Optimizing a **M**is**I**nformation and Mis**B**ehavior (MIB) Attack Targeting Connected Cars

Bruce DeBruhl
Computer Science and Software Engineering
California Polytechnical State University
San Luis Obispo, California
Email: bdebruhl@calpoly.edu

Patrick Tague
Electrical and Computer Engineering
Carnegie Mellon University
Pittsburgh, Pennsylvania
Email: tague@cmu.edu

***Abstract*—Autonomous driving features can mitigate traffic fatalities, create more enjoyable commutes, and increase fuel efficiency. For example, collaborative adaptive cruise control (or platooning) uses sensor-based distance measurement and vehicle-to-vehicle communications to automatically control inter-vehicle spacing. This can have tremendous benefits but is also safety critical. Therefore, it is essential to understand and mitigate potential platooning vulnerabilities.**

In this work, we design an attack that we call the insider *M*is*I*nformation and mis*B*ehavior (MIB) attack. During this attack, a malicious vehicle uses misinformation, erroneous V2V communications, and misbehavior, erratic driving, to cause predictable, dangerous, behavior. Although this attack can be applied broadly, we use it to design three optimal attacks were an attacker causes a collision without being damaged. Finally, we simulate these attacks and discuss trade-offs in there design parameters.

I. Introduction

Autonomous and semi-autonomous driving algorithms have gained increasing visibility in academic and industrial research [10], [6]. In general, these algorithms can increase passenger safety and reducing driver frustration. One of the most popular semi-autonomous driving features is adaptive cruise control (ACC). ACC uses sensor-based distance measurements to automatically regulate inter-vehicle spacing. ACC can even apply braking automatically to reduce accidents caused by inattentive drivers.

The minimum spacing between ACC-equipped vehicles is limited due to system lag. In particular, brake lag is the time it takes for a vehicle's braking mechanism to actuate after the corresponding control signal has been received. To overcome this limitation, collaborative adaptive cruise control (CACC or platooning) has been proposed. CACC uses vehicle to vehicle (V2V) communications to allow for feedforward control which is able to mitigate the impact of brake lag [12]. Since CACC can mitigate the effect of brake lag, it can even increase the density of vehicles on a road [8].

CACC is safety critical, so it is essential to understand vulnerabilities that can be exploited. However, there has been limited research on the impact of insider attacks against CACC controllers. Gerdes et al. [3] designed an insider attack that decreases the efficiency of vehicles using CACC by 20-30%. We have previously demonstrated trivial attack scenarios and develop a model-based detection scheme to detect attacks [1]. Lastly, Haas [5] explored a jamming attack against a CACC controller which could cause an accident.

In this work, we design a novel insider attack that we call the insider *M*is*I*nformation and mis*B*ehavior attack, or insider MIB attack. In this attack, the attacker uses erroneous V2V communications (misinformation) and erratic driving (Misbehavior) to cause predictable behavior in following vehicles. This attack can be used for multiple theoretic attacks. For example, we could consider a malicious logistics company attacking competitors [11], politically motivated attacks, insurance fraud [2], resource depletion attacks [3], or a general byzantine attack. Although some of these attacks may seem far-fetched, we argue that a priori understanding of any possible attack is helpful to develop secure algorithms.

For a specific demonstration of this attack, we develop a MIB attack that causes a collision between cars following the attacker without attacker being damaged. To design the attack, we formulate and solve a discrete-time optimization function with a defined target end-state [7]. To solve this optimization function we leverage an off-the-shelf optimization software [4]. The solution is used by a simulated attacker to mount an optimal insider MIB attack for a defined end-state.

We design three attack instantiations using this approach. First, the "follower attack" is mounted by the third car of a five car platoon and causes the fourth and fifth car to collide. Second, the "distant follower attack" is mounted by the third car in a seven car platoon and causes the sixth and seventh car to collide. An attacker mounting the distant follower attack has the advantage of being further away from the accident. Third, the "sandwich attack" is mounted by the third car in a six car platoon. This attacker causes the fourth, fifth, and sixth car to simultaneously collide. In these three attacks, the attacker accomplishes their goal without being involved in the accident. These attacks not only causes major damage, but would cause a non-linear chaotic event that would lead to unpredictability for other highway users.

To summarize, we make the following contributions.

- We introduce the insider MIB attack, where a malicious vehicle uses erroneous V2V communications (misinformation) and erratic driving (misbehaves) to cause predictable, undesired behavior.

- We develop a constrained, discrete-time, quadratic optimization problem to design a novel insider MIB attacks against platoons of cars.
- We demonstrate this attack with three instantiations that cause accidents without the attacker being involved.
- We discuss design parameter trade-offs.

II. Platooning Model

In this section, we describe our platoon model based on our previous work [1]. We consider a platoon of G cars which we refer to as C_0 through C_{G-1}. We assume that the lead car is C_0 and the remainder of cars are lined up incrementally in unchanging order [12], [9]. For all cars C_i, this allows us to assume that it is behind C_{i-1} and in front of C_{i+1}.

We notate the dynamics of car C_i including inter vehicle spacing, velocity, and acceleration as d_i, v_i, and a_i respectively. We also define $d_{r,i}$ as the desired reference distance for a vehicle. For the lead car, we define $d_0 = d_{r,0} = 0$. For all other vehicles, we define the desired reference distance using a constant headway policy [12]. In this policy, the reference distance is linearly proportionate to the vehicle's velocity with a constant offset. The constant offset is used to define a minimum distance between vehicles at standstill. We denote this as $d_{r,i} = h_{d,i}v_i + L_i$, where $h_{d,i}$ is the the headway and L_i is the constant offset. Without loss of generality, we assume $L_i = 0$. We can then define the error for the system as $e_i = d_i - d_{r,i}$.

We use the double integrator model [9], [1] for platoon dynamics. We include a lag constant η_i to model the delay between the control signal and actuation. Therefore, given an acceleration input u_i, car C_i has the following dynamics

$$\begin{aligned} \dot{a}_i &= -\eta_i^{-1}a_i + \eta_i^{-1}u_i \\ \dot{v}_i &= a_i \\ \dot{e}_i &= v_{i-1} - v_i - h_{d,i}a_i. \end{aligned} \tag{1}$$

Given the vehicle dynamics, we define a controller that uses local sensing and V2V communication to implement platooning (as illustrated in Figure 1). We use local-sensing to estimate the error e_i and its derivative for a proportional-derivative feedback back controller. We define this as

$$u_{fb,i} = k_p e_i + k_d \dot{e}_i, \tag{2}$$

where k_p and k_d are the proportional and derivative gain.

We use V2V communications, like DSRC or 5G, for feedforward control. We define our control input as

$$\dot{u}_{ff,i} = -\lambda^{-1}u_{ff,i} + \lambda^{-1}u_{ff,i-1}, \tag{3}$$

where $u_{ff,i-1}$ is the feedforward control signal received from the proceeding car. Assuming all cars are benign then $u_{ff,i-1} = u_{i-1}$ when packets are received and constant otherwise. We define λ as the update weight for our feedforward controller. We define our final controller as

$$u_i = u_{fb,i} + k_{ff,i}u_{ff_i} \tag{4}$$

where $k_{ff,i}$ is a tunable gain factor.

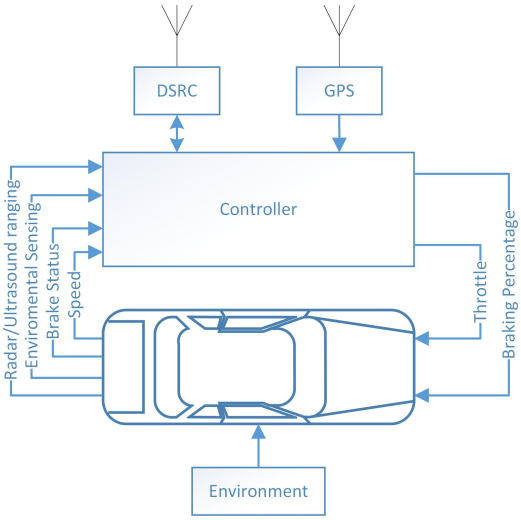

Fig. 1. In this figure, we illustrate a single platoon participant. In this work, we design an attack using misinformation and misbehavior to cause an accident.

To model our closed-loop system, we define C_i's state as a vector $x_i := [a_i, v_i, e_i u_{ff,i}]^T$. Assuming $\dot{e}_i = v_i - v_{i-1}$ we write our control input as

$$u_i = \begin{pmatrix} 0, k_d, k_p, k_{ff} \end{pmatrix} x_i + \begin{pmatrix} 0, -k_d, 0, 0 \end{pmatrix} x_{i-1}. \tag{5}$$

Substituting (5) into (1) and (3) we define our dynamics for car C_2 to C_{G-1} as

$$\dot{x}_i = A_{i,cl}x_i + A_{i-1,cl}x_{i-1} + A_{i-2,cl}x_{i-2} \tag{6}$$

where

$$A_{i,cl} = \begin{pmatrix} -\frac{1}{\eta_i} & \frac{k_d}{\eta_i} & \frac{k_p}{\eta_i} & \frac{k_{ff}}{\eta_i} \\ 1 & 0 & 0 & 0 \\ -h_{d,i} & -1 & 0 & 0 \\ 0 & 0 & 0 & -\frac{1}{\lambda} \end{pmatrix} \tag{7}$$

$$A_{i-1,cl} = \begin{pmatrix} 0 & -\frac{k_d}{\eta_i} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & \frac{k_d}{\lambda} & \frac{k_p}{\lambda} & \frac{k_{ff}}{\lambda} \end{pmatrix} \tag{8}$$

$$A_{i-2,cl} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & -\frac{k_d}{\lambda} & 0 & 0 \end{pmatrix}. \tag{9}$$

We previously demonstrated the behavior of this system [1].

III. Insider MIB Attack

In this section, we define an insider MIB attack. We assume that the lead car in the platoon does not change speed. With this assumption the state update equations for car C_0 and C_1 are defined as

$$\begin{aligned} \dot{x}_0 &= 0_4 x_0 \\ \dot{x}_1 &= 0_4 x_1 \end{aligned} \tag{10}$$

where 0_4 is a 4x4 zero matrix.

For demonstration, we define C_2 as the inside attacker. We define two control inputs for the attacker. First, misbehavior is defined by the attacker controlling there acceleration u_a such that

$$\dot{a}_a = -\frac{a_a}{\eta_a} + \frac{u_a}{\eta_a}. \tag{11}$$

Secondly, the attacker controls misinformation with the feed-forward signal $u_{ff,a}$. This is sent to the car directly following the attacker, which we call the victim and as defined as

$$\dot{u}_{ff,v} = -\frac{u_{ff,v}}{\lambda} + \frac{u_{ff,a}}{\lambda} \tag{12}$$

We then define the attackers state update equations as

$$\dot{x}_a = A_a x_a + A_{a-1} x_{a-1} + B_a u_a \tag{13}$$

where

$$A_a = \begin{pmatrix} -\frac{1}{\eta_i} & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ -h_{d,i} & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$A_{a-1} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad B_a = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Similarly, we define the victims state vector update equations as

$$\dot{x}_v = A_v x_v + A_{v-1} x_{v-1} + B_v u_{ff,a} \tag{14}$$

where

$$A_v = A_{i,cl}$$

$$A_{v-1} = \begin{pmatrix} 0 & -\frac{k_d}{\eta_i} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad B_v = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \frac{1}{\lambda} \end{pmatrix}.$$

We then define the state of the whole system as

$$X_a = \left(x_0, x_1, x_a, x_v, \ldots, x_{K-1}\right)^T \tag{15}$$

and the misinformation and misbehavior (MIB) control input as

$$U_a = \left(0, 0, u_a, u_{ff,a}, 0, \ldots\right). \tag{16}$$

Therefore, the attack system update equation is

$$\dot{X}_a = AX_a + BU_a \tag{17}$$

where

$$A = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & \ldots \\ 0 & 0 & 0 & 0 & 0 & 0 & \ldots \\ 0 & A_{a-1} & A_a & 0 & 0 & 0 & \ldots \\ 0 & 0 & A_{v-1} & A_v & 0 & 0 & \ldots \\ 0 & 0 & A_{i-2,cl} & A_{i-1,cl} & A_{i,cl} & 0 & \ldots \\ 0 & 0 & 0 & A_{i-2,cl} & A_{i-1,cl} & A_{i,cl} & \ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

and

$$B = \left(0, 0, B_a, B_v, 0, 0, 0, \ldots\right)^T.$$

Based on DSRC's transmission time, we discretize this system with a sampling rate of 100ms to arrive at

$$X_a[k+1] = A_{ad} X_a[k] + B_{ad} U_a[k]. \tag{18}$$

In the insider MIB attack, the attacker chooses values for U_a to design their attack.

IV. Optimal MIB Attack

In this section, we use equation (18) to derive a discrete-time, constrained, quadratic optimization function that can be solved to find attack inputs for an insider MIB attack. To make the model realistic, we bound the velocity and acceleration for each vehicle as

$$0\frac{m}{s} \leq v_i \leq 44.7\frac{m}{s} \;\forall\, C_i \tag{19}$$

$$-8.94\frac{m}{s^2} \leq a_i \leq 6.39\frac{m}{s^2} \;\forall\, C_i. \tag{20}$$

We define the initial state of the system for all vehicles as $a_i[0] = 0\frac{m}{s^2}$, $v_i[0] = 25\frac{m}{s}$, and $e_i[0] = 0m$.

We design an optimal MIB attack to reach a desired end state while minimizing deviation from nominal. We use a quadratic cost function defined as

$$J[X_a, U_a] = (X_a[N] - X_r)' Q (X_a[N] - X_r) + \sum_{i=1}^{N-1} (X_a[i] - X_r)' Q (X_a[i] - X_r) + U_a[i]' R U_a[i] \tag{21}$$

where X_r is the nominal state of the system. The matrices Q and R are optimization weights which we represent by the identity matrix.

We use the cost function to define our optimization as

$$\begin{aligned} \underset{U_a}{\text{minimize}} \quad & J[X_a, U_a] \\ \text{subject to} \quad & X_a[i+1] = AX_a[i] + BU_a[i] \;\forall\, i \in [0, N-1] \\ & L_b \leq S_1 X_a[i] \leq G_b \;\forall\, i \in [0, N] \\ & \psi_d = S_2 X_a[N] \end{aligned} \tag{22}$$

where L_b and G_b are the lower and upper bounds for the system state. The lower bounds include minimum acceleration, velocity, and allowable error. We refer to the minimum allowable error as the safety distance which must be selected to ensure that no vehicles collides prior to the desired end state. The upper bounds include maximum acceleration, velocity, and allowable error. The matrix S_1 is used to select the states that must be bound. We define ψ_d as the desired end state for the system. By defining the end state carefully an attacker can cause following vehicles to collide in controlled ways. Similar to above S_2 is used to select the end state variable that are important.

The above constrained optimization problem can be solved with various analytical and empirical tools. In this work, we leverage the efficiency and ease of use of the Gurobi commercial solver [4].

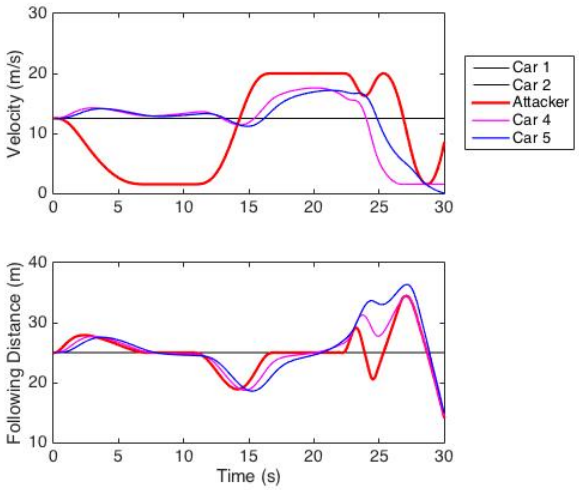

Fig. 2. In this figure, we show the results for a follower attack mounted by the third car in a 5 car platoon. The attack causes car 4 and 5 to collide at 30 seconds with a velocity of 15 $\frac{m}{s}$. The attacker never gets closer than 2.5 meters to any other car in the platoon. A video of the simulation of this attack is available at goo.gl/B7w9Bg.

V. Attacks and Simulations

In this section, we solve (22) with various values of ψ_d to demonstrate potential MIB attacks against platoons. We design three attacks, solve them with the Gurobi optimizer, and then simulate them in Matlab. In the first attack, the follower attack, we optimize the attacker's misbehavior and misinformation to cause the two cars directly following the attacker to collide. The attacker is able to accomplish this without being involved in the accident and while guaranteeing a minimum distance from the colliding vehicles. In the second attack, the distant follower attack, we optimize the attacker's misbehavior and misinformation to cause two cars not directly following the attacker to collide. Lastly, in the sandwich attack, the attacker uses misinformation and misbehavior to cause the three cars directly following it to simultaneously collide. These attacks are all clearly dangerous for platoons of vehicles and demonstrate the need for appropriate cyber-physical intrusion detection techniques.

A. Follower Attack

In the follower attack, the attacker causes a collision between the two cars directly following her at time T while not being involved in the collision. We define ψ_d such that the distance between the victim car C_v and the car following the victim C_{v+1} is zero at time T and their velocity is v_a.

In Figure 2 we show our simulation of this attack in a five car platoon. This attack occurs at 30 seconds with the victim's end velocity at $15\frac{m}{s}$ and attacker never getting closer than 2.5 meters to the vehicle in in front of it. This allows the attacker to use misbehavior and misinformation to cause a high speed accident without being involved. A video of the attack simulation can be found at http://goo.gl/YdpcaZ. In the experiment we found that the attack time, safety distance, maximum attacker's error, and attack velocity v_a must be selected carefully to guarantee a feasible solution.

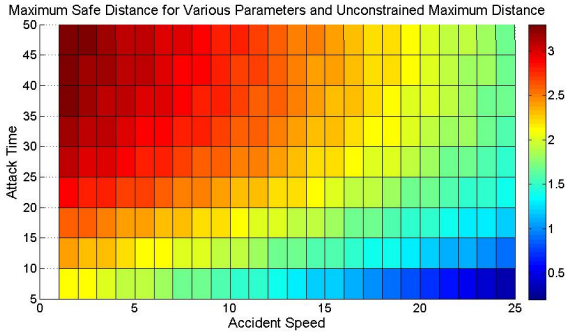

Fig. 3. In this figure, we show the front between maximum safety distance, attack time, and accident speed at collision when the attacker does not have a bound on their maximum following distance.

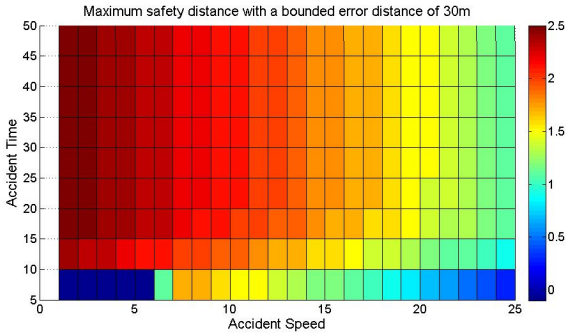

Fig. 4. In this figure, we show the front between maximum safety distance, attack time, and accident speed at collision when the attacker has a bound of 30m of seperation from the platoon.

We demonstrate the feasible reachable space of the follower attack by modeling the trade-offs in parameters. We first consider an attacker that has an unconstrained maximum error (e_a). We then calculate how close an attacker has to get to an accident under various attack times and victim's speed. We use these measurements to develop the heat map shown in Figure 3. In general, the slower the victim's speed and higher the attack time the further the attacker can stay from an accident. It is interesting to note, with many of these parameter combinations the attacker effectively creating a second platoon by selecting an extremely high following distance. In order to limit this effect, we rerun this experiment with a constraint on the attacker's maximum separation from the platoon. In Figure 4, we run the same experiment as above with a maximum following distance for the attacker of 30 meters. In both of these figures, it is clear that the maximum attacker safety distance decreases as the attack speed increases. Their is also some increase with accident time but this effect is less pronounced.

B. Distant follower attack

The distant follower attack if very similar to the follower attack. However, in the distance follower attack the attacker uses misbehavior and misinformation to cause a collision

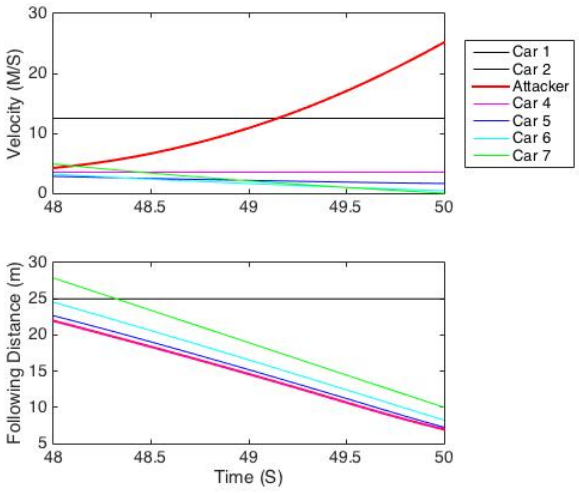

Fig. 5. In this figure, we show the last two seconds of the distant follower attack against a seven car platoon. In the video at https://goo.gl/om5jx6 We simulate this attack against a 6 car platoon.

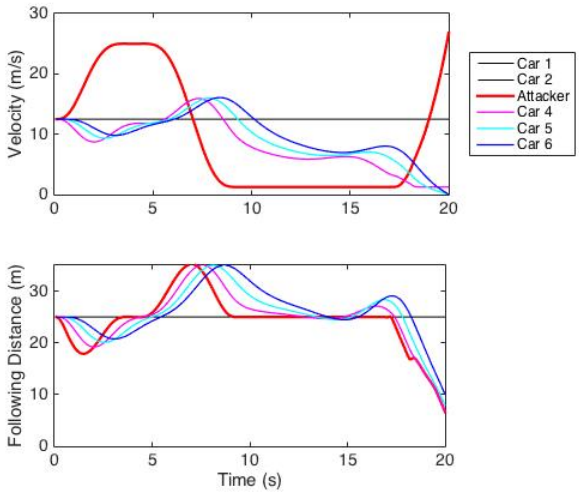

Fig. 6. We demonstrate an attack in a 6 car system where the third car causes a simultaneous double collision between cars 4, 5, and 6. We provide a video simulating this attack at https://goo.gl/Mpc8ZQ.

between two cars following her that are not directly behind her. For example, if the attacker is the third car in the platoon she may cause a collision between the fifth and sixth car. This is beneficial for the attacker because it creates a buffer of another car between them and the collision. We define a time T when the collision occurs and the speed the victim vehicle must be traveling. The attack parameters must be selected considering the tradeoff in their performance. We define one additional constraint parameter, the non-attacker safety distance, which is used to bound how close a car that is not targeted gets to the accident. Using a low non-attacker safety distance increases the risk for other non-target attack. However, using a lower non-attacker safety distance allows us to define a higher attacker safety distance. In Figure 5, we show the velocity and following distance of a platoon of vehicles during the last two seconds of a distant follower attack mounted by the third car in a seven car platoon. In this case, the sixth and seventh car collide at $10\frac{m}{s}$ while the attacker never gets closer then 3 meters to any other vehicle.

C. Sandwich attack

We call the last attack we develop the sandwich attack. In this attack, the attacker causes three following cars to simultaneously collide. This attack also has reachability tradeoffs when selecting constraint parameters as discussed in the previous two attacks. In Figure 6, we simulate this attack mounted by the third car which causes the fourth, fifth, and sixth car to simultaneously collide. This attack damages all the cars, but car 5 would be especially injured. A video of this simulation can be found at http://goo.gl/Mpc8ZQ.

VI. Conclusion

In this paper, we design and demonstrated the insider MIB attack against a CACC algorithm. To demonstrate it, we formulate a discrete-time, constrained, quadratic optimization problem that allows an optimal attack to drive the system to a desired end state. We show that potential end states allow for the attacker to cause a collision while not being involved. We demonstrate three examples of this optimal attack and simulate them. These types of attacks, amongst others, are critical to consider in future platoon designs.

References

[1] Bruce DeBruhl, Sean Weerakkody, Bruno Sinopoli, and Patrick Tague. Is your commute driving you crazy?: a study of misbehavior in vehicular platoons. In *Proceedings of the 8th ACM Conference on Security and Privacy in Wireless and Mobile Networks*, page 22. ACM, 2015.

[2] ESURANCE. Car insurance fraud: Staged car accidents. *Online at https://www.esurance.com/info/car/staged-car-accidents*.

[3] Ryan M Gerdes, Chris Winstead, and Kevin Heaslip. Cps: an efficiency-motivated attack against autonomous vehicular transportation. In *Proceedings of the 29th Annual Computer Security Applications Conference*, pages 99–108. ACM, 2013.

[4] Inc. Gurobi Optimization. Gurobi optimizer reference manual, 2015.

[5] Jason J Haas. The effects of wireless jamming on vehicle platooning, 2009.

[6] Aharon Bar Hillel, Ronen Lerner, Dan Levi, and Guy Raz. Recent progress in road and lane detection: a survey. *Machine Vision and Applications*, 25(3):727–745, 2014.

[7] Donald E Kirk. *Optimal control theory: an introduction*. Courier Corporation, 2012.

[8] Michael P Lammert, Adam Duran, Jeremy Diez, Kevin Burton, and Alex Nicholson. Effect of platooning on fuel consumption of class 8 vehicles over a range of speeds, following distances, and mass. Technical report, SAE Technical Paper, 2014.

[9] Fu Lin, Makan Fardad, and Mihailo R Jovanovic. Optimal control of vehicular formations with nearest neighbor interactions. *Automatic Control, IEEE Transactions on*, 57(9):2203–2218, 2012.

[10] Andreas Mogelmose, Mohan M Trivedi, and Thomas B Moeslund. Vision-based traffic sign detection and analysis for intelligent driver assistance systems: Perspectives and survey. *Intelligent Transportation Systems, IEEE Transactions on*, 13(4):1484–1497, 2012.

[11] Jonathon Saul. Global shipping feels fallout from maersk cyber attack. *Reuters*.

[12] DVAHG Swaroop. String stability of interconnected systems: An application to platooning in automated highway systems. *California Partners for Advanced Transit and Highways (PATH)*, 1997.