IEEE/ASME TRANSACTIONS ON MECHATRONICS
https://ieeexplore.ieee.org/document/9373943
DOI: 10.1109/TMECH.2021.3064816

Covert Attacks through Adversarial Learning: Study of Lane Keeping Attacks on the Safety of Autonomous Vehicles

Faezeh Farivar, *Senior Member, IEEE*, Mohammad Sayad Haghighi, *Senior Member, IEEE*, Alireza Jolfaei, *Senior Member, IEEE*, Sheng Wen, *Member, IEEE*

***Abstract*—Road management systems are to improve in terms of integrity, mobility, sustainability and safety by the adoption artificial intelligence and Internet of Things services. This paper introduces the concept of covert attacks on autonomous vehicles which can jeopardize the safety of passengers. Covert attacks are designed to manipulate the output of a cyber physical system through network channels in a way that while the changes are not easily noticeable by human beings, the system is negatively affected in the long run. We argue that future smart vehicles are prone to worms of such kind which can use adversarial learning methods to adapt themselves to hosts and remain stealth for a long period. As a case study, we design and launch a covert attack on the lane keeping system of autonomous vehicles. In the studied scenario, an intelligent adversary manipulates sensor readings (lane position, curvature, etc.) in order to deceive the controller to drive the vehicle closer to the boundaries. The worm/attacker interactively learns the host vehicle behaviors in terms of lateral deviation and maneuverability and tries to increase the errors to the extent that remains unnoticeable to the driver. This process is carried out by using actor-critic learning based on the Newton-Raphson method in the sample studied scenario. We additionally show how an intrusion detection system can be designed for such covert attacks to alert the driver. We use the GPS data as well as offline maps to reconstruct the road curves and match it against the readings in the case study. A simulation testbed is developed based on the map of Nurburgring-Grand Prix track to evaluate the developed models. Results confirm the validity and effectiveness of the proposed models.**

***Index Terms*—Unmanned autonomous systems, Cyber physical systems, Fault diagnosis & prognosis, IoT in industry, Intelligent control, Lane Keeping, Covert Attack, Adversarial Machine Learning, Security, Vehicle Safety.**

I. INTRODUCTION

AUTONOMOUS driving has attracted a great deal of attention from car manufacturers, drivers, and several technology developers. Nowadays, for standard passenger vehicles, Adaptive Cruise Control (ACC), Lane Keeping Assistance, Lane Departure Warning and Collision Detection are not considered fancy additions anymore. In addition to the above, autonomous vehicles work based on numerous other technologies such as stereo cameras, radar, GPS and artificial intelligence (AI). Most advanced cars employ control in simple scenarios. For example, STOP and GO ACC of Audi enables the car to follow other cars at low speeds, even in dense traffic. Lane departure prevention system of Mercedes-Benz and autopilot of Tesla integrate ACC with auto-steering to attain a certain level of autonomous driving [1].

On the other hand, road management systems and vehicle-to-vehicle communication are becoming more efficient towards having smarter transportation [2]. Integrating smart technologies, employing AI and intelligent control systems as well as Cyber Physical Systems (CPS) and Internet of Things (IoT) are making road mobility systems safer and more sustainable.

Since smart vehicles are equipped with more sensors and network connectivities, the issue of cyber threats is getting more serious too. Security attacks launched on autonomous vehicles can not only cause damage to road mobility systems, but also endanger the lives of human beings.

Nowadays, hackers are more into CPS targets. Stuxnet experience showed cyber attacks can have real life impacts. Hackers are also targeting smart vehicles now. In July 2015, Fiat Chrysler recalled 1.4 million cars after two security experts showed how to hack a Jeep Cherokee remotely through e.g. the car's entertainment system which was presumably connected to the data network [3]. These incidents drew the attention of academics to the security of smart vehicles [4].

Reference [5] has categorized CPS threats and defined a new family which is so called the *covert attacks*. Covert attacks are part of the deception attacks family, in which an attacker takes control of parts of the CPS, but keeps the attack impact small so that it does not draw the attention of human observers. There are two main subtypes of covert attacks. The first kind tries to cause steady state errors on the output of the system to, for example, degrade or exhaust the system over the long run. The second kind aims to (repeatedly) create transient noises/effects on the system output to do the same job.

In this study, we take the concept of covert attacks to the realm of autonomous vehicles. After introducing the general idea, as an example, lane keeping (LK) in autonomous driving will be attacked. In our LK scenario, road boundaries are detected by a stereo camera and are fed to an adaptive model predictive control (MPC) system to handle the steering wheel

This work was in part supported by a grant from the Institute for Research in Fundamental Sciences (IPM).

F. Farivar is with the Department of Computer and Mechatronics Engineering, Science and Research Branch, Islamic Azad University, Tehran 1477893855, Iran, and also with the School of Computer Science, Institute for Research in Fundamental Sciences (IPM), Tehran 19395-5746, Iran, Email: f.farivar@srbiau.ac.ir.

M. Sayad Haghighi is with the School of Electrical and Computer Engineering, College of Engineering, University of Tehran, Iran, Email: sayad@ut.ac.ir (corresponding author).

A. Jolfaei is with the Department of Computing, Macquarie University, Sydney, NSW 2109, Australia , Email: alireza.jolfaei@mq.edu.au.

S. Wen is with the School of Software and Electrical Engineering, Swinburne University of Technology, Australia, Email: swen@swin.edu.au.

angle. By adopting the adversarial machine learning concept, we argue that it is possible to compromise the communication channel connecting the sensors and the controller and manipulate the control inputs (sensor readings) in a way that the controller decisions are changed in the attacker's favor. We conduct an experiment in which the attacker slightly displaces the detected lanes though intelligent and adaptive modification of camera outputs. The adversary learns the probability of density function (pdf) of lateral deviations from the road center line and exerts the maximum possible strength of attack that remains covert to the passengers/driver but statistically increases the chance of having accidents. The attack strength is tuned by an actor-critic learning system which uses Newton-Raphson method. The learning attack example does not depend on the road profile or vehicle properties and can adapt itself to different car/road conditions.

After that, we show how an intrusion detection system (IDS) can be designed for such stealthy attacks to alert the driver in a timely manner. In the case of LK attack, the IDS works based on reconstruction of road curves by using GPS data and comparing them against offline (navigational) map of roads.

To evaluate both the attack and detection strategies, the proposed idea is put to test in a road simulator that we have designed based on Nurburgring-Grand Prix track. The simulator has been developed under MATLAB environment. The results confirm that the proposed covert attack can be successful in penetrating the system of autonomous vehicles and that the proposed IDS can detect such covert behavioral changes. The main contributions of this paper can be summarized as follows:

- Introducing the concept of covert/stealth cyber attacks on Intelligent Transportation Systems (ITS) and more specifically, on autonomous vehicles.
- Designing a novel covert attack for lane keeping systems based on the notions of adversarial learning.
- Designing an IDS for covert attacks based on behavioral pattern changes and sensor inconsistencies. For the LK attack, this is done by reconstruction of road curves and relying on local information.
- Road testing the attack and detection strategies on Nurburgring-Grand Prix track by using a simulator.

The rest of this paper is organized as follows: In Section II, related studies are reviewed. In Section III, the proposed covert attack framework on ITSs is explained. In Section IV, a covert attack on LK and the corresponding intrusion detection system are described. Moreover, the actor-critic learning model of an intelligent adversary is presented. Section V presents the experiments and discusses the simulation results. The paper is concluded in Section VI.

II. RELATED WORK

We divide the related studies into two categories; the first one is on the lane keeping problem in autonomous vehicles and the second is on its security.

There are several dynamics like longitudinal and lateral speeds, yaw angle error, and yaw rate which help to properly steer the vehicle for lane keeping. For example, reference [6] studied the lateral and longitudinal stability for a distributed drive electric vehicle using MPC, or [7] proposed a planning algorithm for safe lane changing of autonomous vehicles on a straight multi-lane driveway.

Many control strategies have been developed for LK systems, e.g. robust control of vehicles for lateral motion regulation in the presence of uncertainties [8]. In addition, there are many studies on assistive LK systems which can switch between driver and LK control system. But among these, MPC has been used more in autonomous vehicles [6, 9]. It has a history of use in other control applications too [10].

In [11], the authors study autonomous control of a car whose lateral and longitudinal controllers are treated (almost) separately. Longitudinal speed and brake control are executed by a feedback PID loop around speed error plus a feedforward term based on a time-averaged car pitch.

All the above studies develop control systems to maintain safety and sustainability of vehicles in the absence of cyber attacks. However, with all the digital connections a smart vehicle makes, it is unreasonable to rule out the possibility of security breaches. Paper [12] classified the attacks and defenses in autonomous vehicles. Autonomous vehicles are prone to cyber attacks because of two reasons. First, the external communication channels between autonomous vehicles and outside environment are increased. Examples are inter-vehicular communications via vehicular ad-hoc networks (VANET) or vehicle to infrastructure communications. Therefore, the vehicle network might be compromised by an adversary which can lead to untrusted data injection and decease of travel safety. Second, the interconnections of electronic control units and other components via in-vehicle communication channels like the controller area network (CAN) bus.

Paper [13, 14] presents several types of attacks which may disturb the operation of VANETs. Two groups of attacks launched by internal malicious vehicles and external entities are introduced. The VANET attacks studied include message spoofing and replay attacks, integrity and impersonation attacks, denial of service (DoS) attacks, and De-anonymization attack [15].

Reference [16] presumed that data streams emanate from vehicles and go to side units on the road. A group structure is suggested to interact with the leader, and the road side unit. A lightweight permutation mechanism retains the confidentiality and privacy of the sensory data.

In [17], two covert attacks on adaptive cruise control systems are proposed. The first one is on the acceleration of the ego car, and the other manipulates the reference signal. Both increase the risk of accidental crash. Moreover, an IDS, based on artificial neural networks, is introduced for anomaly detection. A similar work was done on a DC motor as the plant in [18]. The IDS detected significant deviations from the normal behavior learned during the safe period at the beginning of system operation.

In [19], a resilient attitude tracking control strategy is proposed for aviation systems when the network communication is under DoS attack. In combination with two types of flows or jumps and a general explicit characterization of frequency/duration properties for DoS attacks, the authors develop a hybrid formalism for attitude tracking control.

In [20], a DoS attack detection algorithm for VANETs is proposed. Malicious and irrelevant nodes were detected and isolated from the routing network. In [21], an intrusion detection algorithm is provided with a feature extraction method and a hierarchical classifier for malicious activity detection in VANETs. The two key features that are extracted are the ”differences in traffic flow” and the ”position”. The classifier operation involves the processes of relabeling and recalculation. Reference [22] used multi-stage attention-based convolutional neural networks to detect anomalies and attacks in automated vehicles.

In [23, 24], an intelligent classic control system is presented for compensation of scalar attacks and faults on nonlinear cyber physical systems which are prone to attacks or faults in the forward link. More specifically, in [23], the method is tested on a car cruise control system under attack. The designed control system is a combination of sliding mode control, which is a robust control employed in many studies, as well as a Gaussian RBF neural network [25] for estimation of the attack effect.

In addition, designing anomaly/intrusion detection systems is another issue in autonomous vehicles. Different types of IDSs for this aim are developed using e.g. rolling window and residual detectors based on Kalman filter for real-time attacks [26], deep learning techniques [27, 28], etc.

III. THE PROPOSED COVERT ATTACK FRAMEWORK FOR ITS

In this section, we present how covert attacks through adversarial learning can intrude in autonomous vehicles. The proposed framework for ITS covert attacks is depicted in Fig. 1. As illustrated in this figure, the attacker takes control of the feedback link, either by constantly sitting on the line or planting a malicious code (e.g. worm) in the system. Either way, the adversary can manipulate the sensory data gathered by the output sensors of the system/plant. Hence, the feedback information is tampered with before is received by the control system. The purpose of the attacker is to deceive the controller, though the controller tries to decrease errors to achieve the control objectives for the plant. But the controller inputs are fabricated or manipulated. Therefore, the control effort will not necessarily serve the original control purpose in the closed loop system, but rather serves the adversary’s purpose.

To explain the proposed idea further, we choose an example in smart vehicles, i.e. lane keeping (LK) control system. In the LK scenario, the goal is to drive the vehicle between a specified pair of lines, inside lane boundaries. A controller works to provide the steering command input to the autonomous driving system. Stereo cameras and lane sensors are employed for data collection. In the attack scenario, the adversary make changes to the sensory data or the attacker-planted malicious module manipulates signals such that the resulted variations remain unnoticeable to the driver.

In the context of LK, attack can be translated into excessive deviations from the lane center, to the extent that the vehicle enters the next lanes, or partially goes off the track. In the stealth case, this should not lead to uncontrollability and the amount of lateral deviation should be controlled so that the effect remains covert. One way to achieve this is to learn the probability density function of lateral deviations (from the road center line) and statistically increasing its mean value so that the tail of the pdf enters unsafe zones. This increases the chance of having accidents, but the amount of these excessive deviations must be small so that it can be considered a covert attack. This can be done by a generic learning worm/malicious code which does not rely on the vehicle model/manufacturer specifications and not make prior assumptions about the road profile either. It can adapt itself to different car/road conditions. More details on such an attack are presented in the case study of the next section.

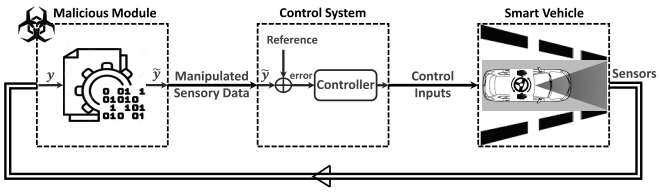


Fig. 1. The proposed covert attack scenario in which an attacker-planted malicious module manipulates the control loop.

IV. COVERT ATTACK ON LANE KEEPING BY ADVERSARIAL LEARNING

In this section, we discuss a case study and explain how a covert attack can be launched on the LK system as an example. Moreover, we show how such covert attacks can be detected.

A. Autonomous Lane Keeping System

A lane keeping control (LKC) system in autonomous vehicles provides safe travel between two lines, which usually form a highway lane. Upon vehicle departure from the marked lane, steering is automatically adjusted by the LKC system to maintain a safe travel within the lane boundaries. In LKC system, there are no steering command inputs from the driver and the vehicle is deemed autonomous in the lane keeping task. However, the assumption is that the driver can intervene and take the control back if he/she finds the situation hazardous. The vehicle detects boundaries of the lane and the road curvature. The LKC system works based on several inputs such as road curvature, car lateral deviation, relative yaw angle, etc. Since these measurements are made by stereo cameras as well as lane sensors, the system works under uncertain conditions caused by e.g. noise, missing, incomplete, or inaccurate readings. Hence, the LKC system should be robust to uncertainties. In this condition, the main controller uses estimations to make its decision. Lateral dynamics of the vehicle are lateral deviation (e_1) and relative yaw angle (e_2), as shown in Fig. 2. The first is the lateral deviation from the center line of the lane, and the second is the relative yaw that is the orientation error with respect to the road. Dynamic errors of lateral dynamic are as follows:

$$\begin{aligned} \dot{e}_1 &= v_x e_2 + v_y \\ \dot{e}_2 &= \dot{\psi} - \kappa v_x \end{aligned} \tag{1}$$

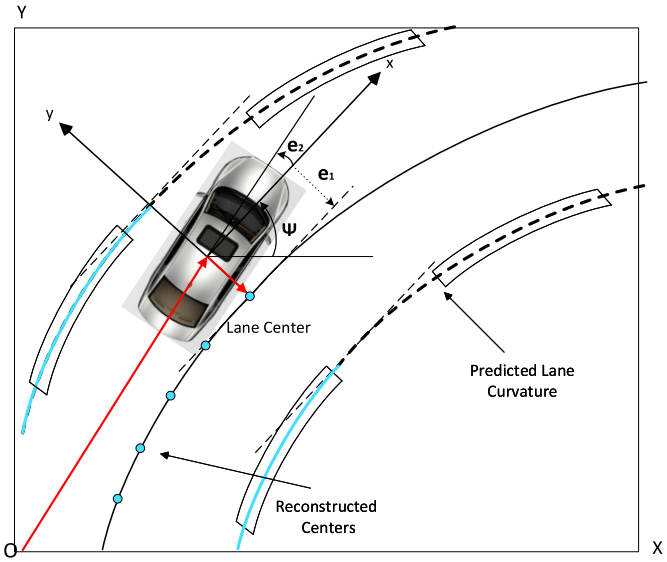


Fig. 2. Lateral dynamics of the vehicle in lane keeping scenario, e_1 is the lateral deviation and e_2 is the relative yaw angle.

where v_x and v_y are vehicle longitudinal and lateral velocities. $\dot{\psi}$ is the yaw angle rate and κ is the road curvature.

B. Covert Attack on LK by an Intelligent Adversary

In this case study, we assume that the adversary penetrates the vehicle sensors/systems or remotely manages to manipulate the measurements. Hence, the main controller is deceived by receiving incorrect signals. The goal of the covert LK attacker is to manipulate the signals in a way that the the probability of deviating from the lane is increased to the extent that remains hidden from a general human observer's point of view but statistically increases the chance of having crashes and accidents. The attacker can do this by remotely hacking/intruding into the car digital systems (e.g. by gaining access to the CAN bus) or spreading a worm that intrudes into smart vehicles automatically and plants itself where accessing the sensor signals is possible.

Notice that since each vehicle has different characteristics (weight, tire friction, etc.), correct manipulation of signals shall be learned for each vehicle individually. This is very much like adversarial machine learning in which the input to the system (normally classifier) is manipulated (by learning) so that a wrong decision is made by the system. To realize one instance of LK covert attacks, the malicious entity can misreport the lateral deviation (e_1) from the lane detection sensors. This is one of the key inputs to the MPC. We propose the adversarial change is made dynamically as:

$$\tilde{e}_1(t) = e_1(t) - \alpha\ \kappa(t) \tag{2}$$

where $\kappa(t)$ is the curvature value at t and α is the strength of attack which is to be tuned/learned based on vehicle characteristics. Both $\kappa(t)$ and $e_1(t)$ are available to (and possibly modifiable by) the intruder or worm as they are derived from lane sensor measurements. There is no shift for the signal if $\kappa(t)$ is zero, since such manipulations are easily detectable by the driver on a straight line. The adversary or the implanted worm learns to adjust α so that while remaining in the stealth zone, the chance of accidents is increased. As a result, the expectation of lateral deviation values is increased, and more technically speaking, Eq. (2) makes the e_1 pdf tail grow longer.

Assume that there is no attack and the adversary monitors the signals. Over the time (before launching the attack), the adversary learns the pdf of lateral deviations, as symbolically shown in Fig. 3a. This is the pdf of lateral deviations over many turns. The attacker's goal is to increase the probability that these deviations go beyond the safe threshold (σ_T) to a a certain amount, as depicted in Fig. 3b. The probability of experiencing excessive deviations is denoted by P_E.

$$P_E(\sigma_T) = \int_{\sigma_T}^{\infty} f_{e_1}(e)\ de \tag{3}$$

Here, $f_{e_1}(e)$ is the pdf of lateral deviations absolute values which is a nonlinear function of the environment (road, vehicle, etc.) and the controller. By manipulating the lateral deviation data through Eq. (2), the attacker is able to make changes in the distribution of e_1 which in turn leads to non-zero values of P_E, as shown in Fig. 3b. The adversary achieves the mentioned goal by tuning the learning parameter α in our example. The parameter is learned by using the Newton-Raphson method. In the next subsection, the learning process of the parameter α is explained.

C. Actor-Critic Learning of α in the Covert LK Attack

As mentioned in the previous subsection, the parameter α should be learned such that P_{E_A} is increased to an amount desirable to the attacker. This could be as low as 0.01 in a covert attack. In order to find a proper α, the problem is formulated as the root finding problem, and then it can be solved by using the Newton-Raphson method. To explain the algorithm, assume that the current parameter is α_c, which gives a probability of P_{E_C}. The problem is to find the value for α that makes P_{E_C} converge to P_{E_A}. Thus, we define the probability function as $g(\alpha) = P_{E_A} - P_{E_C}$. Then, we employ Newton-Raphson method to find the root of $g(\alpha)$ as follows:

$$\alpha_{n+1} = \alpha_n - \frac{g(\alpha_n)}{\dot{g}(\alpha_n)} \tag{4}$$

where $\dot{g}(.)$ is the gradient of $g(.)$. The dependency of P_E on α is not linear, nor necessarily monotonic. This is because there

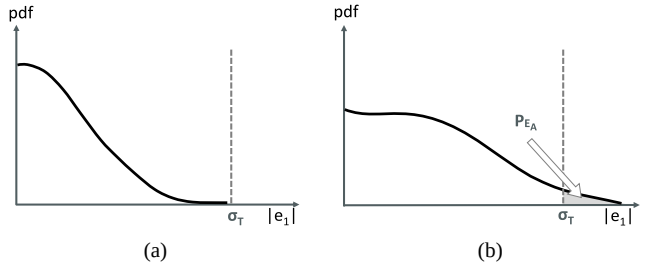


Fig. 3. The probability density function of lateral deviation values ($f_{e_1}(e)$) (a) when there is no attack and the vehicle works normally, and (b) when the adversary has manipulated the sensor readings in a cover attack try.

are nonlinear components in the loop (e.g. the MPC controller or the road and the car themselves). However, the intuition is that with the increase of α, P_E should increase too, perhaps statistically. To avoid the potential problem of having many roots for a not-necessarily monotonic $g(\alpha)$, we use an isotonic regression of it instead. We denote this regression by $\tilde{g}(.)$ It can approximate the observed points as shown in Fig. 4. The isotonic regression is distinguished by a dashed red line. The slope, or $\tilde{g}(.)$ at the each point, can be thought of as $\tan(\theta)$, where θ is shown in the figure. This quantity is the gradient of $\tilde{g}(.)$ at α_C. At each α_C, we can calculate the (average) difference of P_{E_A} and P_{E_C} and find a linear estimation of the amount that we should go forward to reach P_{E_A}. This quantity is denoted by x in the figure. Newton-Raphson works similarly and α is updated based on the relation presented in Eq. (4), however, by using $\tilde{g}(.)$ rather than $g(.)$. When the method converges and finds the root of $\tilde{g}(.)$, the proper attack parameter is obtained.

In this study, the critic calculates $g(\alpha_C)$ and its gradient by the isotonic regression method. The actor adapts the parameter by using the presented adaptation law during the learning process in order to converge the parameter. Afterwards, the adversary employs the parameter in Eq. (2) to make changes to the sensor measurements. In the simulations section, we show how this affects the error distribution.

D. Covert Intrusion Detection System

In this section, we develop an IDS to detect the covert attacks on autonomous driving vehicles. Anomaly detection is the key to disclose covert attacks as long as there is a reference for normal behavior to rely on. In the specific case of LK attack, we use GPS and the offline map. GPS signals are not really a necessity, but they make the development of IDS easier. We assume that the GPS data of the autonomous driving vehicle is available and not compromised. Checking the integrity of GPS outputs is possible by using the vehicle local motion vectors, however, we do not get into that here.

The IDS we employ reconstructs the road curves by using the sensory data and compares the obtained map with the one stored offline. Modern cars usually have an offline copy of the map for navigational purposes. Since normal visual sensors have errors and make mistakes even in benign conditions, there is always temporal noise in the reconstruction, thus instantaneous detection of covert LK attack is not possible in this method. Since the aim of covert attacks is to create damage in the long run, it is justifiable to make the detection decision in a reasonable time period rather than in an instant. We use statistical means to do so and apply thresholds on pdfs to detect behavioral changes.

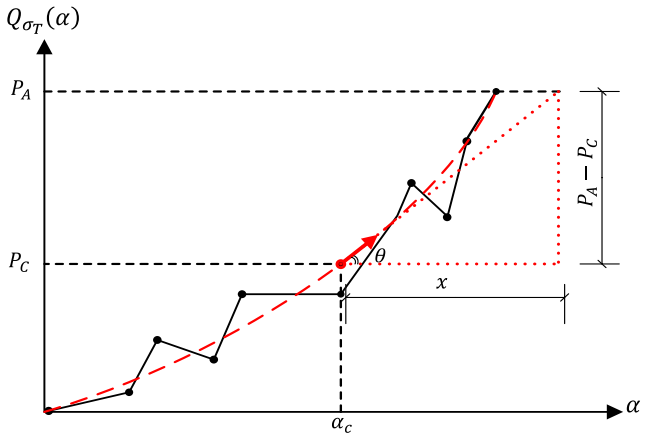


Fig. 4. Using Newton-Raphson method for parameter tuning.

Let us denote the offline map curve by κ_m and the reconstructed one by κ_r. Referring to Fig. 2, since GPS is available, we have the vehicle location in a global coordinates system (the long red vector in the figure which we refer to as $\vec{L}_c$). Also, from the lane sensors (whether attacked or not), we find the road direction as a local vector (denoted by $\vec{x}$) anchored at $\vec{L}_c$. Note that $\vec{x}$ and $\vec{y}$ form a local coordinates system at the car location oriented towards the direction of movement. Now, $\vec{e}_1$, which is the lateral deviation vector measured in the local coordinates system by the vehicle lane sensors, can be used to estimate the road center point, that is,

$$\vec{C}_r = \vec{L}_c + \vec{e}_1 \tag{5}$$

where $\vec{C}_r$ is the road center point, $\vec{e}_1$ is the lateral deviation vector, and $\vec{L}_c$ is the vehicle location in a global coordinates system. Equation (5) can be calculated at sensors sampling times, enabling us to have a reconstructed curvature (κ_r) based on the samples of road centers. Reconstruction of the curvature (κ_r) is achieved using the measured samples. Then, spatial matching of κ_r and the reference (κ_m) centers can be done and the error distribution can be calculated. This error shows the difference of the actual road center and the estimated one.

To summarize, the IDS identifier reconstructs the road curvature by using the (GPS) sensor data and compares it with the road curve obtained from the stored offline map. The instantaneous residual error is calculated as $(\kappa_r - \kappa_m)$ where κ_r is the reconstructed road curvature that the vehicle is being driven on and κ_m is the reference curvature found in the stored maps. The statistical data of the residual errors collected during the process can be used to estimate f_{e_1}. The alert threshold could be set to σ_T. However, we know that measurements are not always correct and might be noisy. Therefore, hard thresholding with instantaneous decision making might not be preferred. In such cases, IDS may use the same relation of Eq. (3) to find the probability of excessive deviations and put a (presumably very low) detection threshold on that probability.

V. EVALUATION RESULTS

In this section, we simulate the covert attack instance introduced in the previous section and also evaluate the developed IDS. As mentioned before, an intelligent adversary, either directly or by planting a malicious agent, manipulates the sensory data that goes to the MPC controller. Experiments are conducted on Nurburgring-Grand Prix track in MATLAB 2020a. First, we present the vehicle model and environment used in our experiments. Then, we explain the details.

$$\frac{d}{dt}\begin{bmatrix} e_1 \\ \dot{e}_1 \\ e_2 \\ \dot{e}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & -\frac{2C_{\alpha f}+2C_{\alpha r}}{m\ v_x} & \frac{2C_{\alpha f}+2C_{\alpha r}}{m} & \frac{-2C_{\alpha f}l_f+2C_{\alpha r}l_r}{m\ v_x} \\ 0 & 0 & 0 & 1 \\ 0 & -\frac{2C_{\alpha f}l_f-2C_{\alpha r}l_r}{I_z\ v_x} & \frac{2C_{\alpha f}l_f-2C_{\alpha r}l_r}{I_z} & \frac{2C_{\alpha f}l_f^2-2C_{\alpha r}l_r^2}{I_z\ v_x} \end{bmatrix} \begin{bmatrix} e_1 \\ \dot{e}_1 \\ e_2 \\ \dot{e}_2 \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{2C_{\alpha f}}{m} \\ 0 \\ \frac{2C_{\alpha f}l_f}{I_z} \end{bmatrix} \delta + \begin{bmatrix} 0 \\ -\frac{2C_{\alpha f}l_f-2C_{\alpha r}l_r}{m\ v_x} - v_x \\ 0 \\ -\frac{2C_{\alpha f}l_f^2+2C_{\alpha r}l_r^2}{I_z\ v_x} \end{bmatrix} \dot{\psi}_{des} \quad (6)$$

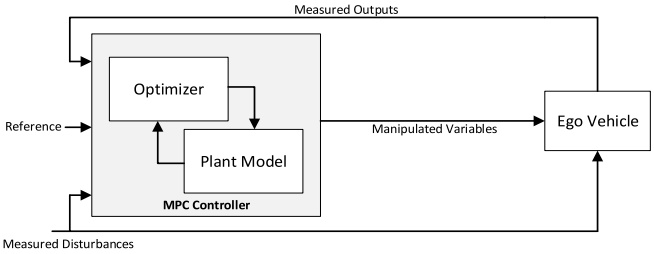


Fig. 5. Model predictive control system [31].

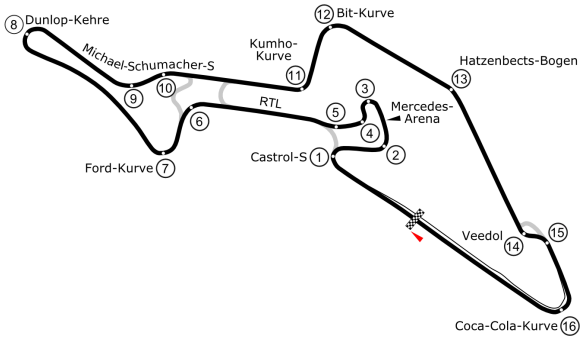


Fig. 6. Map of the modern Nurburgring-Grand Prix race course [32].

Vehicle Model: A lateral dynamic model of the vehicle contains the position and orientation errors with respect to the road. A proper steering angle (δ) should be prescribed by the controller. As defined before, e_1 and e_2 are the two errors involving in the lateral dynamic model. The dynamic equations of the system are obtained as a linear time invariant (LTI) model (as introduced in [29]). Considering the equations of the lateral translation motion y and the yaw angle ψ (as in [29]), the lateral dynamic equations of the vehicle can be presented in the linear state space form as shown in Eq. (6), where δ is the steering angle that is the control input of the system, and $\dot{\psi}_{des} = \kappa v_x$ [29]. The vehicle mass is m and I_z is the yaw moment of inertia. The longitudinal distance from the center of gravity to the front and rear wheels are l_f and l_r, respectively. And finally, $C_{\alpha f}$ and $C_{\alpha r}$ are the cornering stiffness of the front and rear tires [30]. These parameters vary from vehicle to vehicle.

Main Controller. In this study, MPC is employed as the main controller which contains two main parts as depicted in Fig. 5. The first one is an optimizer which finds the optimal value by minimizing a cost function on all constraints. The second part is the plant model (in our case, it is the vehicle LTI model of Eq. (6)). To design the MPC controller, the mentioned linear model of the vehicle is taken into account, but the validation of MPC is carried out using its nonlinear model. The MPCs optimal output is applied as a control input to the vehicle. It is developed using a finite horizon in an iterative manner. The goal is to provide a sequence of control inputs such that the predicted system's outputs follow the reference or set point.

In our scenario, the LK control has been modeled as a reference path tracking problem for the MPC. The objective is to minimize the lateral deviation e_1 and the relative yaw angle e_2, while the longitudinal speed is constant and the steering angle is the control output handled by the MPC. In the LK scenario, inputs of the MPC are the longitudinal velocity, lateral deviation, relative yaw, and predicted curvature. In this scenario, the predicted curvature is found as:

$$\hat{\kappa} = \kappa + \dot{\kappa}\ v_x\ h \quad (7)$$

where v_x is the longitudinal velocity, κ is the current curvature and h is the prediction horizon. Curvature and its derivative are estimated by using the lane sensor measurements. The optimal steering angle is achieved by the adaptive MPC. It is applied to the autonomous driving vehicle as a control input.

Environment. Nurburgring Grand Prix track whose model is used for simulations was built in 1984, and extended in 2001 to a modern racing course. Its map is illustrated in Fig. 6. Its width is between 10 to $25m$ and its length is 5.148 Km. There are ten right and seven left turns/curves [32, 33].

To test the proposed attack and detection approaches, the map of Nurburgring Grand Prix track is imported to the simulator with a virtual lane width of $8.15m$.

Experiments. The vehicle parameters used in the simulations are as follows: $m = 1575\ kg$, $I_z = 2875\ m.N.s^2$, $l_f = 1.2\ m$, $l_r = 1.6\ m$, $C_{\alpha f} = 19000\ N/rad$ and $C_{\alpha r} = 33000\ N/rad$. We set the longitudinal velocity to be constantly $13.9\ m/s$ and h to 30 time steps. The vehicle and lane widths are $1.82\ m$ and $8.15\ m$, respectively. The adversary wants the

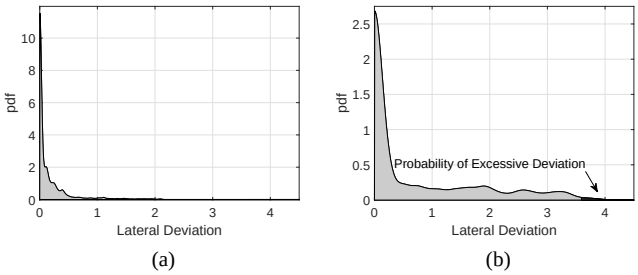
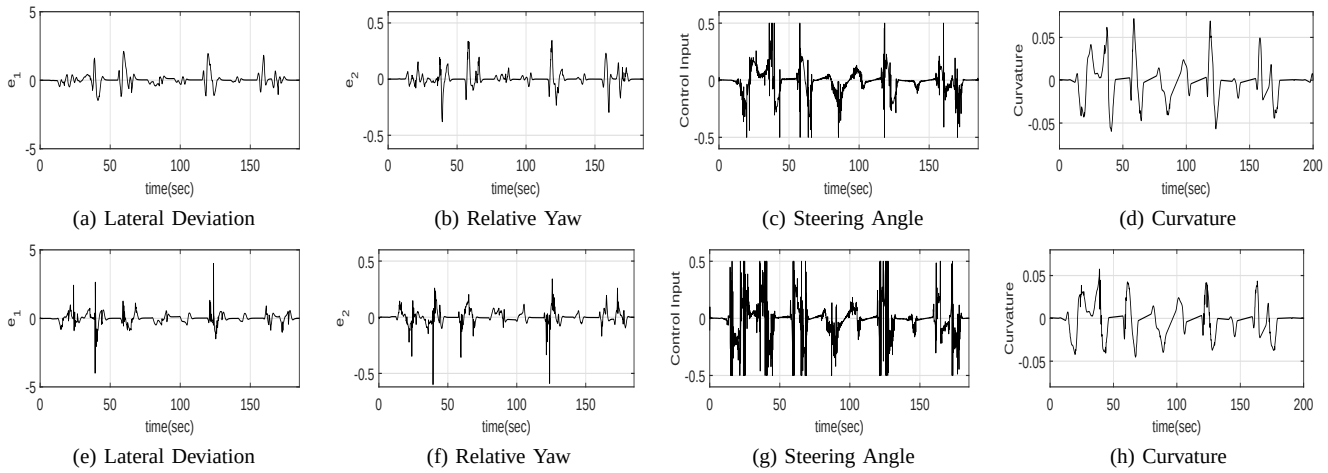


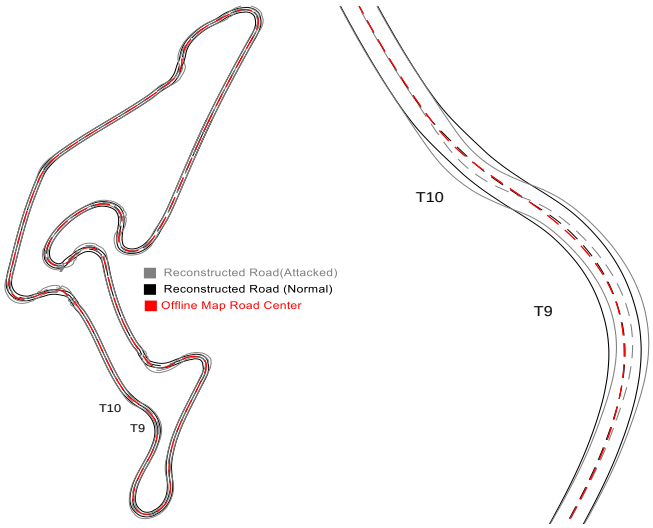
Fig. 7. The pdf of lateral deviations (a) The autonomous vehicle is in the safe and secure condition. (b) The vehicle is under a learning-based covet attack which leads to statistically excessive lateral deviations.



(a) Lateral Deviation (b) Relative Yaw (c) Steering Angle (d) Curvature

(e) Lateral Deviation (f) Relative Yaw (g) Steering Angle (h) Curvature

Fig. 8. Lateral dynamics and control input of the autonomous driving vehicle. The plots placed on the top pertain to the benign condition and the ones at the bottom show the system under the covert attack.



(a) (b)

Fig. 9. (a) The reconstructed roads (curves) in both normal and attack conditions. The center line of the offline map has also been plotted. (b) A closer look to the same reconstructed paths at T9 and T10 turns.

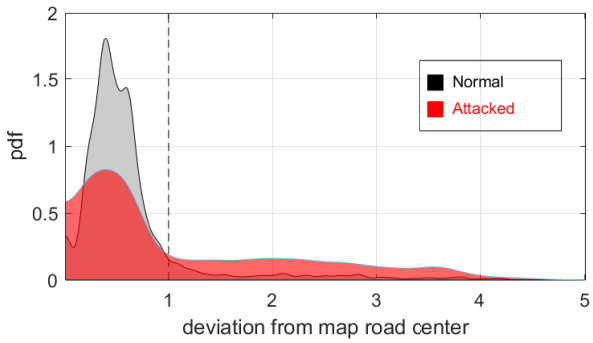


Fig. 10. Probability density function of the lateral distance between the offline map/lane road center and the reconstructed one. The error distributions are clearly different in normal and attack scenarios.

lateral deviation to be more than $\sigma_T = 8.15/2 - 1.82/2 + 0.5m$ with a probability of $P_{E_A} = 0.01$. This means that the attacker wants the vehicles body to go at least $50\ cm$ into the next/opposite lane 1% of the times during cornerings.

First, the autonomous driving vehicle is in the safe and secure condition. We made the vehicle go around the track and from a bird's eye, collected the lateral deviations to estimate the pdf. The normal lateral error pdf is depicted in Fig. 7a. Then, the adversary with $\sigma_T = 3.67\ m$ and $P_{E_A} = 0.01$ starts manipulating the lateral deviation reports coming out of the lane sensor by using Eq. (2). The adjustable parameter of the formula is obtained by the actor-critic learning method explained in Eq. (4). Ultimately, after going around the track multiple times, α^* is learned to be 76. Next, the learning phase is finished and the adversary uses the obtained parameter in order to make the covert attack permanent. The pdf of $|e_1|$ when the vehicle is under attack is illustrated in Fig. 7b. The area colored in black highlights the probability of excessive lateral deviations, i.e. $P_{E_A} = 0.01$. We additionally collected the details of lateral deviation $e_1(t)$, relative yaw $e_2(t)$, and the steering angle $\delta(t)$ which are illustrated in Fig. 8. The plots placed on the top pertain to the benign condition and the ones at the bottom show the system under the covert attack.

The designed covert attack increases the probability of excessive lateral deviations as well as accidents. Thus, this type of attack has influence on the road safety and sustainability. We have also road tested our IDS to detect such attacks. Fig. 9a shows the original track (map) as well as the reconstructed versions of it under normal and attack conditions. Deviations resulted from the attack (with an α value of 76) are visible clearly at the corners. The local distortions in the reconstruction are natural and are due to bad sensor readings, usually in the sharp corners where the camera loses visuals on one of the two lane lines. Fig. 9b shows a closer view to the same reconstructed path at T9 and T10 turns. We have calculated the pdf of spatial distances between the map road centers and those of the reconstructed roads. Fig. 10 clearly shows that the

covert attack has made the tail of the pdf grow longer. Since the vehicle could have learned the normal pdf during the safe times before attack, it can detect abnormal deviations rather easily. The area under the curve from an imaginary line at "1" to infinity can be used to detect the covert attack in this case.

VI. CONCLUSION

In this paper, the concept of covert cyber attack on intelligent transportation systems is introduced. As an instance, a covert adversarial learning attack is designed to disturb the lane keeping system of autonomous vehicles through manipulation of sensor readings. The control system is thus deceived to create high lateral deviation errors. The attacker/worm gradually learns the host vehicle attributes based on the actor-critic learning method. A novel intrusion detection system (IDS) is also developed using GPS data and offline maps which works based on the notion of curve reconstruction. Nurburgring-Grand Prix track is used for simulations and the results confirm the effectiveness of both attack and detection strategies. At the defense side, even at a target attack probability of 1%, the changes in the statistics are big enough for a confident raise of the IDS flag. The proposed attack framework is inspired by adversarial machine learning and has this advantage that it is generic and can be instantiated in different vehicular scenarios. It urges the need for a twin generic protection or detection mechanism, which is the subject of our future research.

REFERENCES

[1] C. Dong, J. M. Dolan, and B. Litkouhi, "Smooth behavioral estimation for ramp merging control in autonomous driving," in *IEEE Intelligent Vehicles Symposium (IV)*, 2018, pp. 1692–1697.

[2] N. Toorchi, M. A. Attari, M. S. Haghighi, and Y. Xiang, "A markov model of safety message broadcasting for vehicular networks," in *IEEE Wireless Communications and Networking Conference*, 2013.

[3] BBC News. (2015, accessed on 01.011.2019) Fiat chrysler recalls 1.4 million cars after jeep hack. [Online]. Available: https://www.bbc.com/news/technology-33650491

[4] M.-J. Kang and J.-W. Kang, "Intrusion detection system using deep neural network for in-vehicle network security," *PloS one*, vol. 11, no. 6, p. e0155781, 2016.

[5] A. O. de Sá, L. F. R. da Costa Carmo, and R. C. Machado, "Covert attacks in cyber-physical control systems," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 4, pp. 1641–1651, 2017.

[6] K. Shi, X. Yuan, G. Huang, and Z. Liu, "Compensation-based robust decoupling control system for the lateral and longitudinal stability of distributed drive electric vehicle," *IEEE/ASME Transactions on Mechatronics*, vol. 24, no. 6, pp. 2768–2778, 2019.

[7] B. Qiao and X. Wu, "Lane change control of autonomous vehicle with real-time rerouting function," in *IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*, 2019, pp. 1317–1322.

[8] X.-H. Chang, Y. Liu, and M. Shen, "Resilient control design for lateral motion regulation of intelligent vehicle," *IEEE/ASME Transactions on Mechatronics*, vol. 24, no. 6, pp. 2488–2497, 2019.

[9] Y. Zhang, Q. Lin, J. Wang, S. Verwer, and J. M. Dolan, "Lane-change intention estimation for car-following control in autonomous driving," *IEEE Transactions on Intelligent Vehicles*, vol. 3, no. 3, pp. 276–286, 2018.

[10] M. R. Kandroodi and B. Moshiri, "Identification and model predictive control of continuous stirred tank reactor based on artificial neural networks," in *International Conference on Control, Instrumentation and Automation*, 2011, pp. 338–343.

[11] L. B. Cremean, T. B. Foote, J. H. Gillula, G. H. Hines, D. Kogan, K. L. Kriechbaum, J. C. Lamb, J. Leibs, L. Lindzey, C. E. Rasmussen *et al.*, "Alice: An information-rich autonomous vehicle for high-speed desert navigation," *Journal of Field Robotics*, vol. 23, no. 9, pp. 777–810, 2006.

[12] V. L. Thing and J. Wu, "Autonomous vehicle security: A taxonomy of attacks and defences," in *IEEE iThings, GreenCom, CPSCom and SmartData*, 2016, pp. 164–170.

[13] B. Mokhtar and M. Azab, "Survey on security issues in vehicular ad hoc networks," *Alexandria engineering journal*, vol. 54, no. 4, pp. 1115–1126, 2015.

[14] D. Djenouri, L. Khelladi, and N. Badache, "Security issues of mobile ad hoc and sensor networks," in *IEEE Communications Surveys Tutorials*, vol. 7, no. 4. IEEE Communications Society, 2005, pp. 2–28.

[15] M. Sayad Haghighi and Z. Aziminejad, "Highly anonymous mobility-tolerant location-based onion routing for vanets," *IEEE Internet of Things Journal*, vol. 7, no. 4, pp. 2582–2590, 2019.

[16] A. Jolfaei and K. Kant, "Privacy and security of connected vehicles in intelligent transportation system," in *IEEE/IFIP International Conference on Dependable Systems and Networks*, 2019, pp. 9–10.

[17] F. Farivar, M. S. 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*, in press, 2020.

[18] F. Farivar, S. Barchinezhad, M. Sayad Haghighi, and A. Jolfaei, "Detection and compensation of covert service-degrading intrusions in cyber physical systems through intelligent adaptive control," in *IEEE International Conference on Industrial Technology*, 2019.

[19] Y. Tang, D. Zhang, X. Jin, D. Yao, and F. Qian, "A resilient attitude tracking algorithm for mechanical systems," *IEEE/ASME Transactions on Mechatronics*, vol. 24, no. 6, pp. 2550–2561, 2019.

[20] S. Kumar and K. S. Mann, "Detection of multiple malicious nodes using entropy for mitigating the effect of denial of service attack in vanets," in *International Conference on Computing Sciences*, 2018, pp. 72–79.

[21] J. Liang, J. Chen, Y. Zhu, and R. Yu, "A novel intrusion detection system for vehicular ad hoc networks (vanets) based on differences of traffic flow and position," *Applied Soft Computing*, vol. 75, pp. 712–727, 2019.

[22] A. R. Javed, M. Usman, S. U. Rehman, M. U. Khan, and M. S. Haghighi, "Anomaly detection in automated vehicles using multistage attention-based convolutional neural network," *IEEE Transactions on Intelligent Transportation Systems*, in press, 2020.

[23] F. Farivar, M. S. Haghighi, A. Jolfaei, and M. Alazab, "Artificial intelligence for detection, estimation, and compensation of malicious attacks in nonlinear cyber-physical systems and industrial iot," *IEEE transactions on industrial informatics*, vol. 16, no. 4, pp. 2716–2725, 2019.

[24] F. Farivar and M. N. Ahmadabadi, "Continuous reinforcement learning to robust fault tolerant control for a class of unknown nonlinear systems," *Applied Soft Computing*, vol. 37, pp. 702–714, 2015.

[25] F. Farivar, M. A. Shoorehdeli, M. Nekoui, M. Teshnehlab, "Synchronization of underactuated unknown heavy symmetric chaotic gyroscopes via optimal gaussian radial basis adaptive variable structure control," *IEEE Transactions on control systems technology*, vol. 21, no. 6, pp. 2374–2379, 2013.

[26] M. H. Basiri, J. G. Thistle, J. W. Simpson-Porco, and S. Fischmeister, "Kalman filter based secure state estimation and individual attacked sensor detection in cyber-physical systems," in *American Control Conference*, 2019, pp. 3841–3848.

[27] M. Farahani, M. Farahani, M. Manthouri, and O. Kaynak, "Short-term traffic flow prediction using variational lstm networks," *arXiv preprint arXiv:2002.07922*, 2020.

[28] B. Du, H. Peng, S. Wang, M. Z. A. Bhuiyan, L. Wang, Q. Gong, L. Liu, and J. Li, "Deep irregular convolutional residual lstm for urban traffic passenger flows prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 3, pp. 972–985, 2019.

[29] R. Rajamani, *Vehicle dynamics and control*. Springer Science & Business Media, 2011.

[30] Z. Nie and H. Farzaneh, "Adaptive cruise control for eco-driving based on model predictive control algorithm," *Applied Sciences*, vol. 10, no. 15, p. 5271, 2020.

[31] K. Kone, "Lateral and longitudinal control of an autonomous racing vehicle." Ph.D. dissertation, Politecnico di Torino, 2019.

[32] E. Neussner, "The new nurburgring, the most modern grand prix race course in the world," *Straße Und Autobahn*, 1984.

[33] Nuburgring track. [Online]. Available: www.nuerburgring.de/en, Accessed on 04.03.2021