Secure Control of Connected and Automated Vehicles Using Trust-Aware Robust Event-Triggered Control Barrier Functions

H M Sabbir Ahmad
Boston University
sabbir92@bu.edu

Ehsan Sabouni
Boston University
esabouni@bu.edu

Akua Dickson
Boston University
akuad@bu.edu

Wei Xiao
Massachusetts Institute of Technology
weixy@mit.edu

Christos G. Cassandras
Boston University
cgc@bu.edu

Wenchao Li
Boston University
wenchao@bu.edu

***Abstract*—We address the security of a network of Connected and Automated Vehicles (CAVs) cooperating to safely navigate through a conflict area (e.g., traffic intersections, merging roadways, roundabouts). Previous studies have shown that such a network can be targeted by adversarial attacks causing traffic jams or safety violations ending in collisions. We focus on attacks targeting the V2X communication network used to share vehicle data and consider as well uncertainties due to noise in sensor measurements and communication channels. To combat these, motivated by recent work on the safe control of CAVs, we propose a trust-aware robust event-triggered decentralized control and coordination framework that can provably guarantee safety. We maintain a trust metric for each vehicle in the network computed based on their behavior and used to balance the tradeoff between conservativeness (when deeming every vehicle as untrustworthy) and guaranteed safety and security. It is important to highlight that our framework is invariant to the specific choice of the trust framework. Based on this framework, we propose an attack detection and mitigation scheme which has twofold benefits: (i) the trust framework is immune to false positives, and (ii) it provably guarantees safety against false positive cases. We use extensive simulations (in SUMO and CARLA) to validate the theoretical guarantees and demonstrate the efficacy of our proposed scheme to detect and mitigate adversarial attacks. The code for the simulated scenarios can be found in this *link*.**

I. Introduction

The emergence of Connected and Automated Vehicles (CAVs) and advancements in traffic infrastructure [1] promise to offer solutions to transportation issues like accidents, congestion, energy consumption, and pollution [2], [3]. To achieve these benefits, secure and efficient traffic management is crucial, particularly at bottleneck locations such as intersections, roundabouts, and merging roadways [4].

We focus on *decentralized* algorithms as they provide manifold benefits, including added security since an attacker can only target a limited number of agents; in contrast, in a centralized scheme an attack on the central entity can potentially compromise every agent/CAV. Security of Autonomous Vehicles (AVs) has been extensively studied in existing literature [5]–[7] whereby the attacks can be broadly categorized into in-vehicle network attacks and V2V or V2X communication network attacks. There has been significant research done [8]–[10] from a control point of view with the aim of designing efficient real-time controllers for CAVs. However, ensuring security in the implementation of these controllers has received little attention, with the literature mostly limited to the security of Cooperative Adaptive Cruise Control (CACC) [11]–[15]. These studies do not extend to the more critical parts of a traffic network such as intersections or roundabouts, where the repercussions of an attack are more severe, yet the literature addressing security in these cases is limited.

The authors in [16] propose a technique based on public key cryptography, while [17] assesses cybersecurity risks on cooperative ramp merging by targeting V2I communication with road-side units (RSU). More comprehensive studies of the security of decentralized control and coordination algorithms for CAVs can be found in [18], [19]. In [19], an attack resilient control and coordination algorithm has been proposed using Control Barrier Functions (CBFs) without any mitigation technique. Moreover, the framework in [19] only uses V2X communication without local perception, which we deem highly useful for added security. It is also not robust to uncertainties in state estimates/measurements, which poses a security limitation as many stealthy attacks are designed to go through a Bad Data Detector (BDD) undetected.

The notion of trust/reputation has been applied to multi-agent systems including Intelligent Transportation Systems (ITS) in [20]–[23]. In [24] a novel trust-based CBF framework is proposed for multi-robot systems (MRSs) to provide safe control against adversarial agents; however, this cannot be directly applied to a traffic network as it is limited to a specific characterization of agents that does not apply to a

This work was supported in part by in part by NSF under grants CPS 1932162, ECCS-1931600, DMS-1664644, CNS-2149511, and by ARPA-E under grant DE-AR0001282.

Symposium on Vehicles Security and Privacy (VehicleSec) 2024
26 February 2024, San Diego, CA, USA
ISBN 979-8-9894372-7-6
https://dx.doi.org/10.14722/vehiclesec.2024.23xxx
www.ndss-symposium.org

road network. The authors in [25] used a trust framework to address the security of CACC. Finally, [26] employed a trust framework to address Sybil attacks within traffic intersections using a macroscopic network model. However, it is constrained by the accuracy of the traffic density estimation model in detecting fake Vehicles (CAVs) and also offers no guarantees on preventing false positives (i.e., detecting all fake vehicles accurately and not detecting any real vehicle as fake).

The main contributions of this paper are summarized below:

1) We propose a novel robust trust-aware event-triggered control and coordination framework that guarantees safe coordination for CAVs in conflict areas in the presence of adversarial attacks. Our proposed formulation is robust against stealthy attacks that can pass through BDDs undetected. The benefit of event-triggered control lies in reducing the communication load, thus improving robustness against attacks.
2) We propose an attack detection and mitigation scheme based on the trust score of CAVs that can alleviate the effect of the attack, particularly the case of traffic holdup by restoring normal coordination. Our proposed scheme guarantees safety against false positive (FP) cases, which may arise due to a poor choice (or, design) of the trust framework.

Although, our framework views security as a specification in a control and coordination problem, it is important to note that various network security measures like cryptographic techniques can complement this framework. The paper is organized in six sections. The next section provides some background, followed by the threat model in Section III. In Section IV, we present the robust event-triggered control and coordination framework, which is followed by the attack mitigation in Section V. We present simulation results in Section VI. Finally, the conclusion is included in Section VII.

II. Background

We present a resilient control and coordination approach that includes an attack detection and mitigation scheme for secure coordination of CAVs in conflict areas using the signal-free intersection presented in [27] as an illustrative example. Figure 1 shows a typical intersection with multiple lanes. Here, the Control Zone (CZ) is the area within the circle. containing eight entry lanes labeled from O_1 to O_8 and exit lanes labeled from l_1 to l_8 each of length L which is assumed to be the same here. Red dots show all the merging points (MPs)where potential collisions may occur. All the CAVs have the following possible movements: going straight, turning left from the leftmost lane, or turning right from the rightmost lane.

The vehicle dynamics for each CAV in the CZ take the following form

$$\begin{bmatrix} \dot{x}_i(t) \\ \dot{v}_i(t) \end{bmatrix} = \begin{bmatrix} v_i(t) \\ u_i(t) \end{bmatrix}, \tag{1}$$

where $x_i(t)$ is the distance along the lane from the origin at which CAV i arrives, $v_i(t)$ and $u_i(t)$ denote the velocity and control input (acceleration/deceleration) of CAV i, respectively, $v_{max} > 0$ denotes the maximum speed and $u_{min} < 0$ is the minimum control allowed in the CZ.

A road-side unit (RSU) acts as a coordinator which receives and stores the state and control information $[x_i(t), v_i(t), u_i(t)]^T$ from CAVs through vehicle-to-infrastructure (V2X) communication. Additionally, it also stores and updates the a trust metric for each CAV in the CZ. It is assumed that the coordinator knows the entry and exit lanes for each CAV upon their arrival and uses it to determine the list of MPs in its planned trajectory. It facilitates safe coordination by providing each CAV with relevant information about other CAVs in the network, particularly those that are at risk of collision.

A. Constraints/rules in the Control Zone

Let t_i^0 and t_i^f denote the time that CAV i arrives at the origin and leaves the CZ at its exit point, respectively. In the following section we summarize the rules that the CAVs in the CZ have to satisfy so as to maintain a safe flow in the intersection.

Constraint 1 (Rear-End Safety Constraint): Let i_p denote the index of the CAV which physically immediately precedes CAV i in the CZ (if one is present). It is required that CAV i conforms to the following constraint:

$$x_{i_p}(t) - x_i(t) - \varphi v_i(t) - \Delta \geq 0, \quad \forall t \in [t_i^0, t_i^f] \tag{2}$$

where φ denotes the reaction time and $\Delta \in \mathbb{R}_{>0}$ is a given minimum safe distance which depends on the length of these two CAVs.

Constraint 2 (Safe Merging Constraint): Every CAV i should leave enough room for the CAV preceding it upon arriving at a MP, to avoid a lateral collision i.e.,

$$x_{i_m}(t_i^m) - x_i(t_i^m) - \varphi v_i(t_i^m) - \Delta \geq 0, \tag{3}$$

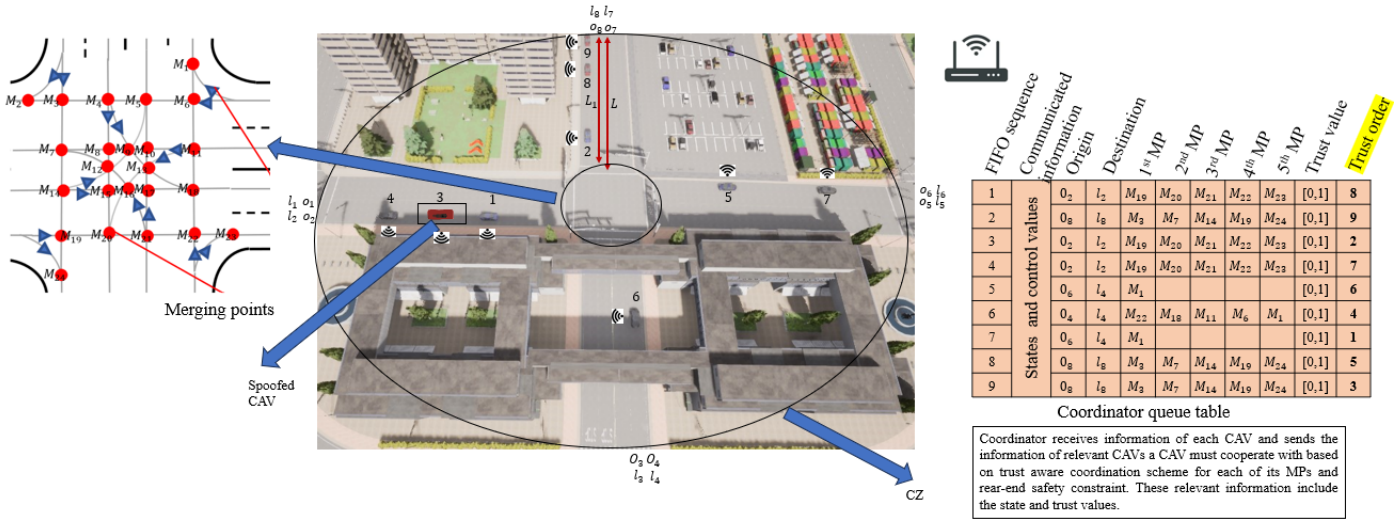
where i_m is the index of the CAV that may collide with CAV i at the merging points $m_i = \{1, ..., n_i\}$ where n_i is the total number of MPs that CAV i passes in the CZ.

Constraint 3 (Vehicle limitations): Finally, there are constraints on the speed and acceleration for each $i \in S(t)$:

$$v_{min} \leq v_i(t) \leq v_{max}, \forall t \in [t_i^0, t_i^f] \tag{4}$$

$$u_{min} \leq u_i(t) \leq u_{max}, \forall t \in [t_i^0, t_i^f] \tag{5}$$

where $v_{min} \geq 0$ denote the minimum speed, and $u_{max} > 0$ denote the maximum control allowed in the CZ respectively. v_{max} and u_{min} are as defined before. The coordinator finds CAV i_p and CAV i_m for each CAV $i \in S(t)$ from their trajectory and communicates it to CAV i. The determination depends on the policy adopted for sequencing CAVs whose relative performance has been studied in [28]. A common sequencing scheme is the First In First Out (FIFO) policy whereby CAVs exit the CZ in the order they arrive.



FIFO sequence	States and control values	Communicated Information Origin	Destination	1st MP	2nd MP	3rd MP	4th MP	5th MP	Trust value	Trust order
1		0_2	l_2	M_{19}	M_{20}	M_{21}	M_{22}	M_{23}	[0,1]	8
2		0_8	l_8	M_3	M_7	M_{14}	M_{19}	M_{24}	[0,1]	9
3		0_2	l_2	M_{19}	M_{20}	M_{21}	M_{22}	M_{23}	[0,1]	2
4		0_2	l_2	M_{19}	M_{20}	M_{21}	M_{22}	M_{23}	[0,1]	7
5		0_6	l_4	M_1					[0,1]	6
6		0_4	l_4	M_{22}	M_{18}	M_{11}	M_6	M_1	[0,1]	4
7		0_6	l_4	M_1					[0,1]	1
8		0_8	l_8	M_3	M_7	M_{14}	M_{19}	M_{24}	[0,1]	5
9		0_8	l_8	M_3	M_7	M_{14}	M_{19}	M_{24}	[0,1]	3

Fig. 1. The multi-lane intersection problem. Collisions may happen at the MPs (red dots shown in above figure).

B. Decentralized control formulation:

Under this formulation, each CAV i determines its control policy in a *decentralized manner* based on some objective that includes minimizing travel time and energy consumption, maximizing comfort, etc., governed by the dynamics (1). Expressing energy through $\frac{1}{2}u_i^2(t)$ we use $\alpha \in [0,1]$ as a relative weight between the time and energy objectives, which can be properly normalized by setting $\beta := \frac{(1-\alpha)\max\{u_{max}^2, u_{min}^2\}}{2\alpha}$ to penalize travel time relative to the energy cost of CAV i. Then, we can formulate an Optimal Control Problem (OCP) as follows:

$$J_i(u_i(t), t_i^f) := \beta(t_i^f - t_i^0) + \int_{t_i^0}^{t_i^f} \frac{1}{2}u_i^2(t)dt \tag{6}$$

subject to Constraints (1)-(3).

C. Trust framework

Let $\mathcal{B}$ be a set of indices associated with the behavioral specifications that are used to evaluate the trust of a vehicle. The behavior specifications used in our experiments are listed in [19]. For example, conformity to the underlying physical model is a specification that each CAV has to satisfy all the time. For each CAV $i \in S(t)$, $\forall t \in [t_i^0, t_i^f]$ the coordinator assigns positive evidence $r_{i,j}(t)$ and negative evidence $p_{i,j}(t)$ for conformance and violation respectively of every specification $j \in \mathcal{B}$ respectively (where $0 \leq r_{i,j}(t) \leq r_{max}, 0 \leq p_{i,j}(t) \leq p_{max}$), which it uses to update the trust $\tau_i(t)$. We define $R_i(t)$ and $P_i(t)$ as cumulative positive and negative evidence for CAV i at time t discounted by trust of other CAVs (if the check involves another CAV, as in (2) and (3), as they can be untrustworthy). We also define a time discount factor $\gamma \in (0,1)$ as shown below. In addition, we use a non-informative prior weight h_i as in [21], [29]. Let the set of checks for every CAV involving another CAV(s) be denoted by $\mathcal{B}_a \subset \mathcal{B}$. The set of other CAVs involved in check $j \in \mathcal{B}_a$ when applied to CAV i, is denoted as $S_{i,j}(t) \subseteq S(t)/\{i\}$. Then, the trust metric is updated as follows:

$$\tau_i(t) = \frac{R_i(t)}{R_i(t) + P_i(t) + h_i} \quad \forall i \in S(t) \tag{7}$$

$$\begin{aligned} R_i(t) =& \gamma R_i(t-1) + \sum_{j \in \mathcal{B}\setminus\mathcal{B}_a} r_{i,j}(t) + \sum_{j\in\mathcal{B}_a} \prod_{k\in S_{i,j}} \tau_k(t) r_{i,j}(t) \\ P_i(t) =& \gamma P_i(t-1) + \underbrace{\sum_{j \in \mathcal{B}\setminus\mathcal{B}_a} p_{i,j}(t) + \sum_{j\in\mathcal{B}_a} \prod_{k\in S_{i,j}} \tau_k(t) p_{i,j}(t)}_{p_i(t)} \\ & \forall i \in S(t), \forall t \in [t_i^0, t_i^f] \end{aligned} \tag{8}$$

Finally, we define a lower trust threshold $\delta \in (0, 1/2)$, and a higher trust threshold $1-\delta$ for subsequent sections. It is important to emphasize that, in practice, the magnitude of negative evidence is different and significantly higher compared to the magnitude of positive evidence. Note that, every vehicle is deemed untrustworthy i.e. $\tau_i(t_i^0) = 0 \; \forall i$ upon arrival in the CZ. The coordinator updates the trust for every CAV based on the outcome of the behavior specification checks. The zero trust model is used to prioritize safety in our proposed framework.

III. Threat Model

The adversarial effects of malicious attacks, as highlighted in [18], consist of creating traffic jams across multiple roads due to the cooperative aspect of the control scheme, and, in the worst case, accidents. This warrants making the control robust against these attacks. We consider the attacker models presented in [18] in what follows.

Definition 1: (Safe coordination) Safe coordination is defined as the ability to guarantee the satisfaction of (2) and (3) for every CAV $i \in S(t)$ $\forall t$ while also conforming to (4) and (5).

Definition 2: (Adversarial agent) An agent is called *adversarial* if it has one of the following objectives: (i) prevent safe coordination, (ii) reduce traffic throughput.

Assumption 1: Adversarial agents do not collide with other CAVs, nor do they attempt to cause collisions between CAVs and themselves to avoid inflicting loss on themselves.

Sybil attack A single adversarial agent (could be a CAV or attacker nearby the CZ) may spoof one or multiple unique identities and register them in the coordinator queue table as detailed in [19]. Let $S_x(t)$ and $S_s(t)$ be the set of the indices of normal and fake CAVs in the FIFO queue of the coordinator unit. Therefore at any time t, there are $N(t) = |S_x(t)|+|S_s(t)|$ CAVs which communicate their state and control information to the coordinator. A Sybil attack is one where the $S_s(t) \subset S(t)$ is a nonempty set that is located in the coordinator queue table, but unknown to the coordinator.

Assumption 2: There is an upper bound on the maximum number of fake CAVs that an adversary can spoof during a Sybil attack due to resource and energy limitations.

Assumption 3: (Bad data detection) The CAVs are equipped with BDDs whereby $\|\boldsymbol{x}_i(t) - \hat{\boldsymbol{x}}_i(t)\|_\infty \leq \epsilon$, $\forall t, \forall i \in S(t)$ where $\hat{\boldsymbol{x}}_i(t)$ is the measured/estimated state of CAV i at time t and $\|\boldsymbol{x}\|_\infty$ is the infinity norm of the state vector.

Stealthy attack An attack is stealthy if $\|\boldsymbol{x}(t) - \hat{\boldsymbol{x}}(t)\|_\infty \leq \epsilon_1$. Such attacks can be injected through targeting V2I and in-vehicular networks as well as onboard sensing systems.

Specifically, we consider bias injection attacks as defined below:

Bias Injection (BI) attack An adversarial agent may attempt to violate safe coordination amongst CAVs, or affect the traffic by targeting one or more CAVs using Person-In-The-Middle attack by adding bias to the data sent by the CAVs to the RSU, or the data sent by the RSU to the CAVs containing state information of the relevant CAVs, or both of them. Let $\boldsymbol{y_i}(t)$, $i \in S(t)$ be the data (of CAV i, or data for CAV i containing the information of the relevant CAVs) injected by the adversary during the attack; and $\boldsymbol{z_i}(t)$ be the actual data (of CAV i sent to the RSU or data for CAV i containing the information of the relevant CAVs sent by the RSU). Then, during the BI attack, $\boldsymbol{y_i}(t) = \boldsymbol{z_i}(t) + \boldsymbol{g_i(t)}$ where $\|\boldsymbol{g_i}(t)\|_\infty \leq \epsilon_1$ is the mapping used by the adversary to generate false data being stealthy.

Assumption 4: We assume that the coordinator is trustworthy i.e., it is not targeted by attacks.

IV. Safe and Resilient Control Formulation using Trust Aware CBFs

A. Trust-aware coordination

The RSU assigns each CAV a unique index based on a passing sequence policy and this information is tabulated and stored according to the assigned indices as shown in Fig. 1. For example, under a FIFO passing sequence the coordinator assigns $N(t) + 1$ to a new CAV upon arriving in the CZ. Similarly, each time a CAV i leaves the CZ, it is dropped from the table and all CAV indices larger than i decrease by one.

The coordinator computes and updates the trust metric for each CAV in the CZ as shown in figure 1. The trust metric is incorporated to the selected passing sequence to identify the CAVs any given CAV has to cooperate within the CZ. The cooperation with a CAV involves either constraint (2), or (3). According to this method, for every CAV $i \in S(t)$ and for every MP $j \in m_i$, the coordinator identifies the indices of all CAVs that precede CAV i at j based on the selected passing sequence until the first CAV whose trust value is greater than or equal to $1 - \delta$. This leads to a new set $S_{i,j}(t) \subset S(t)$ containing all the CAV indices identified during the search process. The coordinator follows the same search process for every MP in m_i corresponding to (3). Therefore, for each CAV i, the coordinator identifies $S_i^p(t) \subset S(t)$, and $S_i^M(t) = \cup_{j \in m} S_{i,j}(t)$ (where $S_i^p(t)$ is the set for (2) and $S_i^M(t)$ correspond to the set of indices for every MP) and the information is communicated to the CAV. For the example in Fig. 1, note that for CAV 4 we have $i_p = 3$, however since $\tau_3 < 1 - \delta$, the search process will continue and return $S_{4,p} = \{1, 3\}$.

Local sensing We also assume that each CAV has a vision-based perception capability defined by a radius and angle pair denoted as (r, θ), (where $r \in \mathbb{R}^+$,$\theta \in [0, 2\pi]$). The incorporation of local sensing into CAV $i \in S(t)$ adds additional constraints of the form (2) to the control problem, besides the constraints corresponding to $S_i^p(t)$ and $S_i^M(t)$ returned by *trust-based search*. Every CAV $i \in S(t)$ is able to estimate the states of every observed CAV j within its sensing range. CAV i is able to estimate the state of the preceding CAV (if there is one and it is within sensing range) and in the vicinity of MPs in its own trajectory; in particular, the CAV that will precede i immediately at its next MP should be visible to CAV i.

We consider state estimates and communication information from the coordinator to be noisy as defined below:

$$\hat{\boldsymbol{x}}_i(t) = \boldsymbol{x}_i(t) + \boldsymbol{w}_i(t) \tag{9}$$

where $\boldsymbol{w}_i(t) = [w_i^{(x)}(t), w_i^{(v)}(t)]^T$ is random measurement noise with bounded support $\|\boldsymbol{w}_i\|_\infty \leq \epsilon_1$, $\forall i \in S(t)$. We can set $\epsilon = \epsilon_1$ (same as the bound for stealthy attacks) to make the controller robust to both noise and stealthy attacks.

The OCBF Controller. This approach uses the OCP formulation in (6) with each state constraint $b_q(\boldsymbol{x}(t)) \geq 0$ mapped onto a new constraint which has the property that it implies $b_q(\boldsymbol{x}(t)) \geq 0$ and it is *linear* in the control input. The function $b_q(\boldsymbol{x}(t))$ is called a Control Barrier Function (CBF) [30]. We use such CBFs so as to ensure the constraints (2), (3), (4) and (5) are satisfied subject to the vehicle dynamics in (1) by defining $f(\boldsymbol{x}_i(t)) = [v_i(t), 0]^T$ and $g(\boldsymbol{x}_i(t)) = [0, 1]^T$. Each of these constraints can be easily written in the form of $b_q(\boldsymbol{x}_{i,j}(t)) \geq 0$, $q \in \{1, 2, 3, 4\}$ where n stands for the number of constraints only dependent on state variables $\boldsymbol{x}_{i,j}(t) = [\boldsymbol{x}_i(t), \boldsymbol{x}_j(t)]^T$. The general form of the *transformed* CBF-based constraints is:

$$L_f b_q(\boldsymbol{x}_{i,j}(t)) + L_g b_q(\boldsymbol{x}_{i,j}(t)) u_i(t) + \kappa_q(b_q(\boldsymbol{x}_{i,j}(t))) \geq 0 \tag{10}$$

where L_f, L_g are the Lie derivatives of a function along the system dynamics defined by f, g above and κ_q is a class $\mathcal{K}$ function. By combining the OCP formulation in (6) with the CBF-based constraints of the form (10) instead of the original ones, we obtain the Optimal control with CBFs (termed OCBF) approach detailed in [10].

Finally, the road speed limit can be included as a reference $v_i^{ref}(t)$ treated by the controller as a soft constraint using a Control Lyapunov Function (CLF) [28] by setting $V(\boldsymbol{x}_i(t)) = (v_i(t) - v_i^{ref}(t))^2$, rendering the following constraint:

$$L_f V(\boldsymbol{x}_i(t)) + L_g V(\boldsymbol{x}_i(t))\boldsymbol{u}_i(t) + c_i V(\boldsymbol{x}_i(t)) \leq e_i(t), \quad (11)$$

where $e_i(t)$ makes this a soft constraint. The significance of CBFs in this approach is twofold: first, their forward invariance property [30] guarantees that all constraints they enforce are satisfied at all times if they are initially satisfied; second, CBFs impose *linear* constraints on the control which is what enables the efficient solution of the tracking problem through a sequence of Quadratic Programs (QPs) thus computationally efficient and suitable for real-time control.

B. Trust-Aware CBFs

The choice of the class $\mathcal{K}$ function in (10) determines the rate at which an agent/CAV reaches the boundary of the safety set. Thus, the choice of this function provides a tradeoff between conservativeness and safety. We can choose a conservative candidate function to prioritize safety by considering all agents to be untrustworthy. However, in view of the available trust metric, we incorporate it in the function with the aim of balancing this tradeoff. The underlying idea is that the degree of conservativeness of a CBF constraint corresponding to a CAV i with respect to CAV j can be adjusted by incorporating the *trust* of CAV j, τ_j, in it as shown below:

$$L_f b_q(\boldsymbol{x}_{i,j}(t)) + L_g b_q(\boldsymbol{x}_{i,j}(t))u_i(t) + \kappa_{q,\tau_j}(b_q(\boldsymbol{x}_{i,j}(t))) \geq 0. \quad (12)$$

An example for the choice of a class $\mathcal{K}$ function is $\kappa_q(b_q(\boldsymbol{x}_{i,j}(t))) = c_{i,j}\tau_j(t)b_q(\boldsymbol{x}_{i,j}(t))$, where $c_{i,j} \in \mathbb{R}^+$ is a scaling factor.

C. Robust Trust-Aware CBFs

In the presence of noisy measurements (estimates) as in (9) the corresponding CBF constraint in (10) can be rewritten as follows due to (9):

$$L_f b_q(\hat{\boldsymbol{x}}_{i,j}(t) - \boldsymbol{w}_{i,j}(t)) + L_g b_q(\hat{\boldsymbol{x}}_{i,j}(t) - \boldsymbol{w}_{i,j}(t))u_i(t) + \kappa_{q,\tau_j}(b_q(\hat{\boldsymbol{x}}_{i,j}(t) - \boldsymbol{w}_{i,j}(t))) \geq 0. \quad (13)$$

where $\boldsymbol{w}_{i,j}(t) = [\boldsymbol{w}_i(t), \boldsymbol{w}_j(t)]^T$. For example, the CBF constraint corresponding to (2) is as follows:

$$v_{i_p}(t) - v_i(t) - \varphi u_i(t) - \kappa_{q,\tau_{i_p}}(x_{i_p}(t) - x_i(t) - \varphi v_i(t) - \Delta) \geq 0 \quad (14)$$

In the presence of noise $\boldsymbol{w}_i(t)$, according to (9) this becomes:

$$\hat{v}_{i_p}(t) + w_{i_p}^{(v)}(t) - \hat{v}_i(t) - w_i^{(v)}(t) - \varphi u_i(t) - \kappa_{q,\tau_{i_p}}(\hat{x}_{i_p}(t) + w_{i_p}^{(x)}(t) - \hat{x}_i(t) - w_i^{(x)}(t) - \varphi\hat{v}_i(t) - \varphi w_i^{(v)}(t) - \Delta) \geq 0$$

Obviously, the random noise $\boldsymbol{w}_{i,j}(t)$ is unknown, hence, we use the bound ϵ_1 on the noise to derive the following lemma for the robust trust-aware CBF.

Lemma 1: Given a constraint $b_q(\boldsymbol{x}(t))$ associated with the set $\mathrm{C} := \{\boldsymbol{x} \in \mathbb{R}^n : b_q(\boldsymbol{x}) \geq 0\}$ and $\|\boldsymbol{w}_{i,j}\|_\infty \leq \epsilon_1$, any Lipschitz continuous controller $u(t)$ that satisfies

$$\min_{\{\boldsymbol{w}_{i,j}(t):\|\boldsymbol{w}_{i,j}(t)\|_\infty \leq \epsilon_1\}} [L_f b_q(\hat{\boldsymbol{x}}_{\boldsymbol{i,j}}(t) - \boldsymbol{w}_{i,j}(t))] + L_g b_q(\hat{\boldsymbol{x}}_{\boldsymbol{i,j}}(t) - \boldsymbol{w}_{i,j}(t))u_i(t) + \kappa_{q,\tau_j}(b_q(\hat{\boldsymbol{x}}_{\boldsymbol{i,j}}(t) - \boldsymbol{w}_{i,j}(t)))] \geq 0 \quad (15)$$

renders the set C forward invariant $\forall t \geq t_0$ for the system (1).

Proof: The satisfaction of (15) guarantees the satisfaction of the constraint (13) (and (10)) since it is a lower bound for (10) which according to Theorem 1 in [10] makes the set C forward invariant $\forall t \geq t_0$ w.r.t (1). ■

Based on the information in the table (as shown in Fig. 1) the coordinator communicates the state information and the trust value of the CAVs in S_i^p and S_i^M corresponding to constraints (2) and (3) respectively to each CAV i in the CZ.

The OCBF problem corresponding to (6) is formulated as:

$$\min_{u_i(t),e_i(t)} J_i(u_i(t), e_i(t)) := \int_{t_i^0}^{t_i^f} [\frac{1}{2}(u_i(t) - u_i^{ref}(t))^2 + \lambda e_i^2(t)]dt \quad (16)$$

subject to vehicle dynamics (1), the CBF constraints (15), $\forall q = \{1, ..., n\}$ and CLF constraint (11). In this approach, u_i^{ref} is generated by solving the *unconstrained* optimal control problem in (6) which can be analytically obtained. The resulting control reference trajectory is optimally tracked subject to the constraints.

D. Event-triggered Control

A common way to solve (16)) is to discretize $[t_i^0, t_i^f]$ into intervals $[t_i^0, t_i^0 + \Delta], ..., [t_i^0 + k\Delta, t_i^0 + (k+1)\Delta], ...$ with equal length Δ and solving (16) over each time interval. The decision variables $u_{i,k} = u_i(t_{i,k})$ and $e_{i,k} = e_i(t_{i,k})$ are assumed to be constant on each interval and can be easily calculated at time $t_{i,k} = t_i^0 + k\Delta$ through solving a QP at each time step:

$$\min_{u_{i,k},e_{i,k}} [\frac{1}{2}(u_{i,k} - u_i^{ref}(t_{i,k}))^2 + \lambda e_{i,k}^2] \quad (17)$$

subject to the CBF constraints (15), $\forall q = \{1, ..., n\}$, CLF constraint (11) and dynamics (1), where all constraints are linear in the decision variables.

This is referred to as the *time-driven* approach. The main problem with this approach is that there is no guarantee for the feasibility of each CBF-based QP, as it requires a small enough discretization time which is not always possible to achieve. Also, it is worth mentioning that synchronization is required amongst all CAVs which can be difficult to impose in real-world applications. Therefore, to tackle these issues we adopt an *event-triggered* control scheme inspired by [31]. Under this scheme, the control for a CAV is updated by solving the QP (17) upon the occurrence of any of a predefined set of events (not in the original time-driven fashion) with the goal of ensuring that the state trajectory of the CAV satisfies all the constraints between two consecutive events. We will formulate such a framework for a CAV i w.r.t to another CAV j for a constraint $q \in \{1, \ldots, 4\}$, corresponding to (2), (3) and (4), which generalizes to every other CAV and constraints. Let $t_{i,k}$, and $t_{i,k+1}$ (where $k = 1, 2, ...$), be the time for the k-th and $(k+1)$-th event during which vehicle i solves its QP (17). The goal is to guarantee that the state trajectory does not violate any safety constraints within the interval $(t_{i,k}, t_{i,k+1}]$. We define C_i to be the feasible set of constraints (only dependent on our states (4)) and involving states of another CAV (2),(3)) defined

as:

$$C_i \equiv \Big\{ \boldsymbol{x}_{i,j} \in \mathbf{X}^2 : b_q(\boldsymbol{x}_{i,j}) \geq 0 \text{ and } b_q(\boldsymbol{x}_i) \geq 0, \\ j \in S_i^P \cup S_i^M \Big\} \tag{18}$$

We define a compact convex set on the state space of CAV i at time $t_{i,k}$ such that:

$$X_i(t_{i,k}) = \Big\{ \boldsymbol{y}_i \in \mathbf{X} : |\boldsymbol{y}_i - \boldsymbol{x}_i(t_{i,k})| \leq \boldsymbol{s}_{\boldsymbol{x}_i} \Big\} \tag{19}$$

where $\boldsymbol{s}_i \in \mathbb{R}^2_{>0}$ is a parameter vector. Similarly, we define a compact convex set on the trust metric:

$$\mathcal{T}_j(t_{j,k}) = \Big\{ \tau_j \in [0,1] : |\tau_j - \tau_i(t_{j,k})| \leq s_{\tau_j} \Big\} \tag{20}$$

Intuitively, this choice reflects a trade-off between computational efficiency and conservativeness. A larger choice of value makes the controller conservative requiring less frequent control update thus being more computationally efficient, and vice versa. As we use robust CBFs, we need to modify the previously defined sets to adjust the bounds on noisy states as in (1). At first, we define the feasible set of constraints as following:

$$\hat{C}_i \equiv \Big\{ \hat{\boldsymbol{x}}_{i,j} \in \mathbf{X}^2 : \min_{\{\boldsymbol{w}_{i,j} : \|\boldsymbol{w}_{i,j}\|_\infty \leq \epsilon_1\}} b_q(\hat{\boldsymbol{x}}_{i,j} - \boldsymbol{w}_{i,j}) \geq 0 \text{ and } \\ \min_{\{\boldsymbol{w}_i \| \boldsymbol{w}_i\|_\infty \leq \epsilon_1\}} b_q(\hat{\boldsymbol{x}}_i - \boldsymbol{w}_i) \geq 0 \Big\} \tag{21}$$

Note that $\hat{C}_i \subset C_i$ because $b_q(\hat{\boldsymbol{x}}_{i,j} - \boldsymbol{w}_{i,j}) \geq \min_{\{\boldsymbol{w}_{i,j} : \|\boldsymbol{w}_{i,j}\|_\infty \leq \epsilon_1\}} b_q(\hat{\boldsymbol{x}}_i - w_i) \geq 0$. The minimum can be derived in closed form as shown below:

$$\begin{aligned} &\min_{\{\boldsymbol{w}_{i,j} : |\boldsymbol{w}_{i,j}\|_\infty \leq \epsilon_1\}} b_q(\hat{\boldsymbol{x}}_{i,j} - \boldsymbol{w}_{i,j}) \\ &= \min_{\{\boldsymbol{w}_{i,j} : |\boldsymbol{w}_{i,j}\|_\infty \leq \epsilon_1\}} \hat{x}_j - w_j^{(x)} - \hat{x}_i + w_i^{(x)} - \varphi(\hat{v}_i - w_i^{(v)}) - \Delta \\ &= \hat{x}_j - \hat{x}_i - \varphi v_i - \Delta - \epsilon_1(2+\varphi) \end{aligned} \tag{22}$$

$$\begin{aligned} &\min_{\{\boldsymbol{w}_i : |\boldsymbol{w}_{i,j}\|_\infty \leq \epsilon_1\}} b_3(\hat{\boldsymbol{x}}_i - \boldsymbol{w}_i) \\ &= \min_{\{\boldsymbol{w}_i : |\boldsymbol{w}_{i,j}\|_\infty \leq \epsilon_1\}} \hat{v}_i - w_i^{(v)} - v_{min} \\ &= \hat{v}_i - v_{min} - \epsilon_1 \end{aligned} \tag{23}$$

$$\min_{\{\boldsymbol{w}_i : |\boldsymbol{w}_{i,j}\|_\infty \leq \epsilon_1\}} v_{max} - \hat{v}_i + w_i^{(v)} = v_{max} - \hat{v}_i - \epsilon_1 \tag{24}$$

We can similarly define $\hat{X}_j(t_{i,k})$:

$$\hat{X}_i(t_{i,k}) = \Big\{ \hat{\boldsymbol{y}}_i \in \mathbf{X} : |\hat{\boldsymbol{y}}_i - \hat{\boldsymbol{x}}_i(t_{i,k})| \leq \boldsymbol{s}_{\boldsymbol{x}_i} - 2[\epsilon_1, \epsilon_1]^T \Big\} \tag{25}$$

where $\hat{\boldsymbol{y}}_i = \boldsymbol{y}_i + \boldsymbol{w}_i$. Note that $\hat{X}_i(t_{i,k}) \subset X_i(t_{i,k})$ since

$$\begin{aligned} |\boldsymbol{y}_i - \boldsymbol{x}_i(t_{i,k})| &= |\hat{\boldsymbol{y}}_i - \boldsymbol{w}_i - \hat{\boldsymbol{x}}_i(t_{i,k}) + \boldsymbol{w}_i(t_{i,k})| \\ &\leq |\hat{\boldsymbol{y}}_i - \hat{\boldsymbol{x}}_i(t_{i,k}) + 2\|\boldsymbol{w}\|_\infty [1,1]^T| \\ &= |\hat{\boldsymbol{y}}_i - \hat{\boldsymbol{x}}_i(t_{i,k})| + 2[\epsilon_1, \epsilon_1]^T \end{aligned}$$

Thus, $|\hat{\boldsymbol{y}}_i - \hat{\boldsymbol{x}}_i(t_{i,k})| \leq \boldsymbol{s}_{\boldsymbol{x}_i} - 2[\epsilon_1, \epsilon_1]^T \Rightarrow |\boldsymbol{y}_i - \boldsymbol{x}_i(t_{i,k})| \leq \boldsymbol{s}_{\boldsymbol{x}_i}$.

Next, we seek a bound and a control law that satisfies the safety constraints within this bound. This can be accomplished by considering the minimum value of each component of (15) as shown next. For the first term, let

$$b_{q,f_i}^{min}(t_{i,k}) = \min_{\substack{\hat{\boldsymbol{y}}_i \in S_i(t_{i,k}) \\ \hat{\boldsymbol{y}}_j \in S_{i,j}(t_{i,k}) \\ \{\boldsymbol{w}_{i,j} : \|\boldsymbol{w}_{i,j}\|_\infty \leq \epsilon_1\}}} L_f b_q\big(\hat{\boldsymbol{y}}_{i,j}(t_{i,k}) - \boldsymbol{w}_{i,j}\big) \tag{26}$$

where $\hat{\boldsymbol{y}}_{i,j}(t_{i,k}) = [\hat{\boldsymbol{y}}_i(t_{i,k}), \hat{\boldsymbol{y}}_j(t_{i,k})]^T, S_i(t_{i,k}) := (\hat{C}_i \cap \hat{X}_i(t_{i,k}))$, and $S_{i,j}(t_{i,k}) := \hat{X}_j(t_{i,k})$. Similarly, we can define the minimum value of the third term in (15):

$$b_{\kappa_q}^{min}(t_{i,k}) = \min_{\substack{\hat{\boldsymbol{y}}_i \in S_i(t_{i,k}) \\ \hat{\boldsymbol{y}}_j \in S_{i,j}(t_{i,k}) \\ \tau_j \in \mathcal{T}_j(t_{i,k}) \\ \{\boldsymbol{w}_{i,j} : \|\boldsymbol{w}_{i,j}\|_\infty \leq \epsilon_1\}}} \kappa_{q,\tau_j}\Big(b_q(\hat{\boldsymbol{y}}_{i,j}(t_{i,k}) - \boldsymbol{w}_{i,j})\Big). \tag{27}$$

For the second term in (15), if it is not constant then the limit value $b_{2,g_i}^{min}(t_{i,k}) \in \mathbb{R}$ can be determined as follows:

$$b_{q,g_i}^{min}(t_{i,k}) = \begin{cases} \min\limits_{\substack{\hat{\boldsymbol{y}}_i \in S_i(t_{i,k}) \\ \hat{\boldsymbol{y}}_j \in S_{i,j}(t_{i,k}) \\ \{\boldsymbol{w}_{i,j} : \|\boldsymbol{w}_{i,j}\|_\infty \leq \epsilon_1\}}} L_g b_q(b_q(\hat{\boldsymbol{y}}_{\boldsymbol{i},\boldsymbol{j}}(t_{i,k}) - \boldsymbol{w}_{i,j}), \\ \text{if } u_{i,k} \geq 0 \\ \max\limits_{\substack{\hat{\boldsymbol{y}}_i \in S_i(t_{i,k}) \\ \hat{\boldsymbol{y}}_j \in S_{i,j}(t_{i,k}) \\ \{\boldsymbol{w}_{i,j} : \|\boldsymbol{w}_{i,j}\|_\infty \leq \epsilon_1\}}} L_g b_q(b_q(\hat{\boldsymbol{y}}_{\boldsymbol{i},\boldsymbol{j}}(t) - \boldsymbol{w}_{i,j})), \\ \text{otherwise,} \end{cases} \tag{28}$$

where the sign of $u_{i,k}$ can be determined by simply solving the CBF-based QP (16) at time $t_{i,k}$.

Thus, the condition that can guarantee the satisfaction of a CBF constraint in the interval $(t_{i,k}, t_{i,k+1}]$ is given by

$$b_{q,f_i}^{min}(t_{i,k}) + b_{q,g_i}^{min}(t_{i,k}) u_{i,k} + b_{\kappa_q}^{min}(t_{i,k}) \geq 0, \tag{29}$$

for $q \in \{1, \ldots, 4\}$. Note that the minimizations in (26), (28) and (27) are simple linear programs whose closed form solution can be easily derived. In order to apply this condition to the QP (17), we just replace (15) by (29) as follows:

$$\min_{u_{i,k}, e_{i,k}} \big[\frac{1}{2}\big(u_{i,k} - u_i^{ref}(t_{i,k})\big)^2 + \lambda e_{i,k}^2\big] \quad \text{s.t. } (11), (29), (5) \tag{30}$$

Finally, we can determine $t_{i,k+1}$, the next time that a solution of the QP (30) must be solved, as follows:

$$\begin{aligned} t_{i,k+1} = \min \Big\{ t > t_{i,k} : &|\hat{\boldsymbol{x}}_i(t) - \hat{\boldsymbol{x}}_i(t_{i,k})| \geq \boldsymbol{s}_{\boldsymbol{x}_i} - 2[\epsilon_1, \epsilon_1]^T \\ &\text{or } |\hat{\boldsymbol{x}}_j(t) - \hat{\boldsymbol{x}}_j(t_{i,k})| \geq \boldsymbol{s}_{\boldsymbol{x}_j} - 2[\epsilon_1, \epsilon_1]^T, \ \forall j \\ &\text{or } |\tau_j(t) - \tau_j(t_{i,k})| \geq \boldsymbol{s}_{\tau_j} \forall j \Big\}, \quad t_{i,1} = 0 \end{aligned} \tag{31}$$

The following theorem formalizes our analysis by showing that if new constraints of the general form (29) hold, then our original CBF constraints (12) also hold. The proof follows the same lines as that of a more general theorem in [32].

Theorem 1: Given a CBF $b_q(\boldsymbol{x}_{i,j}(t))$ with relative degree one, let $t_{i,k+1}$, $k = 1, 2, \ldots$ be determined by (31) with $t_{i,1} = 0$ and $b_{q,f_i}^{min}(t_{i,k})$, $b_{\gamma_q}^{min}(t_{i,k})$, $b_{q,g_i}^{min}(t_{i,k})$ obtained through (26), (27), and (28). Then, any control input $u_{i,k}$ that satisfies (29) for all $q \in \{1, \ldots, 8\}$ within the time interval $[t_{i,k}, t_{i,k+1})$

renders the set $\hat{C}_i$ and therefore C_i forward invariant for the dynamic system defined in (1).

Proof: The satisfaction of (29) satisfies the constraint (12) which in turn satisfies (15) which makes C_i forward invariant based on Lemma 1. ■

Corollary 1: The satisfaction of (29) corresponding to (2), (3) and (4), subject to (5) guarantees satisfaction of the constraints (2) and (3) for $\|\boldsymbol{x}_i(t)-\hat{\boldsymbol{x}}_i(t)\|_\infty \leq \epsilon_1 \; \forall t, \forall i \in S(t)$.

Proof: The satisfaction of (29) makes the set $\hat{C}_i$ and correspondingly set C_i for (2), (3) and (4) forward invariant from Theorem (1) for any $\|\boldsymbol{w}_i(t)\|_\infty = \|\boldsymbol{x}_i(t)-\hat{\boldsymbol{x}}_i(t)\|_\infty \leq \epsilon_1$ guaranteeing their satisfaction $\forall t, \forall i \in S(t)$. ■

Corollary 2: The trust-based coordination in conjunction with control using robust trust-aware CBFs guarantees safe navigation of CAVs against Sybil attacks and Stealthy attacks.

Proof: The trust-based search guarantees safe coordination against Sybil attacks as proved in [19] (in Theorem 1) and in conjunction with robust trust-aware CBFs from Corollary 1 makes our control and coordination framework safe against Sybil and Stealthy attacks. ■

V. Attack Detection and Mitigation

Our proposed robust control scheme offers provably safe coordination against adversarial attacks. However, there are scenarios where attackers may target the network performance by causing traffic holdup. This is possible with Sybil attacks, as illustrated in [18], necessitating attack mitigation besides safety guarantees. The problem of detection involves the identification of adversarial (or, spoofed) CAVs accurately and mitigation can be defined as reestablishing the normal cooperation in the network close to what it would be in the ideal scenario without any attack. Resilience is necessary to ensure safe coordination until the attack is detected and in the presence of any false identification of adversarial (or, spoofed) CAVs. In this section, we present our proposed mitigation framework based on the trust framework with the aforementioned objective.

A. Determination of Fake CAVs

Initially, every CAV is considered untrustworthy (i.e., $\tau_i(t_i^0) = 0$). Upon arrival in the CZ, the coordinator monitors the trust for each CAV and, if it detects any CAV $i \in S(t)$ s.t. $\tau_i(t) \leq 1-\delta$ and $\tau_i(t) \leq \tau_i(t-1)$, it initiates an observation window for that particular CAV of length η. If the trust for CAV i is non-increasing and stays below the threshold of $1-\delta$ during the observation window then the coordinator proceeds to the mitigation step.

B. Robust Mitigation

The most trivial strategy that can be adopted is to rescind cooperation with the fake CAVs; however, it is essential to note that our framework can output false positives (although highly unlikely if the priorities of the behavioral specifications are chosen as mentioned in [19]). Therefore, we offer a soft mitigation scheme; we call it "soft" because it is a passive scheme that relies on the local sensory information of the CAVs. This will become apparent in the remainder of the section. We define a *rescheduling zone* in the CZ of length L_1 as shown in Fig. 1. It has been shown that any passing sequence can be rescheduled in this area in [31]. Then, we present the following definitions.

Definition 3: (Explicitly constrained agent) An agent i is called *explicitly constrained* by an agent j at time t if it has a constraint directly involving states of agent j at that time.

Definition 4: (Implicitly constrained agent) An agent i is called *implicitly constrained* by an agent j at time t if there is any other agent k in the environment constrained by j, which constrained agent i.

We mitigate the effect of fake CAVs by unconstraining the CAVs that are explicitly constrained by them (including the physically following CAVs if they are within their perception range and do not actually see any vehicle ahead) by solving the Integral Linear Program (ILP) defined below. Let the set of the ordered indices of detected fake CAVs that we want to mitigate be denoted as $S_f(t)$. We define the index $k_{min} = \min S_f(t)$ as the index of the first (fake) CAV in the queue to re-sequence from and $S_+(k_{min}) = \{k_{min}, \ldots, N(t)\}$. Then, the ILP is formulated as follows:

$$\max_{i \in S_f(t)} \sum_{i \; \in S_f(t)} a_i \tag{32}$$

$$a_j - a_k \geq \nu, \; \forall k \in \bar{S}_f(t) \cap S_j^p(t), \text{ and } j \in S_+(k_{min}) \tag{33}$$

$$a_j - a_k \geq \nu, j \in S_+(k_{min}), k \in S_j^M(t) \tag{34}$$

$$a_j \neq a_k \; j, k \in S_+(k_{min})$$

$$\{a_{k_{min}}, \ldots, a_N(t)\} \in S_+(k_{min}); \nu \geq 1 \tag{35}$$

where (33) correspond to constraint (2), (34) correspond to constraint (3), $\{a_{k_{min}}, \ldots, a_{N(t)}\}$ are the new indices of the CAVs in $S_+(k_{min})$.

Based on the above definitions we now outline the scenarios that are of importance to us and derive an approximate solution of (32) for them.

1) **No CAVs are constrained by CAVs in** $S_f(t)$: In this case, the solution of (32) will reschedule the CAVs starting from index $k = \min S_f(t)$ in $S_f(t)$ by moving them at the end of the queue and move the remaining CAVs with original index $i \geq k$ and $i \notin S_f(t)$ ahead in the queue to fill their places in their current order. This process will be repeated $\forall k \in S_f(t)$.
2) **There are CAVs in** $S_f(t)$ **which physically precede another CAV in the CZ**: First, let us consider CAV $k \in S_f(t)$ and j is the index of physically immediately following CAV, and let $S_j^c(t) \subseteq S(t)$ be the set of CAVs explicitly and implicitly constrained by j. At first, the CAVs with indices between k to $j-1$ are moved ahead in the queue by incrementing their index by 1, then, we set $k \leftarrow j-1$ where $j-1 > k$. The reason for moving k down the queue up to $j-1$ is because k can be a real CAV which has been falsely identified as a fake CAV. Finally, remove $\{j-1, j\} \cup S_j^c(t)$ from the queue, rearrange the queue by incrementing the indices of the remaining CAVs appropriately in the queue, and

add $\{j-1, j\} \cup S_j^c(t)$ in the queue. Then, repeat the process for the remaining CAVs in $S_f(t)$. Finally, update $S_f(t)$ accordingly. The final step is needed to move any CAVs $k > j-1$ that are not explicitly constrained, or implicitly constrained by the immediately preceding CAV of CAV $j+1$ ahead of CAV $j-1$ in the queue.

Observe that the rear-end constraints are excluded for the CAVs that are physically immediately following any CAV $k \in S_f(t)$ in (35) to allow CAVs that are physically immediately behind the CAVs in $S_f(t)$ to overtake them *only if* they are not visible when within sensing range. This is necessary to guarantee safety for FP cases which will be described later.

Moreover, observe that, for CAV $k \in S_f(t)$, upon rescheduling, the index of its immediately following CAV will become $k+1$. Once within the sensing range of CAV $k+1$, if CAV k is not visible, it changes its control in (17) by removing the CBF constraint corresponding to CAV i to complete the overtake. The coordinator detects the overtake completion by checking the satisfaction of the inequality in (36), upon which it completes the final step of the (32) by swapping the indices of CAV k and $k+1$ with each other. This step is performed $\forall k \in S_f(t)$ and repeated by following the scenarios mentioned previously (i.e., the solution of ILP) until all fake CAVs reach the end of the queue.

$$\hat{x}_i(t) - \hat{x}_{i_p}(t) - \varphi \hat{v}_{i_p}(t) - \Delta \geq 0 \tag{36}$$

For FP cases, notice that for any $(j-1) \in S_f(t)$, j is the CAV physically preceding it and every CAV $j^+ > j$ that is not explicitly or implicitly constrained by j is scheduled ahead of them after the first iteration of the algorithm. There will be no further rescheduling for $j-1$, i.e., there will be no vehicles overtaking it in the same road.

Lemma 2: The proposed mitigation scheme guarantees safety for real CAVs even if they are falsely identified as fake CAVs due to a Sybil attack.

Proof: In the rescheduling zone, any real CAV $i \in S(t) \backslash S_f(t)$ only overtakes a CAV $k \in S_f(t)$ if it does not observe k through its local perception. Similarly, any CAV $i \in S(t) \backslash S_f(t)$ only ignores the CBF condition in its control and jumps ahead of a CAV in $S_f(t)$ in the intersection if it does not observe it through its local vision. This makes our proposed mitigation scheme soft (or passive) and guarantees safety for false positive cases i.e., real CAVs which have been misidentified as fake CAVs. ■

The fake CAVs are removed from the coordinator queue in one of two ways: (i) the attacker stops sending information about a fake CAV, and (ii) the fake CAV leaves the CZ.

VI. Simulation Results

In this section, we present simulation results for the application of our proposed trust-aware robust CBF based event-triggered control and coordination scheme, including results for mitigation applied to various attacks mentioned in Section III. Throughout, we set $\delta = 0.1$ and $\eta = 40$. The positive and negative evidence magnitudes for the tests in the order they are mentioned in Section II-C are: $r_i(t) = [0.6, 0.6, 0.6, 0.6]$ and $p_i(t) = [1000, 100, 50, 1]$ $\forall i \in S(t)$ and $\forall t$. The intersection dimensions are: $L = 400$m, $A = 300$m^2; and the remaining parameters are $\varphi = 1.8$s, $\Delta = 3.78$m, $\beta_1 = 1, u_{\max} = 4.905$m/s^2, $u_{\min} = -5.886$m/s^2, $v_{\max} = 108$km/h, $v_{\min} = 0$km/h. Finally, we also used a realistic energy consumption model from [33] to supplement the simple surrogate L_2-norm (u^2) model in our analysis: $f_{\text{v}}(t) = f_{\text{cruise}}(t) + f_{\text{accel}}(t)$ with

$$f_{\text{cruise}}(t) = \omega_0 + \omega_1 v_i(t) + \omega_2 v_i^2(t) + \omega_3 v_i^3(t),$$
$$f_{\text{accel}}(t) = \big(r_0 + r_1 v_i(t) + r_2 v_i^2(t)\big) u_i(t).$$

where we used typical values for parameters $\omega_1, \omega_2, \omega_3, r_0, r_1$ and r_2 as reported in [33]. The simulation was done in Sumo and Carla, where we used Sumo to generate various traffic scenarios and Carla to validate and evaluate the performance of our proposed schemes.

Trust-aware CBFs: We present results comparing trust aware robust CBFs with ordinary CBFs using the event-triggered control framework. The results are summarized in Table I, containing simulations for 30 vehicles with a Poisson traffic arrival process whose rate was set to 400 vehicles/hour. In the ordinary CBF case, the class K function is set to be linear in its argument: $\kappa_q = \kappa'_q . b_q(.)$ where $\kappa'_q = 0.1$. α is similar to as defined in (6). We can see the benefits of incorporating the trust metric into CBFs, as there is a mixture of low-trust and high-trust vehicles. As can be seen, integrating trust makes the CBFs less conservative reducing the average travel times of the CAVs in the network and increasing average acceleration, thus improving the throughput of the network. Finally, we notice that this also improves the average fuel consumption of the vehicles in the network.

TABLE I. Event-triggered control performance comparison with and without trust based CBF

	Item	CBF with trust	CBF without trust
$\alpha = 0.9$	Ave. Travel time	25	30.10
	Ave. $\frac{1}{2}u^2$	1.2	3.10
	Ave. Fuel consumption	17.73	18.50
$\alpha = 0.75$	Ave. Travel time	22.58	27.70
	Ave. $\frac{1}{2}u^2$	3.80	3.16
	Ave. Fuel consumption	17.36	18.55
$\alpha = 0.6$	Ave. Travel time	22.2	27.59
	Ave. $\frac{1}{2}u^2$	5.65	4.75
	Ave. Fuel consumption	17.49	18.65

Bias Injection Attack In order to highlight the robustness of our scheme against stealthy attacks and noise/estimation uncertainties we simulated an attack scenario by combining Sybil attack with BI attack. We compare our framework against the non-robust framework proposed in [19] and the results are shown in fig 4. As can be seen, the attack violates constraint (2) as shown in the plot of the constraint value (top left) which becomes negative due to the attack. This results in safety violation resulting in collision as shown in the image (on the left). On the other hand our proposed framework ensure safe coordination as can be verified from the plot and the image (on the right).

Mitigation: The ultimate goal of having mitigation in place is to avoid accidents and minimize the effects of attacks on the performance of the traffic network (i.e., average travel time, average energy consumption, and average fuel consumption). We present our empirical results in Fig. 2 by injecting different proportions of fake CAVs during the attack and for each scenario performing 5 runs whose average and standard deviation

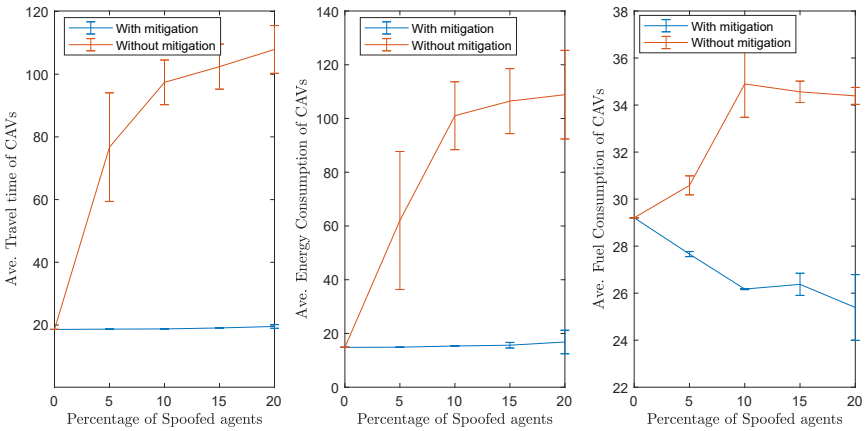


Fig. 2. The values of average travel time, average energy, and average fuel consumption for real CAVs for different proportions of fake CAVs over 5 runs with and without our proposed mitigation scheme.

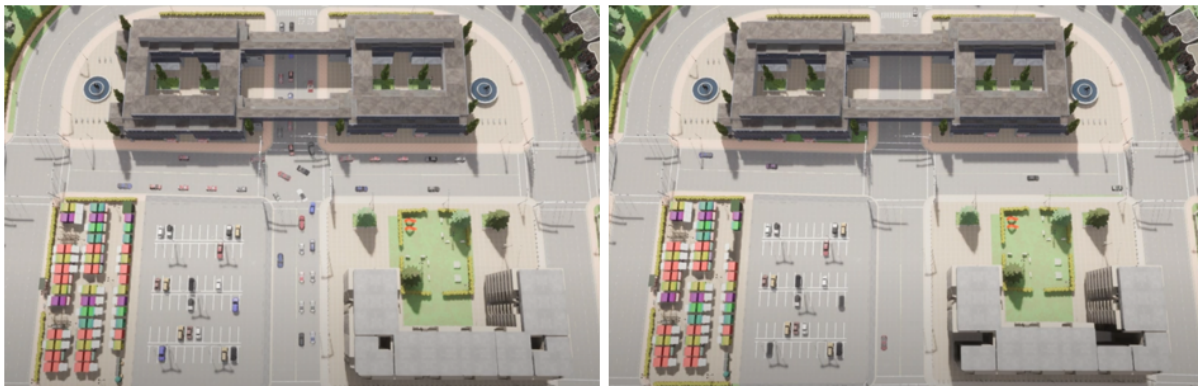


Fig. 3. The figure shows the performance of the network during a Sybil attack containing six spoofed CAVs without (left) and with (right) our proposed attack mitigation scheme. The picture was taken after 1 minute of running the simulation. The spoofed CAVs were located in three of the eight lanes.

are shown in the plots. We considered the strategic attacker model presented in [18]. It is important to note that this model assumes that the attacker has no access to the RSU. We varied the location of the spoofed CAVs, their initial states, and the proportion of spoofed CAVs across the runs. As can be seen, with our proposed mitigation scheme the average travel time was reduced to almost the same value as the scenario with no attack, thus validating the efficacy of the mitigation scheme in maintaining network performance. In addition, the average energy was also reduced to almost what it was without an attack. Moreover, we notice that the average fuel consumption improves with our proposed mitigation scheme.

Additionally, we provide a simulation scenario from CARLA during a Sybil attack in Fig. 3. The two figures show the network performance with and without our proposed mitigation scheme after 1 min. of starting the simulation. As can be seen, the absence of mitigation causes traffic holdup which is eased with our proposed mitigation scheme.

False positive case. As mentioned, our choice of trust framework does not result in false positive cases. However, as our proposed method is invariant to the specific choice of the trust framework, we conducted experiments to analyze scenarios when a real CAV gets falsely identified as spoofed due to a poorly chosen trust framework. We conducted our experiments for various degrees of accuracy of the onboard vision system. For each scenario, we ran 100 experiments and computed the percentage of safe scenarios, with results shown in Fig. 5. The experiments were run under different traffic conditions by varying the location of the falsely identified CAV (as spoofed) at the intersection for various values of states for the preceding vehicle(s). An experiment was deemed "safe" if there were no collisions between real CAVs upon triggering mitigation. Our experiments show that we can guarantee safety with $95\% - 99.96\%$ accuracy when the accuracy of the onboard object detection pipeline varies from $85\% - 95\%$.

VII. Conclusion

We have addressed Stealthy attacks, specifically Bias Injection attacks and Sybil attacks, on the cooperative control of a network of CAVs in a roadway with conflicting traffic. We propose a decentralized event-triggered control framework using robust trust-aware CBFs. Our proposed framework provides twofold benefits. Firstly, it guarantees provably safe coordination in the presence of adversarial attacks. Secondly, CBFs require choosing a class $\mathcal{K}$ function that inherently poses a trade-off between conservativeness and safety. We combine a trust metric associated to each CAV to balance this trade-off where the trust of each CAV is intended to reflect the

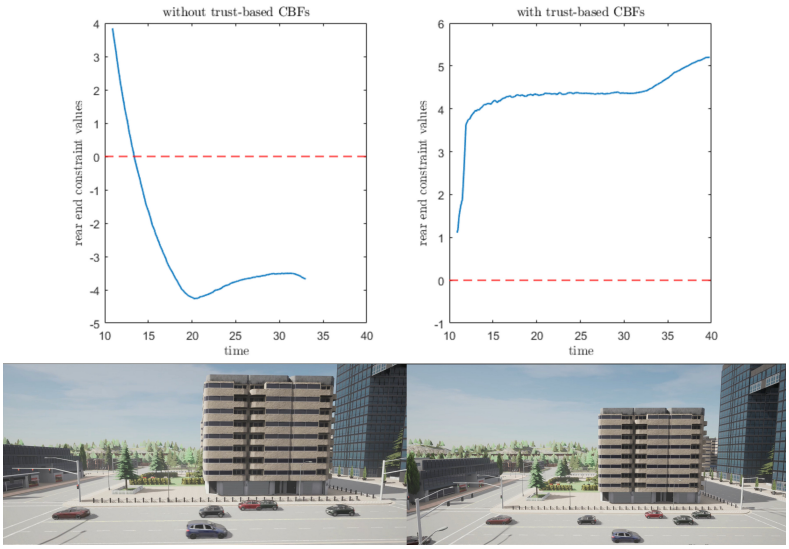


Fig. 4. Results illustrating the merit of our proposed robust trust-aware event-triggered control scheme. The result was generated by simulating an attack scenario combining BI attack with Sybil attack. As can be seen, the framework in [19] results in safety violation (left) which is prevented by our proposed robust trust-aware event- triggered control scheme. The images shown above are from CARLA simulations.

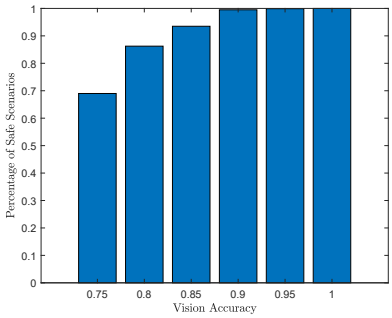


Fig. 5. Percentage of safe scenarios over 100 runs for different degrees of accuracy of the onboard vision system.

normalcy of a CAV. It is important to note that our proposed framework is invariant to the specific implementation of the trust framework. In addition, we propose a soft attack mitigation scheme to restore normal operation of the road network in the presence of attacks. Our proposed mitigation scheme can guarantee safety coordination against false positive cases. Our simulation results, obtained using the SUMO and CARLA simulators, highlight the merits of our proposed control and coordination scheme and validate their efficacy. In future work, we will extend our work by considering sensor attacks, in particular attacks on the Vision, Radar or LIDAR systems, along with attacks on in-vehicular networks.

REFERENCES

[1] D. W. L. Li and D. Yao, "A survey of traffic control with vehicular communications," *IEEE Trans. on Intelligent Trans. Sys*, vol. 15, no. 1, pp. pp. 425–432, 2013.

[2] D. de Waard, C. Dijksterhuis, and K. Brookhuis, "Merging into heavy motorway traffic by young and elderly drivers," *Accident Analysis & Prevention*, vol. 41, no. 3, pp. pp. 588–597, 2009.

[3] I. Kavalchuk, A. Kolbasov, K. Karpukhin, A. Terenchenko *et al.*, "The performance assessment of low-cost air pollution sensor in city and the prospect of the autonomous vehicle for air pollution reduction," in *IOP Conference Series: Materials Science and Engineering*, vol. 819, no. 1. IOP Publishing, 2020, p. 012018.

[4] V. A. van den Berg and E. T. Verhoef, "Autonomous cars and dynamic bottleneck congestion: The effects on capacity, value of time and preference heterogeneity," *Transportation Research Part B: Methodological*, vol. 94, pp. 43–60, 2016.

[5] R. M. Shukla and S. Sengupta, "Analysis and detection of outliers due to data falsification attacks in vehicular traffic prediction application," in *2018 9th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference*, 2018, pp. 688–694.

[6] X. Sun, F. R. Yu, and P. Zhang, "A survey on cyber-security of connected and autonomous vehicles (cavs)," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 7, pp. 6240–6259, 2022.

[7] M. Pham and K. Xiong, "A survey on security attacks and defense techniques for connected and autonomous vehicles," *Computers & Security*, vol. 109, p. 102269, 2021.

[8] H. Xu, S. Feng, Y. Zhang, and L. Li, "A grouping-based cooperative driving strategy for cavs merging problems," *IEEE Trans. on Vehicular Technology*, vol. 68, no. 6, pp. pp. 6125–6136, 2019.

[9] F. Xu and T. Shen, "Decentralized optimal merging control with optimization of energy consumption for connected hybrid electric vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 6, pp. 5539–5551, 2022.

[10] W. Xiao, C. G. Cassandras, and C. A. Belta, "Bridging the gap between optimal trajectory planning and safety-critical control with applications to autonomous vehicles," *Automatica*, vol. 129, p. 109592, 2021.

[11] P. Lu, L. Zhang, B. B. Park, and L. Feng, "Attack-resilient sensor fusion for cooperative adaptive cruise control," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, 2018, pp. 3955–3960.

[12] R. A. Biroon, P. Pisu, and Z. Abdollahi, "Real-time false data injection attack detection in connected vehicle systems with pde modeling," in *2020 American Control Conference (ACC)*, 2020, pp. 3267–3272.

[13] S. Boddupalli, A. S. Rao, and S. Ray, "Resilient cooperative adaptive cruise control for autonomous vehicles using machine learning," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–18, 2022.

[14] A. Alipour-Fanid, M. Dabaghchian, and K. Zeng, "Impact of jamming attacks on vehicular cooperative adaptive cruise control systems," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 11, pp. 12 679–12 693, 2020.

[15] F. Farivar, M. Sayad Haghighi, A. Jolfaei, and S. Wen, "On the security

of networked control systems in smart vehicle and its adaptive cruise control," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 6, pp. 3824–3831, 2021.

[16] A. Jarouf, N. Meskin, S. Al-Kuwari, M. Shakerpour, and C. G. Cassanderas, "Security analysis of merging control for connected and automated vehicles," in *2022 IEEE Intelligent Vehicles Symposium (IV)*, 2022, pp. 1739–1744.

[17] X. Zhao, A. Abdo, X. Liao, M. Barth, and G. Wu, "Evaluating cybersecurity risks of cooperative ramp merging in mixed traffic environments," *IEEE Intelligent Transportation Systems Magazine*, pp. 2–15, 2022.

[18] H. M. Sabbir Ahmad, E. Sabouni, W. Xiao, C. Cassandras, and W. Li, "Evaluations of cyber attacks on cooperative control of connected and autonomous vehicles at bottleneck points," in *2023 Network and Distributed System Security (NDSS) Symposium (Vehiclesec)*, 2023.

[19] H. M. S. Ahmad, E. Sabouni, W. Xiao, C. G. Cassandras, and W. Li, "Trust-aware resilient control and coordination of connected and automated vehicles," in *Proc. of 2023 IEEE International Intelligent Transportation Systems Conference*, 2023.

[20] Q. A. Chen, Y. Yin, Y. Feng, Z. Mao, and H. Liu, "Exposing congestion attack on emerging connected vehicle based traffic signal control," in *Proceedings 2018 Network and Distributed System Security Symposium*, 2018.

[21] M. Cheng, J. Zhang, S. Nazarian, J. Deshmukh, and P. Bogdan, "Trust-aware control for intelligent transportation systems," in *2021 IEEE Intelligent Vehicles Symposium (IV)*, 2021, pp. 377–384.

[22] H. Hu, R. Lu, Z. Zhang, and J. Shao, "Replace: a reliable trust-based platoon service recommendation scheme in vanet," *IEEE Transactions on Vehicular Technology*, vol. 66, pp. 1–1, 01 2016.

[23] H. Hu, R. Lu, and Z. Zhang, "Tpsq: Trust-based platoon service query via vehicular communications," *Peer-to-Peer Networking and Applications*, vol. 10, 01 2017.

[24] H. Parwana, A. Mustafa, and D. Panagou, "Trust-based rate-tunable control barrier functions for non-cooperative multi-agent systems," in *2022 IEEE 61st Conference on Decision and Control (CDC)*, 2022, pp. 2222–2229.

[25] K. Garlichs, A. Willecke, M. Wegner, and L. C. Wolf, "Trip: Misbehavior detection for dynamic platoons using trust," in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, 2019, pp. 455–460.

[26] Y. Shoukry, S. Mishra, Z. Luo, and S. Diggavi, "Sybil attack resilient traffic networks: A physics-based trust propagation approach," in *2018 ACM/IEEE 9th International Conference on Cyber-Physical Systems (ICCPS)*, 2018, pp. 43–54.

[27] H. Xu, C. G. Cassandras, L. Li, and Y. Zhang, "Comparison of cooperative driving strategies for cavs at signal-free intersections," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 7, pp. 7614–7627, 2021.

[28] H. Xu, W. Xiao, C. G. Cassandras, Y. Zhang, and L. Li, "A general framework for decentralized safe optimal control of connected and automated vehicles in multi-lane signal-free intersections," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 10, pp. 17 382–17 396, 2022.

[29] M. Cheng, C. Yin, J. Zhang, S. Nazarian, J. Deshmukh, and P. Bogdan, "A general trust framework for multi-agent systems," in *Proceedings of the 20th International Conference on Autonomous Agents and MultiAgent Systems*, 2021, p. 332–340.

[30] W. Xiao and C. Belta, "Control barrier functions for systems with high relative degree," in *Proc. of 58th IEEE Conference on Decision and Control*, Nice, France, 2019, pp. 474–479.

[31] E. Sabouni, H. S. Ahmad, W. Xiao, C. G. Cassandras, and W. Li, "Optimal control of connected automated vehicles with event-triggered control barrier functions: a test bed for safe optimal merging," in *2023 IEEE Conference on Control Technology and Applications (CCTA)*, 2023, pp. 321–326.

[32] W. Xiao, C. Belta, and C. G. Cassandras, "Event-triggered control for safety-critical systems with unknown dynamics," *IEEE Transactions on Automatic Control*, pp. 1–16, 2022.

[33] M. Kamal, M. Mukai, J. Murata, and T. Kawabe, "Model predictive control of vehicles on urban roads for improved fuel economy," *Control Systems Technology, IEEE Transactions on*, vol. 21, pp. 831–841, 05 2013.