communications, CAV platoons use different wireless networks than the one used in V2V communications. A combination of large-area street network and a cellular Long Term Evolution (LTE) was used in [18]. This communication network consists of vehicle's user equipment (UE) and Base Station (BS). The UE always connects with the closest BS by calculating the signal-to-noise ratio. As in [19], vehicles equipped with broadband wireless access technologies such as LTE can communicate with each other via the Internet. These vehicles can also conduct integration of cloud computing with CAV, which gives vehicles the ability to run virtual computers with strong computation and a large amount of data storage. According to [20], LTE communication covers up to 2 km in range between the vehicle and the roadside infrastructure antenna with a rate of data transmission that is up to 75 Mbps and a 2.6 GHz frequency with a 20 MHz channel bandwidth.

III. Transmissibility Identification of CAV

A. CAV Transmissibility Operators

Consider the state space model (3)-(4), let $v_{\rm a}$ and $v_{\rm b}$ denote two independent sets of velocity outputs obtained from a platoon with n vehicles. Then, define $v_{\rm a}(t) \triangleq C_{\rm a}x(t) \in \mathbb{R}^m$ and $v_{\rm b}(t) \triangleq C_{\rm b}x(t) \in \mathbb{R}^{n-m}$, where m is the number of independent pseudo inputs, $n \geq 2$ is the number of vehicles in the platoon, and $2n$ is the system order, $C_{\rm a} \in \mathbb{R}^{m\times 2n}$, and $C_{\rm b} \in \mathbb{R}^{(n-m)\times 2n}$. Then, the transmissibility $\mathcal{T}$ whose pseudo input is $v_{\rm a}$ and whose pseudo output is $v_{\rm b}$, satisfies [10]

$$v_{\rm b}(t) = \mathcal{T}(\mathbf{p})v_{\rm a}(t), \tag{5}$$

where $\mathcal{T}(\mathbf{p}) \triangleq \Gamma_{\rm b}(\mathbf{p})\Gamma_{\rm a}^{-1}(\mathbf{p})$,

$$\Gamma_{\rm a}(\mathbf{p}) \triangleq C_{\rm a}{\rm adj}(\mathbf{p}I - A)B \in \mathbb{R}^{m\times m}[\mathbf{p}], \tag{6}$$

$$\Gamma_{\rm b}(\mathbf{p}) \triangleq C_{\rm b}{\rm adj}(\mathbf{p}I - A)B \in \mathbb{R}^{(n-m)\times m}[\mathbf{p}], \tag{7}$$

and $\operatorname{adj}\Gamma_{\rm a}$ denotes the adjugate matrix of $\Gamma_{\rm a}$. Since sensor measurements are obtained in discrete time, we consider discrete-time transmissibility operators in the forward-shift operator $\mathbf{q}$, that is, we replace $\mathbf{p}$ by the forward shift operator $\mathbf{q}$ [21].

B. Identification of transmissibilities

Replacing $\mathbf{p}$ in (5) with $\mathbf{q}$ yields, for all $k \geq 0$,

$$v_{\rm b}(k) = \mathcal{T}(\mathbf{q})v_{\rm a}(k), \tag{8}$$

where

$$\mathcal{T}(\mathbf{q}) = \Gamma_{\rm b}(\mathbf{q})\Gamma_{\rm a}^{-1}(\mathbf{q}) \tag{9}$$

$$= \frac{1}{\det\Gamma_{\rm a}(\mathbf{q})}\Gamma_{\rm b}(\mathbf{q})\operatorname{adj}\Gamma_{\rm a}(\mathbf{q}). \tag{10}$$

Note that if $\Gamma_{\rm a}$ has a nonminimum phase (unstable) zero, then $\mathcal{T}$ will be unstable. Also, if $\Gamma_{\rm b}$ has more zeros than $\Gamma_{\rm a}$, then $\mathcal{T}$ will be noncausal. Moreover, transmissibilities are usually identified using the output measurements with no information about the dynamics of the system, and thus, the order of the transmissibility is unknown. Therefore, to identify transmissibilities, we need to consider a model structure that can approximate noncausal and unstable transmissibilities with unknown order. In this paper, we consider noncausal FIR models, which are truncations of the Laurent expansion in an analytic annulus that contains the unit circle [14]. A noncausal FIR model of $\mathcal{T}$ is given by

$$\mathcal{T}(\mathbf{q}, \Theta_{r,d}^{\rm FIR}) = \sum_{i=-d}^{r} H_i\mathbf{q}^{-i}, \tag{11}$$

where r, d denote the order of the causal and noncausal parts of the FIR model of $\mathcal{T}$, respectively, $H_i \in \mathbb{R}^{(n-m)\times m}$ is the ith coefficient of the Laurent expansion of $\mathcal{T}$ in the annulus that contains the unit circle, and $\Theta_{r,d}^{\rm FIR} \triangleq [H_{-d}, \ldots, H_r\,]^T$. Then, the least squares estimate $\hat{\Theta}_{r,d,\ell}^{\rm FIR}$ of $\Theta_{r,d}^{\rm FIR}$ is given by

$$\hat{\Theta}_{r,d,\ell}^{\rm FIR} = (\Phi_{r,d,\ell}\Phi_{r,d,\ell}^{\rm T})^{-1}\Phi_{r,d,\ell}\Psi_{v_{\rm b},\ell}, \tag{12}$$

where ℓ is the number of samples, $\Psi_{v_{\rm b},\ell} \triangleq \begin{bmatrix} v_{\rm b}(r) & \cdots & v_{\rm b}(\ell-d) \end{bmatrix}^{\rm T}$, $\Phi_{r,d,\ell} \triangleq \begin{bmatrix} \phi_{r,d}(r) & \cdots & \phi_{r,d}(\ell-d) \end{bmatrix}$, and $\phi_{r,d}(k) \triangleq \begin{bmatrix} v_{\rm a}(k+d) & \cdots & v_{\rm a}(k-r) \end{bmatrix}^{\rm T}$. The residual of the identified transmissibility obtained using least squares with a noncausal FIR model at time k is given by

$$e_{\rm b}(k|\hat{\Theta}_{r,d,\ell}^{\rm FIR}) = v_{\rm b}(k) - \hat{v}_{\rm b}(k|\hat{\Theta}_{r,d,\ell}^{\rm FIR}), \tag{13}$$

where

$$\hat{v}_{\rm b}(k|\hat{\Theta}_{r,d,\ell}^{\rm FIR}) = \mathcal{T}(\mathbf{q}, \hat{\Theta}_{r,d,\ell}^{\rm FIR})v_{\rm a}(k) = \sum_{i=-d}^{r} \hat{H}_{i,\ell}v_{\rm a}(k-i), \tag{14}$$

$$\mathcal{T}(\mathbf{q}, \hat{\Theta}_{r,d,\ell}^{\rm FIR}) = \sum_{i=-d}^{r} \hat{H}_{i,\ell}\mathbf{q}^{-i}, \tag{15}$$

and $\hat{\Theta}_{r,d,\ell}^{\rm FIR} = [\hat{H}_{-d,\ell}, \ldots, \hat{H}_{r,\ell}\,]^{\rm T}$.

C. Fault Detection

We use least squares to estimate $\hat{\Theta}_{r,d,\ell}^{\rm FIR}$ from CAV velocities $v_{\rm a}$ and $v_{\rm b}$ under healthy conditions. Next, we use the identified transmissibility operator $\mathcal{T}(\mathbf{q}, \hat{\Theta}_{r,d,\ell}^{\rm FIR})$ with the measurement of v_a to obtain the predicted velocity $\hat{v}_{\rm b}(k|\hat{\Theta}_{r,d,\ell}^{\rm FIR})$. If a fault occurs in the platoon, this leads to a change in dynamics of the platoon, which leads to a change in the level of the residuals of the identified transmissibility $e_{\rm b}(k|\hat{\Theta}_{r,d,\ell}^{\rm FIR})$. Based on the change in the residual $e_{\rm b}(k|\hat{\Theta}_{r,d,\ell}^{\rm FIR})$, we can determine whether the system is healthy or faulty. Next, for all $k \geq 0$, we compute

$$E_{\rm b}(k|\hat{\Theta}_{r,d,\ell}^{\rm FIR}, w) \triangleq \sqrt{\sum_{i=k}^{w+k} \|e_{\rm b}(k|\hat{\Theta}_{r,d,\ell}^{\rm FIR})\|}, \tag{16}$$

which is the norm of the residuals over a sliding window of size w steps. Assume that the system operates in a healthy manner for the first L steps, where $L \geq w + d$, and let η be the signal-to- noise ratio, then define the threshold [22]

$$\mu(\hat{\Theta}^{\text{FIR}}_{r,d,\ell}, w, L) \triangleq \frac{\eta}{L+1} \sum_{i=d}^{L} E_{\text{b}}(i|\hat{\Theta}^{\text{FIR}}_{r,d,\ell}, w). \tag{17}$$

IV. Fault Models

In this section, we show common fault models that affect the performance and the security of the CAV platoon system. These faults include disturbances due to physical faults (for example engine bearing knock), cyber-physical attacks, and time delays in communication links. To model these faults in the CAV platoon, Figure 4 shows the modified bond graph model of the CAV including the possible faults.

A. Engine Bearing Knock

Reciprocating engines are typical appliances of slider-crank mechanisms with clearance joint and lubrication. These engines have severe operating conditions including high pressures and temperatures. However, these engines are highly vulnerable to bearing damage, which leads to an incomplete combustion cycle and thus can cause the bearing to knock [11]. For all $i = 1, \ldots, n$, let ρ_i denote the engine constant, which is defined as the nominal value of the ratio of the engine fuel volumetric flow rate to the engine output angular velocity. Then the engine bearing knock fault can be modeled as a deviation from the nominal value of ρ_i, that is, for all $i = 1, \ldots, n$, and for all $t \geq 0$,

$$\tilde{\rho}_i(t) = \rho_i + \delta_{\rho,i}(t), \tag{18}$$

where $\tilde{\rho}_i$ denotes the corrupted value of the engine bearing knock constant, and $\delta_{\rho,i}$ denotes the deviation from the nominal value of the engine bearing knock constant.

B. Spacing Distance

Connected autonomous vehicles platoons have several spacing-distance policies. One possible cyber attack that can affect the system performance is adding disturbances to the spacing distance between two preceding vehicles. This fault can lead to instabilities, inaccuracies, and oscillations in the system performance [3]. For all $i = 1, \ldots, n$, let h_i denote the nominal spacing value between vehicle i and vehicle $i+1$, then for all $t \geq 0$,

$$\tilde{h}_i(t) = h_i(t) + \delta_{f,i}(t), \tag{19}$$

where $\tilde{h}_i$ denotes the corrupted spacing distance, and $\delta_{f,i}$ denotes the deviation from h_i due to a cyber attack. Spacing distance fault can occur due to corrupted measurements of the velocities of the vehicles. Therefore, can be represented by

$$\tilde{v}_i(t) = v_i(t) + \dot{\tilde{h}}_i(t), \tag{20}$$

where for all $i = 1, \ldots, n$, $\tilde{v}_i$ represents the corrupted measurements of the velocity of vehicle i, $\dot{\tilde{h}}_i$ denotes the deviation from v_i due to a cyber attack.

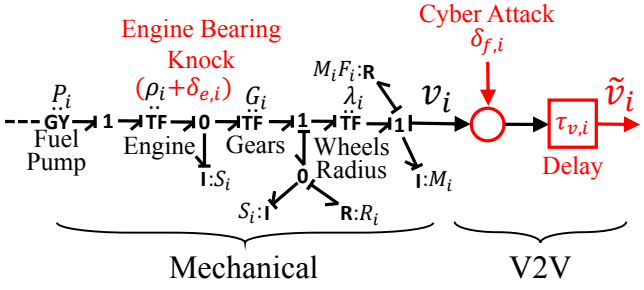


Fig. 4: Modified bond graph model of the CAV platoon with the proposed faults.

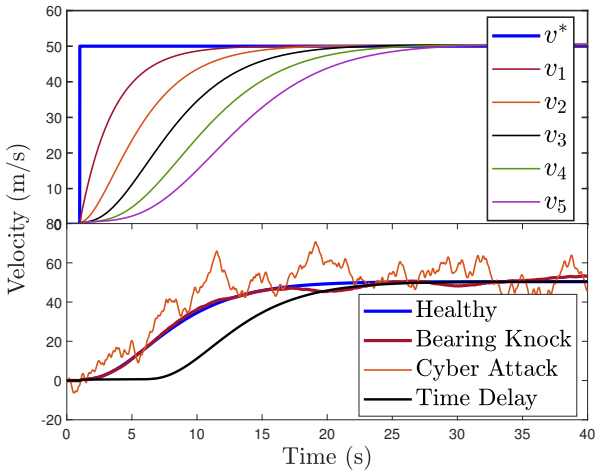


Fig. 5: Example 4.1: The first part shows velocities of five healthy vehicles in a platoon under a step input, where $v^* = 50$ m/s is the desired velocity of the platoon, and the second part shows response of the third vehicle under healthy and faulty conditions, where the faults are introduced individually.

C. Delay

Time delays in connected autonomous vehicles platoons can yield fatal faults as considered in the literature, see for example [12]. Time delays can occur due to a cyber attack or system jamming. For all $i = 1, \ldots, n$, consider the velocity of the ith vehicle v_i, then a delay in v_i yields the corrupted signal

$$\tilde{v}_i(t) = v_i(t - \tau_{v,i}(t)), \tag{21}$$

where $\tau_{v,i}$ is the time-variant communication delay in v_i.

Example 4.1: In this example, we consider a platoon of $n = 5$ vehicles. We consider the bond graph model shown in Figure 3 with the parameters values shown in Table I. The first part of Figure 5 shows the velocities of five healthy vehicles in a platoon under a step input. Next, we add a band-limited white noise to the engine constant of the third vehicle, that is, we set $\delta_{e,3}$ in (18) to be white noise, which

results in an engine bearing knock fault. The second part of Figure 5 compares the healthy and faulty response of the third vehicle due to a step input. Similarly, we consider the cases of introducing white noise to the communication link between the second and third vehicles to emulate a cyber attack. Moreover, we add a time delay of $\tau_{v,3} = 6$ seconds to the response of the third vehicle. The second part of Figure 5 compares the healthy and faulty response of the third vehicle due to a step input under a cyber attack and time-delay faults.

Symbol	Description	Value
C_i	Wireless receiver gain	$2.52V/(m/s)$
Z_i	Pump electrical resistance	0.001Ω
P_i	Pump voltage to fuel ratio	$0.02V/(mg/s)$
ρ_i	Engine fuel to speed ratio	$8mg/rev.$
S_i	Engine shaft moment of inertia	$0.2kg.m^2$
R_i	Engine shaft damping	$\simeq 0$
G_i	Gears Ratio	0.2
I_i	Wheel spin inertia	$\simeq 0$
λ_i	Wheel Radius	$0.3m$
M_i	Vehicle gross mass	$1478kg$
F_i	Friction coefficient	0.6
$\mathcal{M}_i$	Sliding mode constant	70

TABLE I: Simulation parameters for the bond graph model depicted in Figure 3.

V. Sliding Mode Controller

Sliding mode control was used effectively for fault mitigation in CAV networks [23]. In this section, we use a sliding mode controller to mitigate the faults that are detected using the proposed fault detection algorithm.

As shown in Figure 6, for all $i = 2, \ldots, n$, a fault in vehicle i results in producing a corrupted signal $\tilde{v}_i$. Transmissibilities identified under healthy conditions can be used along with measurements from healthy sensors to obtain a prediction $\hat{v}_i$ of v_i, which can then activate the sliding mode controller to produce the control signal ϑ_i. For all $k \geq 0$, the discrete form of the vehicle model in (2) can be written as

$$x_i(k+1) = \mathcal{A}_i x_i(k) + \mathcal{B}_i v_i^*(k), \tag{22}$$

where $\mathcal{A}_i = \exp(A_i T_s), \mathcal{B}_i = A_i^{-1}(\mathcal{A}_i - \mathbf{I})B_i$, T_s is the sampling time and $\mathbf{I} \in \mathbb{R}^{2\times 2}$ is the identity matrix. Considering the sliding mode controller [23] for vehicle i, where $i = 2, \ldots, n$, (2) becomes

$$x_i(k+1) = \mathcal{A}_i x_i(k) + \mathcal{B}_i(v_i^*(k) - \vartheta_i(k|\hat{\Theta}^{\text{FIR}}_{r,d,\ell})), \tag{23}$$

where ϑ_i is the control signal produced by the sliding mode controller. For all vehicles $i = 2, \ldots, n$, let e_i denote the transmissibility residual as defined in (13) where the transmissibility output is v_i, then the control signal is designed to force the states slide along $e_i = 0$ under the presence of internal disturbances or communication fault such as cyber attacks and communication delay. Following [23], we define the control signal as

$$\vartheta_i(k|\hat{\Theta}^{\text{FIR}}_{r,d,\ell}) = \begin{cases} 0, & E_i(k) < \mu_i, \\ \mathcal{P}_i e_i(k|\hat{\Theta}^{\text{FIR}}_{r,d,\ell}) & \\ \quad - \mathcal{M}_i \tanh(e_i(k|\hat{\Theta}^{\text{FIR}}_{r,d,\ell})), & E_i(k) \geq \mu_i, \end{cases} \tag{24}$$

where E_i is the norm of residual over a sliding window as defined in (16) where the transmissibility output is v_i, $\mathcal{M}_i$ and $\mathcal{P}_i$ are positive constants that indicate how the states will slide back to $e_i = 0$, and μ_i represents the average of (17) for all k where vehicle i is healthy. Stability analysis can be shown by following the same stability analysis steps shown in [23].

VI. Simulation Results

Consider a platoon with five identical vehicles with the parameters shown in Table I. We set the desired velocity of the platoon to Gaussian white noise with zero mean and unit variance. For $i = 1, \ldots, 5$ and $j = 1, \ldots, 5$,

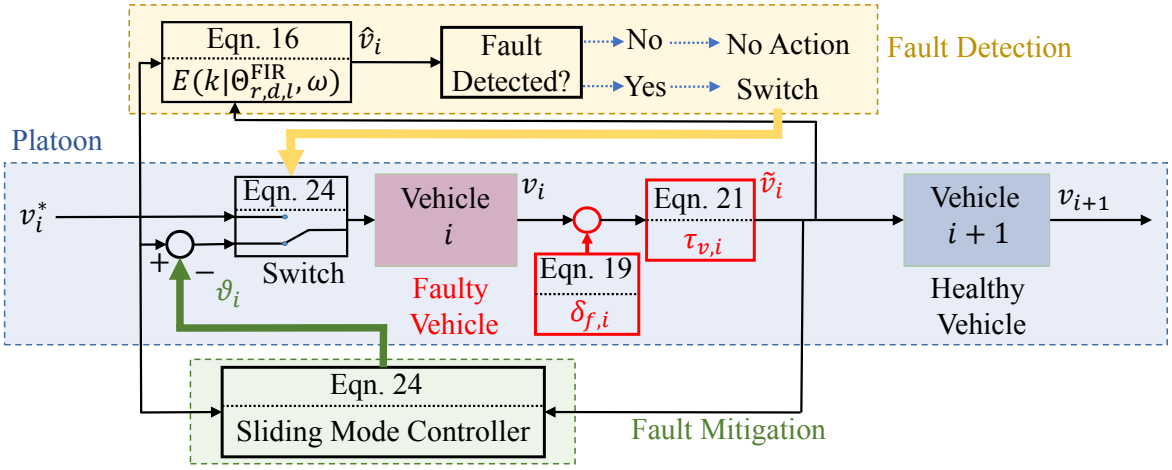


Fig. 6: Block diagram for the proposed fault mitigation algorithm based on transmissibilities. A fault in vehicle i results in producing the corrupted signal $\tilde{v}_i$. Transmissibilities identified under healthy conditions can be used along with measurements from healthy sensors to obtain a prediction $\hat{v}_i$ of v_i, which can then activate the sliding mode controller to produce the control signal ϑ_i.

where $j \neq i$, let $\mathcal{T}_{ij}$ denote the transmissibility operator from vehicle i to vehicle j. Note that using a noncausal FIR model of the transmissibility results in a delay in the predicted output of the transmissibility. Therefore, to avoid delays in the closed-loop system due to using a noncausal model of the transmissibility, we use causal FIR models to identify the transmissibilities that are used for fault mitigation. Fault mitigation can be achieved if the distance between the measured output obtained under healthy conditions and the predicted output obtained using a causal model of the transmissibility is negligible. Figure 7 shows the estimated Markov parameters of the operator $\mathcal{T}_{23}$ identified with $r = 50$ and $d = 0$. Then, the estimated transmissibility is used along with the measurements of v_2 to obtain an estimate of the velocity v_3 as shown in the first part of Figure 8. At $t = 100$ seconds, the cyber attack was introduced, which leads to the prediction error shown in the second part of Figure 8. As shown in Figure 9, at $t = 102$ seconds the error norm exceeds the threshold limit and thus the sliding mode controller is activated, which leads to a drop in the error norm below the threshold limit.

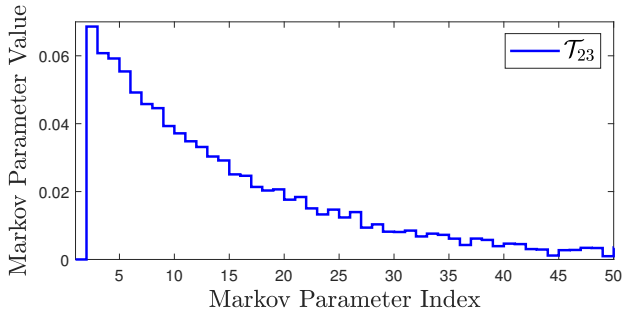


Fig. 7: Simulation results: Estimated Markov parameters for the transmissibility operator $\mathcal{T}_{23}$ obtained using least squares with a causal FIR model with $r = 50$ and $d = 0$.

Next, we introduce the engine bearing knock, cyber attack, and V2V time-delay faults to the system separately. To emulate the engine bearing knock fault, a band-limited white noise is added to the engine's constant of the third vehicle's engine. To emulate a cyber attack fault, a band-limited white noise is added to the communication link between the third and the fourth vehicles. Moreover, to emulate time delay, a time delay of 2 seconds is introduced in the communication link between the third and fourth vehicles. Figure 9 shows the norm of the residuals of the transmissibilities $\mathcal{T}_{12}, \mathcal{T}_{23}, \mathcal{T}_{34}$, and $\mathcal{T}_{45}$ computed using (16) with $w = 100$ steps. Note that at approximately $t = 100$ seconds, the norm of the residual of the operator $\mathcal{T}_{23}$ exceeds the threshold limit, which activates the sliding mode controller. At time $t = 102$ seconds, the norm of the residuals of the operator $\mathcal{T}_{23}$ drops below the threshold limit due to using the sliding mode control.

VII. Experimental Results

To verify the proposed method, we consider the experimental setup shown in Figure 10 consisting of three autonomous Quanser QBot2e robots. Each robot consists of

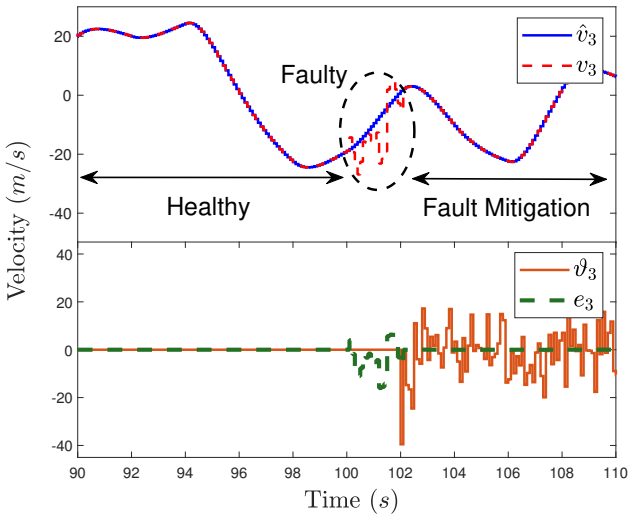


Fig. 8: Simulation results: Measured and predicted values of v_3. At $t = 100$ seconds a cyber attack is introduced, which leads to the shown prediction error e_3. At time $t = 102$ seconds, the sliding mode controller is activated and the shown control signal ϑ_3 is produced.

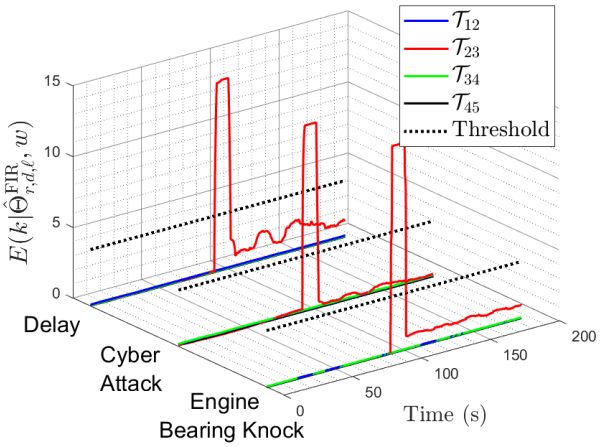


Fig. 9: Simulation results: Norm of the residuals of the transmissibilities $\mathcal{T}_{12}, \mathcal{T}_{23}, \mathcal{T}_{34}$, and $\mathcal{T}_{45}$ computed using (16) with $w = 100$ steps. Note that at approximately $t = 100$ seconds, the norm of the residual of the operator $\mathcal{T}_{23}$ exceeds the threshold limit, which activates the sliding mode controller. Moreover, note that after activating the sliding mode controller the norm of the residual drops below the threshold value.

two coaxial wheels (driven by DC motors). The variance between the wheels velocities gives angular velocity for the robot. The first robot receives the excitation signals from a computer through a wireless connection, and the second robot is connected with the first robot via V2V communication. Similarly, the third is connected with the second robot via V2V communication.

For health monitoring, we consider a one-dimensional

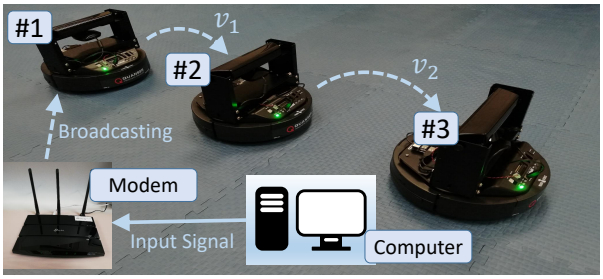


Fig. 10: The experimental setup. The first robot receives the input signal from the computer while the second and third robots receive the input signal from the preceding robot via V2V communication.

motion for the platoon. We first run the setup by sending a zero-mean, unit variance, Gaussian random excitation signal to the first robot, all robots run and move simultaneously depending only on the V2V communications. Note that using a noncausal FIR model of the transmissibility results in a delay in the predicted output of the transmissibility. Therefore, to avoid delays in the closed-loop system due to using a noncausal model of the transmissibility, we use causal FIR models to identify the transmissibilities. We use least squares with a causal FIR model with $r = 50$ and $d = 0$ to identify the transmissibility operators $\mathcal{T}_{12}$, and $\mathcal{T}_{23}$, where $\mathcal{T}_{12}$ is the transmissibility from robot 1 to robot 2 and $\mathcal{T}_{23}$ is the transmissibility from robot 2 to robot 3. The estimated Markov parameters for the transmissibility operator $\mathcal{T}_{12}$ are shown in Figure 11.

A. Noise injection

We consider injecting noise that makes the velocities of the wheels in the second robot not equal, which results in a 2-D motion of the second robot (i.e. a physical fault). Next, the estimated transmissibilities are used with the measurements of v_1 to obtain an estimate of the velocities v_2 as shown in the first part of Figure 12 before $t = 50$ seconds. At $t = 50$ seconds, the noise is injected, which increases the level of the error e_2 as shown in the second part of Figure 12. At $t = 54$ seconds, the cumulative error exceeds the threshold limit as shown in Figure 13. This activates the sliding mode controller, which produces the control signal ϑ_2 shown in the second part of Figure 12. Note that after $t = 54$ seconds, the controller is activated and the measured velocity is close to the predicted velocity.

B. Cyber attack and time-delay faults

Similar results can be obtained for the cyber attack and time-delay faults, which we apply individually. For the cyber attack, a noise signal is injected in the communication link between robot 1 and robot 2. For the time-delay fault, we consider the case of 2 seconds of time delay applied to robot 2. Next, the estimated transmissibilities are used with the measurements of v_1 to obtain an estimate of the velocity v_2 as shown in the first part of Figure 12 for $t < 50$ seconds. At $t = 50$ seconds, the cyber attack and time delay faults are introduced, which leads to the prediction error e_2 shown in the second part of Figure 12.

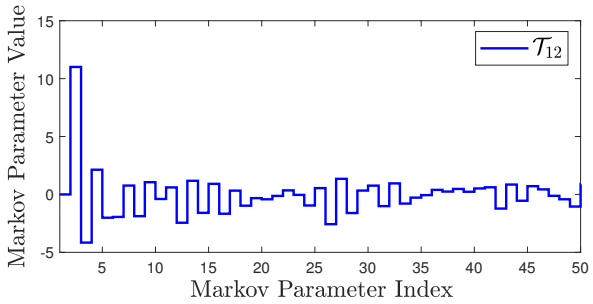


Fig. 11: Experimental results: Estimated Markov parameters for the transmissibility operator $\mathcal{T}_{12}$ obtained using least squares with a causal FIR model with $r = 50$ and $d = 0$.

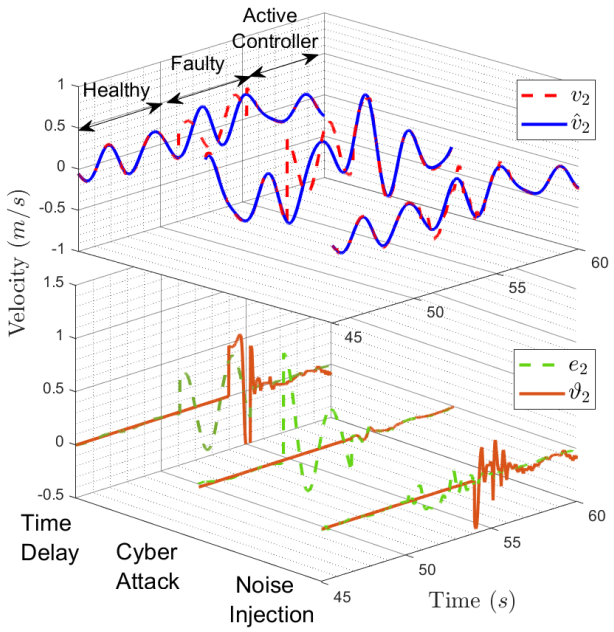


Fig. 12: Experimental results: Measured output velocity and predicted output velocity of v_2. At $t = 50$ seconds the proposed faults are introduced, which leads to an increase in the level of the error e_2. At $t = 54$ seconds the sliding mode controller is activated and the control signal ϑ_2 is produced, which leads to a drop in the error e_2.

Figure 13 shows the norm of the residuals of the transmissibilities $\mathcal{T}_{12}$ and $\mathcal{T}_{23}$ obtained using (16) with $w = 100$ steps. Note from Figure 13 that at $t = 53$ seconds, the cumulative error exceeds the threshold limit, which activates the sliding mode controller that produces the control signals ϑ_2 for the cyber attack and the time-delay as shown in the second part of Figure 12. Note from Figure 13 that after the controller is activated at $t = 53$ seconds, the error in the measured velocities drops below the threshold value. Note that at approximately $t = 50$ seconds, the norm of

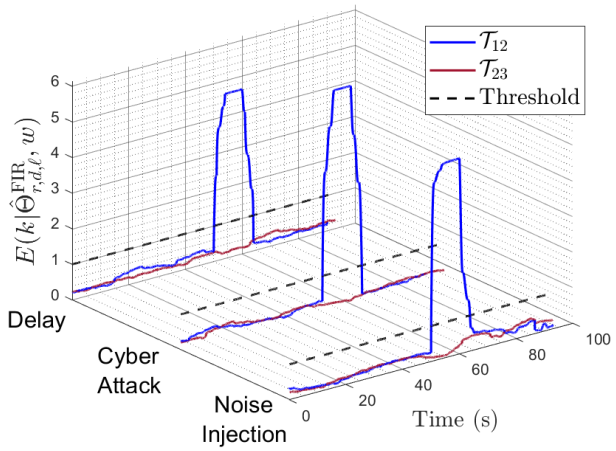


Fig. 13: Experimental results: Norm of the residual of the transmissibilities $\mathcal{T}_{12}$ and $\mathcal{T}_{23}$ computed using (16) with $w = 100$ steps. Note that at approximately $t = 50$ seconds, the norm of the residual of the operator $\mathcal{T}_{12}$ exceeds the threshold limit which activates the sliding mode controller. At approximately $t = 54$ seconds, the norm of residuals of the operator $\mathcal{T}_{23}$ drops below the threshold limit.

residual of the operator $\mathcal{T}_{12}$ reaches the threshold limit. This activates the sliding mode controller in the second robot. At approximately $t = 54$ seconds, the norm of the residual of the operator $\mathcal{T}_{23}$ drops below the threshold limit. Since the norm of the residual computed over a window with a width of 10 seconds, we can conclude that the sliding mode controller took about 4 seconds to mitigate the faults.

VIII. Conclusion

In this paper, we used transmissibility operators for fault detection and mitigation in a network of autonomous vehicles. Transmissibility-based health monitoring uses available sensor measurements for fault detection under unknown excitation and unknown dynamics of the network. After detecting a fault, a sliding mode controller is activated to mitigate the fault. We first consider simulating a network of vehicles, where the model of the network was obtained using the bond graph approach. Next, we considered an experimental setup consisting of three autonomous Quanser QBot2e robots. In both cases, the proposed approach was used effectively for fault detection and mitigation.

Acknowledgement

This work was supported by the Natural Sciences and Engineering Research Council of Canada.

References

[1] R. Hult, G. R. Campos, E. Steinmetz, L. Hammarstrand, P. Falcone, and H. Wymeersch, "Coordination of cooperative autonomous vehicles: Toward safer and more efficient road transportation," *IEEE Signal Processing Magazine*, vol. 33, pp. 74–84, 2016.

[2] S. Darbha, S. Konduri, and P. R. Pagilla, "Benefits of V2V communication for autonomous and connected vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, pp. 1954–1963, 2018.

[3] X. Jin, W. M. Haddad, Z.-P. Jiang, and K. G. Vamvoudakis, "Adaptive control for mitigating sensor and actuator attacks in connected autonomous vehicle platoons," in *IEEE Conference on Decision and Control*, Miami Beach, FL, pp. 2810-2815, 2018.

[4] P. Seiler, A. Pant, and K. Hedrick, "Disturbance propagation in vehicle strings," *IEEE Transactions on automatic control*, vol. 49, pp. 1835–1842, 2004.

[5] Z. A. Biron, S. Dey, and P. Pisu, "Real-time detection and estimation of denial of service attack in connected vehicle systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, pp. 3893–3902, 2018.

[6] J. Lunze, "Adaptive cruise control with guaranteed collision avoidance," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, pp. 1897–1907, 2018.

[7] M. Pirani, E. Hashemi, A. Khajepour, B. Fidan, B. Litkouhi, S.-K. Chen, and S. Sundaram, "Cooperative vehicle speed fault diagnosis and correction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, pp. 783–789, 2018.

[8] A. Khalil, M. Al Janaideh, K. F. Aljanaideh, and D. Kundur, "Fault detection, localization, and mitigation of a network of connected autonomous vehicles using transmissibility identification," in *American Control Conference*, Denver, CO, pp.386-391, 2020.

[9] J. Huang, Y. Wang, and T. Fukuda, "Set-membership-based fault detection and isolation for robotic assembly of electrical connectors," *IEEE Transactions on Automation Science and Engineering*, vol. 15, pp. 160–171, 2016.

[10] K. F. Aljanaideh and D. S. Bernstein, "Time-domain analysis of sensor-to-sensor transmissibility operators," *Automatica*, vol. 53, pp. 312–319, 2015.

[11] J. Chen and R. B. Randall, "Intelligent diagnosis of bearing knock faults in internal combustion engines using vibration simulation," *Mechanism and Machine Theory*, vol. 104, pp. 161–176, 2016.

[12] F. Li, D. Mikulski, J. Wagner, and Y. Wang, "Trust-based control and scheduling for UGV platoon under cyber attacks," *SAE Technical Paper*, 2019.

[13] K. F. Aljanaideh and D. S. Bernstein, "Experimental application of time-domain transmissibility identification to fault detection and localization in acoustic systems," *Journal of Vibration and Acoustics*, vol. 140, pp. 021 017.1–11, 2018.

[14] ——, "Closed-loop identification of unstable systems using noncausal FIR models," *International Journal of Control*, vol. 90, pp. 168–185, 2017.

[15] D. C. Karnopp, D. L. Margolis, and R. C. Rosenberg, *System dynamics: modeling, simulation, and control of mechatronic systems*. John Wiley & Sons, 2012.

[16] S. E. Li, Y. Zheng, K. Li, Y. Wu, J. K. Hedrick, F. Gao, and H. Zhang, "Dynamical modeling and distributed control of connected and automated vehicles: Challenges and opportunities," *IEEE Intelligent Transportation Systems Magazine*, vol. 9, pp. 46–58, 2017.

[17] M. Pirani, E. Hashemi, A. Khajepour, B. Fidan, B. Litkouhi, S.-K. Chen, and S. Sundaram, "Cooperative vehicle speed fault diagnosis and correction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, pp. 783–789, 2018.

[18] J. Pillmann, B. Sliwa, J. Schmutzler, C. Ide, and C. Wietfeld, "Car-to-cloud communication traffic analysis based on the common vehicle information model," in *IEEE Vehicular Technology Conference (VTC Spring)*, pp. 1-5, 2017.

[19] J. A. Guerrero-Ibanez, S. Zeadally, and J. Contreras-Castillo, "Integration challenges of intelligent transportation systems with connected vehicle, cloud computing, and internet of things technologies," *IEEE Wireless Communications*, vol. 22, pp. 122–128, 2015.

[20] L. Kong, M. K. Khan, F. Wu, G. Chen, and P. Zeng, "Millimeter-wave wireless communications for IoT-cloud supported autonomous vehicles: Overview, design, and challenges," *IEEE Communications Magazine*, vol. 55, pp. 62–68, 2017.

[21] R. H. Middleton and G. C. Goodwin, *Digital Control and Estimation: A Unified Approach*. Prentice Hall PTR, USA, 1990.

[22] A. Youssef, C. Delpha, and D. Diallo, "An optimal fault detection threshold for early detection using kullback–leibler divergence for unknown distribution data," *Signal Processing*, vol. 120, pp. 266–279, 2016.

[23] T. Keijzer and R. M. Ferrari, "A sliding mode observer approach for attack detection and estimation in autonomous vehicle platoons using event triggered communication," *arXiv preprint arXiv:1911.00374*, 2019.