Theory and Implementation of Dynamic Watermarking for Cybersecurity of Advanced Transportation Systems

Woo-Hyun Ko, Bharadwaj Satchidanandan, and P R Kumar, *Fellow, IEEE*

***Abstract*—We consider a prototypical intelligent transportation system with a control law that is specifically designed to avoid collisions. We experimentally demonstrate that, nevertheless, an attack on a position sensor can result in collisions between vehicles. This is a consequence of the feeding of malicious sensor measurements to the controller and the collision avoidance module built into the system. This is an instance of the broader concern of cybersecurity vulnerabilities opened up by the increasing integration of critical physical infrastructures with the cyber system. We consider a solution based on "dynamic watermarking" of signals to detect and stop such attacks on cyber-physical systems. We show how dynamic watermarking can handle nonlinearities arising in vehicular models. We then experimentally demonstrate that employing this nonlinear extension indeed restores the property of collision freedom even in the presence of attacks.**

***Index Terms*—Dynamic Watermarking, Secure control, Cyber-physical systems, Cybersecurity, Autonomous Transportation Systems, Driver Assist Systems, Intelligent Transportation Systems**

I. Introduction

RECENTLY there has been great interest in automated as well as semi-automated transportation systems involving various driver assists. These advanced systems rely on sensors to provide state information and situational awareness to the control logic governing the vehicle. However, this has also increased the vulnerability of these advanced transportation systems to cyber attacks. In fact, there have been demonstrated cyber attacks on automobiles in the recent past [1], [2], where two hackers have remotely subverted an automobile, taking control of its steering and braking units. This ultimately led to the automobile manufacturer recalling over a million cars to patch the identified vulnerabilities. Several other similar reports [3]–[5] point to the need for cybersecurity of automated transportation systems. In this paper we demonstrate how the technique of "dynamic watermarking" can be employed to secure an automated transportation system against arbitrary attacks on its sensors.

More broadly, this problem lies in the larger realm of security of cyber-physical systems (CPS). These consist of a physical plant which is to be controlled using sensors, controllers, and actuators that share information over an underlying communication network, such as the Internet.

In previous work, the technique of dynamic watermarking of signals has been proposed to secure such systems [6]–[10]. It has been shown in [9] that, in theory, employing dynamic watermarking and using two specific tests can detect erroneous sensor measurements for a large class of linear systems. In this paper we investigate whether this method can actually be used in a real transportation system to detect attacks on the positioning sensor systems and thereby prevent collisions. For this purpose we extend the theory to allow for several nonlinearities not accounted for in an idealized linear system for which the theoretical results have previously been established. We have implemented the dynamic watermarking method with this nonlinear extension on a laboratory autonomous transportation system. We demonstrate that this method successfully detects and responds to attacks on the position sensors which can otherwise cause collisions.

II. Related Work

Initial work on CPS security involved defining the problem of secure control, and identifying the aspects that distinguish it from the traditional problem of network and information security [11]. Fundamental limitations of different classes of attack detectors, viz. static, dynamic, and active detectors, in detection and identification of attacks are presented in [12]. A static detector/monitor does not know the system dynamics, while a dynamic monitor does. An active monitor is a dynamic monitor that can modify system behavior by injecting additional input signals. Conditions and an algorithm for secure state estimation in the presence of malicious sensors, and an approach to increase this resiliency, have been presented in [13]. Bounds on the state prediction error in the presence of adversaries, and an algorithm that achieves this bound are presented in [14]. Attacks that cannot be detected using input and output data are called zero-dynamics attacks, and [15] presents ways to detect such attacks by perturbing the system parameters. A data verification framework for detecting and removing faulty data from a wireless sensor network is presented in [16], where correlations between the measurements of different sensors are exploited to identify malicious reports.

The aforementioned techniques for CPS security can be classified as passive defense techniques. An alternate paradigm for CPS security is active defense, where the honest actuators actively probe the system by injecting a small, random signal, unknown to other nodes in the system. References [6]–[8],

This material is based upon work partially supported by NSF under Contract Nos. CNS-1646449 and Science & Technology Center Grant CCF-0939370, the U.S. Army Research Office under Contract No. W911NF-15-1-0279, and NPRP grant NPRP 8-1531-2-651 from the Qatar National Research Fund, a member of Qatar Foundation.

[17], are among the first papers to investigate such an active defense to secure CPS. Similar schemes are also investigated in other scenarios [18], [19]. This random signal, known as the "watermark," would evoke a particular response in accordance with the plant dynamics if the sensors were honest. The honest actuators can therefore check for maliciousness of sensors by checking if the measurements reported by the sensors are appropriately correlated with the random signal injected into the system. It has been shown in [9], that by using this approach of dynamic watermarking, and subjecting the reported sequence of measurements to two tests, it can be ensured that the malicious sensors do not distort the measurements beyond adding a zero power signal to the noise already entering the system for broad classes of linear systems.

III. Dynamic Watermarking

The fundamental idea of dynamic watermarking is to have each actuator i superimpose a random signal $e_i[t]$, known as the watermark, on the control policy-specified input. While the statistics of the watermark $\{e_i\}$ are made known to every node in the system, its actual realization is not revealed to any other node $j \neq i$ in the system. To introduce the ideas, we illustrate it for a linear system. Let $\mathbf{e}[t] \sim \mathcal{N}(0, \sigma_e^2 I)$ denote the vector of watermarks superimposed by all the nodes. The watermarks are assumed to be independent and identically distributed (i.i.d.) across time. With the superimposed watermarks, the linear system with vector output $y[t]$, control policy g, input vector u_t^g, and white Gaussian noise $\mathbf{w}[t]$ with mean zero and covariance matrix $\sigma_w^2 I$, evolves as

$$\mathbf{y}[t+1] = A\mathbf{y}[t] + B\mathbf{u}_t^g(\mathbf{z}^t) + B\mathbf{e}[t] + \mathbf{w}[t+1]. \quad (1)$$

Hence the actual sequence of measurements $\{\mathbf{y}\}$ satisfies

$$\{\mathbf{y}[t+1] - A\mathbf{y}[t] - B\mathbf{u}_t^g(\mathbf{z}^t)\} \sim \mathcal{N}(0, BB^T\sigma_e^2 + \sigma_w^2 I),$$

and

$$E[e_i[t](\mathbf{y}[t+1] - A\mathbf{y}[t] - B\mathbf{u}_t^g(\mathbf{z}^t))] = B_{.i}\sigma_e^2,$$

where $B_{.i}$ denotes the i^{th} column of B.

Based on the above observations, each honest actuator $i \in \{1, 2, ..., m\}$ subjects the reported sequence of measurements to the following two tests.

1) **Test 1:** The i-th node checks if the reported sequence of measurements satisfies

$$\lim_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T-1} (\mathbf{z}[k+1] - A\mathbf{z}[k] - Bg_k(\mathbf{z}^k))$$
$$(\mathbf{z}[k+1] - A\mathbf{z}[k] - Bg_k(\mathbf{z}^k))^T = \sigma_e^2 BB^T + \sigma_w^2 I_n \quad (2)$$

2) **Test 2:** The i-th node also checks if the reported sequence of measurements satisfies

$$\lim_{T\to\infty} \frac{1}{T} \sum_{k=0}^{T-1} e_i[k](\mathbf{z}[k+1] - A\mathbf{z}[k] - Bg_k(\mathbf{z}^k))$$
$$= B_{\cdot,i}\sigma_e^2 \quad (3)$$

In [9], it was established that if the reported sequence of measurements passes the above tests, then any malicious sensor present could not have distorted the actual measurements beyond adding a zero power sequence to the process noise.

IV. Attacking an autonomous transportation system by malicious attacks on sensors

Advanced transportation systems, whether fully autonomous or those that provide driver assists, employ sensors to determine the position of a vehicle. Such sensors can be maliciously attacked. We first begin by experimentally demonstrating how such an attack can be launched to create collisions. The experimental setup is a prototype of an intelligent transportation system housed in the Cyberphysical Systems Laboratory at Texas A&M University. At the core of the system is a set of vehicles that are required to follow a particular trajectory within a rectangular area. This can be thought of as the high-level control objective. The system consists of a supervisory layer or a high-level controller which decides the trajectory that each vehicle should follow.

Monitoring the environment is a set of ten cameras which capture an image of the rectangular area including the vehicles once every 100ms. These images are transmitted to the vision sensors in the system which accurately computes, from these raw images, the coordinates (x_i, y_i) and orientation θ_i of each vehicle i at each sampling instant t. The vision sensors then transmit this information to the vision server, which disseminates this information to other modules in the system such as the low-level controller, supervisor, collision avoidance module, etc.

The low-level controller determines the control input to be applied to each vehicle using Model Predictive Control. It computes the control input using the position and orientation information of each vehicle that it obtains from the vision server, and also their reference trajectories that it obtains from the supervisor. The controller then sends this information to the collision avoidance module.

The collision avoidance module uses this control input, the position estimates provided by the vision server, and the dynamic models of the vehicles to predict their positions during the next sampling epoch. If it detects an imminent collision based on this computation, it instructs the actuators to halt the vehicles. If not, it relays the control input computed by the controller as such to the actuator, which then applies that particular input.

The plant model for vehicle i is given by its kinematic equations:

$$x_i[t+1] = x_i[t] + h\cos(\theta_i[t])v_i[t] + h\cos(\theta_i[t])w_{ix}[t], \quad (4)$$
$$y_i[t+1] = y_i[t] + h\sin(\theta_i[t])v_i[t] + h\sin(\theta_i[t])w_{iy}[t], \quad (5)$$
$$\theta_i[t+1] = \theta_i[t] + h\omega_i[t] + hw_{i\theta}[t], \quad (6)$$

where h is the sampling time period (100ms in this case), $v_i[t]$ is the speed of the vehicle at sampling epoch t and is one of the control inputs of the vehicles, while $\omega_i[t]$ is the angular speed of the vehicle at sampling epoch t and is the second control input of the vehicles. Also, $w_{ix}[t]$, $w_{iy}[t]$, and $w_{i\theta}[t]$ are random variables whose variances we denote by σ_x^2, σ_y^2, and σ_θ^2 respectively. They model the ambient noise entering the system as a consequence of small, random drifts in the actual values of the applied control inputs from their set points. We model them as zero-mean, i.i.d. normal random variables. For the purposes of our demonstration, it suffices to use just two vehicles, so that $i \in \{1,2\}$. We make both of them follow an elliptical trajectory, one behind the other. The accompanying video clip in [20] opens by showing the vehicle trajectories in the absence of both attacks and dynamic watermarking.

Our system includes a collision avoidance module [21] that halts the vehicles when an imminent collision is detected. To illustrate its behavior, when a manually controlled vehicle is made to intercept the two vehicles' trajectories, the collision avoidance module detects imminent collisions and commands the actuators to halt the vehicles, thereby avoiding collisions. This can also be seen in the video clip [20]. In a prior work [21], it was shown that this system indeed guarantees collision freedom.

Next, we construct a specific attack which spoofs the collision avoidance module by sending malicious position information to it. Specifically, we introduce maliciousness in the vision sensor which computes the $x-$coordinate of the vehicles' position from the image that it receives from the cameras. The attack strategy is as follows. Let t_A denote the time at which the attack begins. Then, $z_{2x}[t_A] = x_2[t_A] + \tau$, where τ is the bias that the sensor adds to the $x-$coordinate of the vehicle. For $t > t_A$, the malicious sensor reports measurements $\{z_{2x}\}$ generated as

$$z_{2x}[t+1] = z_{2x}[t] + h\cos(\theta_2[t])u^g_{v_2}(\mathbf{z}_1^t, \mathbf{z}_2^t) + h\cos(\theta_2[t])n[t], \tag{7}$$

where $n[t] \sim \mathcal{N}(0, \sigma_x^2)$, and $u^g_{v_i}(\mathbf{z}_1^t, \mathbf{z}_2^t)$ denotes the control policy-specified input for the input $v_i[t]$. Therefore, once the attack begins, wrong position information is sent to the vision server, and consequently to all other modules in the system. In particular, wrong position information is sent to the collision avoidance module, which results in it not detecting imminent collisions. The demonstration of this attack which culminates in the two vehicles colliding with each other can be seen in [20].

For the purposes of implementation, we have chosen a specific attack strategy as described above. However, as we show in the following section, arbitrary attacks causing "excessive" distortion that may be employed by the malicious sensor can also be detected by dynamic watermarking.

V. Protecting transportation systems from malicious attacks on sensors through a nonlinear extension of dynamic watermarking

We now consider how the transportation system can be protected from malicious attacks on sensors. We first extend the theory of dynamic watermarking to nonlinear systems to address the equations of vehicular motion. Consider a controller, as in Section-IV, which computes the control policy-specified input and superimposes the watermark on it. The system evolves as

$$\begin{aligned} x_i[t+1] &= x_i[t] + h\cos(\theta_i[t])u^g_i(\mathbf{z}_1^t, \mathbf{z}_2^t) \\ &\quad + h\cos(\theta_i[t])e_{iv}[t] + h\cos(\theta_i[t])w_{ix}[t], && (8) \\ y_i[t+1] &= y_i[t] + h\sin(\theta_i[t])u^g_i(\mathbf{z}_1^t, \mathbf{z}_2^t) \\ &\quad + h\sin(\theta_i[t])e_{iv}[t] + h\sin(\theta_i[t])w_{iy}[t], && (9) \\ \theta_i[t+1] &= \theta_i[t] + h\omega_i[t] + he_{i\theta}[t] + hw_{i\theta}[t]. && (10) \end{aligned}$$

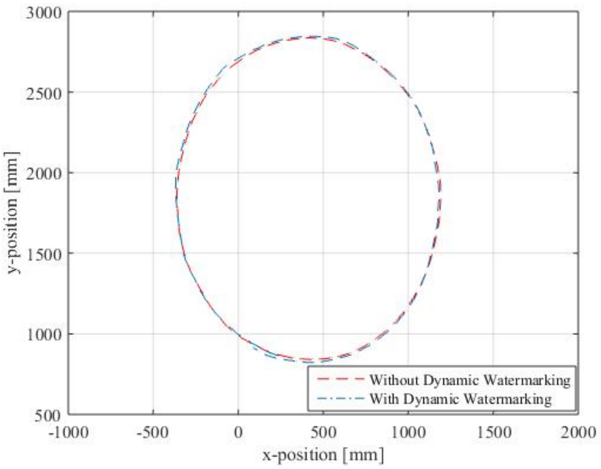

Fig. 1. Trajectories of the vehicles with and without Dynamic Watermarking

Above, $e_{iv}[t] \sim \mathcal{N}(0, \sigma_e^2)$ and i.i.d. across time is the watermark superimposed on the translational velocity control input $v_i[t]$, $e_{i\theta}[t] \sim \mathcal{N}(0, \sigma_\theta^2)$ and i.i.d. across time is the watermark superimposed on the angular velocity control input $\omega_i[t]$, $\mathbf{z}_i[t] = [z_{ix}[t], z_{iy}[t], z_{i\theta}[t]] = [z_{ix}[t], y_i[t], \theta_i[t]]^T$, $z_{ix}[t]$ is vehicle-i's $x-$coordinate reported by the vision sensor at time t differing from its true value $x_i[t]$, while $z_{iy}[k]$ and $z_{i\theta}[k]$ are the values reported for $y_i[k]$ and $\theta_i[k]$ respectively that are equal to their true values at all times. The controller does not know a priori which of the sensors are malicious, if any.

The nominal trajectories of the vehicles in the presence and absence of dynamic watermarking are compared in Fig. 1. The mismatch between them is negligible, and in principle, can be made arbitrarily small. Thus watermarking per se does not cause any significant performance deterioration just by its mere usage.

Section-III outlined the rationale behind conducting the tests (2) and (3) for securing linear systems. Motivated by the same line of reasoning, the controller in the vehicular CPS performs the following two tests to check for maliciousness. The tests are specified only for $\{z_{ix}\}$ below, but analogous tests are also carried out by the controller for $\{z_{iy}\}$ and $\{z_{i\theta}\}$. While we state the following as standalone tests, they are subsumed by (2) and (3). The asymptotic equalities to be checked below are converted to statistical tests on finite time deviation in a standard manner.

1) **Test 1:** The controller checks if

$$\begin{aligned} &\lim_{t\to\infty} \frac{1}{t} \sum_{k=0}^{t-1} (z_{ix}[k+1] - z_{ix}[k] \\ &-h\cos(z_{i\theta}[k])u^g_{v_i}(\mathbf{z}_1^t, \mathbf{z}_2^t) - h\cos(z_{i\theta}[k])e_{iv}[k])^2 = \widetilde{\sigma}_x^2. \end{aligned} \tag{11}$$

2) **Test 2:** The controller checks if

$$\begin{aligned} &\lim_{t\to\infty} \frac{1}{t} \sum_{k=0}^{t-1} (z_{ix}[k+1] - z_{ix}[k] \\ &-h\cos(z_{i\theta}[k])u^g_{v_i}(\mathbf{z}_1^t, \mathbf{z}_2^t))^2 = \sigma_c^2. \end{aligned} \tag{12}$$

where

$$\widetilde{\sigma}_x^2 := \lim_{t\to\infty} \frac{1}{t} \sum_{k=0}^{t-1} (h\cos(\theta_i[k]) w_{ix}[k])^2, \tag{13}$$

and

$$\sigma_c^2 := \lim_{t\to\infty} \frac{1}{t} \sum_{k=0}^{t-1} (h\cos(\theta_i[k]) e_{iv}[k] + h\cos(\theta_i[k]) w_{ix}[k])^2 \tag{14}$$

are the values that would be attained by the LHSs of (11) and (12) respectively had $\{z_{ix}\}$ been equal to $\{x_i\}$. The above quantities are the sum of independent, but non-identically distributed random variables. We assume that for the trajectory followed by the vehicles, the above limits exist. Fig. 2 plots $\frac{1}{t}\sum_{k=0}^{t-1}(h\cos(\theta_i[k])w_{ix}[k])^2$ and $\frac{1}{t}\sum_{k=0}^{t-1}(h\cos(\theta_i[k])e_{iv}[k]+h\cos(\theta_i[k])w_{ix}[k])^2$ as a function of time, and support our assumption that the limits (13) and (14) exist. These quantities were computed experimentally as follows in our demonstration. Since in the absence of attacks, we have

$$\widetilde{\sigma}_x^2 = \lim_{t\to\infty} \frac{1}{t} \sum_{k=0}^{t-1} (x_i[k+1] - x_i[k] - h\cos(\theta_i[k]) u_{v_i}^g(\mathbf{z}_1^k, \mathbf{z}_2^k) \\ -h\cos(\theta_i[k]) e_{iv}[k])^2,$$

and

$$\sigma_c^2 = \lim_{t\to\infty} \frac{1}{t} \sum_{k=0}^{t-1} (x_i[k+1] - x_i[k] - h\cos(\theta_i[k]) u_{v_i}^g(\mathbf{z}_1^k, \mathbf{z}_2^k))^2,$$

evaluating the RHSs of these from the experiment in the absence of attacks yield the desired noise variances.

The following theorem ensures that the the above two tests are sufficient to restrict the malicious sensor to adding an additive distortion that can only have zero power asymptotically.

Theorem 1. *Define*

$$v_x[t+1] := z_{2x}[t+1] - z_{2x}[t] - h\cos(\theta_2[t]) u_2^g(\mathbf{z}_1^t, \mathbf{z}_2^t) \\ -h\cos(\theta_2[t]) e_{2v}[t] - h\cos(\theta_2[t]) w_{2x}[t]. \tag{15}$$

For an honest sensor, $v_x \equiv 0$. If the reported sequence of measurements satisfies (11) and (12), then,

$$\lim_{t\to\infty} \frac{1}{t} \sum_{k=0}^{t-1} v_x^2[k+1] = 0 \tag{16}$$

Proof. Since the sequence of reported measurements $\{z_{2x}\}$ satisfies (11), we have using (15)

$$\lim_{t\to\infty} \frac{1}{t} \sum_{k=0}^{t} (h\cos(\theta_2[k]) w_{2x}[k] + v_x[k+1])^2 = \widetilde{\sigma}_x^2. \tag{17}$$

Expanding the above and using (13), we get

$$\lim_{t\to\infty} \frac{1}{t} \sum_{k=0}^{t} v_x^2[k+1] + 2h\cos(\theta_2[k]) w_{2x}[k] v_x[k+1] = 0. \tag{18}$$

Similarly, since the reported measurements satisfy (12) we have,

$$\lim_{t\to\infty} \frac{1}{t} \sum_{k=0}^{t} (h\cos(\theta_2[k]) w_{2x}[k] + h\cos(\theta_2[k]) e_{2v}[k] + \\ v_x[k+1])^2 = \sigma_c^2. \tag{19}$$

Expanding the above and using (13), (14), and (18), we arrive at

$$\lim_{t\to\infty} \frac{1}{t} \sum_{k=0}^{t} \cos(\theta_2[k]) e_{2v}[k] v_x[k+1] = 0. \tag{20}$$

Now, define the sigma-algebra $\mathcal{F}_k := \sigma(x^k, y^{k-1}, \theta^{k-1}, e_{2v}^{k-2}, z^k)$, $\widehat{w}[k] := E[w[k] \big| \mathcal{F}_{k+1}]$, and $\widetilde{w}[k] := w[k] - \widehat{w}[k]$. Then, for k such that $\cos(\theta_2[k]) \neq 0$, we have $\widehat{w}[k] = \frac{\sigma_w^2}{\sigma_e^2+\sigma_w^2}(e[k] + w[k])$. Now,

$$\sum_{k=0}^{t-1} \cos(\theta_2[k]) w_{2x}[k] v_x[k+1] = \sum_{\substack{k=0,\\ \cos(\theta_2[k])\neq 0}}^{t-1} \cos(\theta_2[k]) w_{2x}[k] v_x[k+1].$$

Expressing $w_{2x}[k]$ as $\widehat{w}[k] + \widetilde{w}[k]$, substituting for $\widehat{w}[k]$ using the aforementioned expression, and rearranging the terms give

$$\sum_{\substack{k=0,\\ \cos(\theta_2[k])\neq 0}}^{t-1} \cos(\theta_2[k]) w_{2x}[k] v_x[k+1] = \\ \frac{\beta}{1-\beta} \sum_{\substack{k=0,\\ \cos(\theta_2[k])\neq 0}}^{t-1} \cos(\theta_2[k]) e_{2v}[k] v_x[k+1] \\ + \frac{1}{1-\beta} \sum_{\substack{k=0,\\ \cos(\theta_2[k])\neq 0}}^{t-1} \cos(\theta_2[k]) \widetilde{w}_{2x}[k] v_x[k+1], \tag{21}$$

where $\beta := \frac{\sigma_w^2}{\sigma_e^2+\sigma_w^2} < 1$. Now, $(\cos(\theta_2[k])\widetilde{w}_{2x}[k], \mathcal{F}_{k+2})$ is a Martingale Difference Sequence. Also, $v_x[k+1] \in \mathcal{F}_{k+1}$, since $v_x[k+1] = z_x[k+1] - z_x[k] - x[k+1] - x[k]$. So, the Martingale Stability Theorem (MST) [22] applies, and we have

$$\sum_{\substack{k=0,\\ \cos(\theta_2[k])\neq 0}}^{t-1} \cos(\theta_2[k]) \widetilde{w}_{2x}[k] v_x[k+1] = o(\sum_{\substack{k=0,\\ \cos(\theta_2[k])\neq 0}}^{t-1} v_x^2[k+1]) + O(1).$$

Substituting the above in (21), and the result in (18), we get

$$\sum_{\substack{k=0,\\ \cos(\theta_2[k])\neq 0}}^{t-1} v_x^2[k+1] + \frac{2h\beta}{1-\beta} \sum_{\substack{k=0,\\ \cos(\theta_2[k])\neq 0}}^{t-1} \cos(\theta_2[k]) e_{2v}[k] v_x[k+1] \\ +o(\sum_{\substack{k=0,\\ \cos(\theta_2[k])\neq 0}}^{t-1} v_x^2[k+1]) + O(1) + \sum_{\substack{k=0,\\ \cos(\theta_2[k])=0}}^{t-1} v_x^2[k+1] = o(t). \tag{22}$$

Dividing the above by t, taking the limit as $t \to \infty$, and invoking (20) completes the proof. □

Now we experimentally demonstrate the performance of dynamic watermarking for the transportation system. The specific

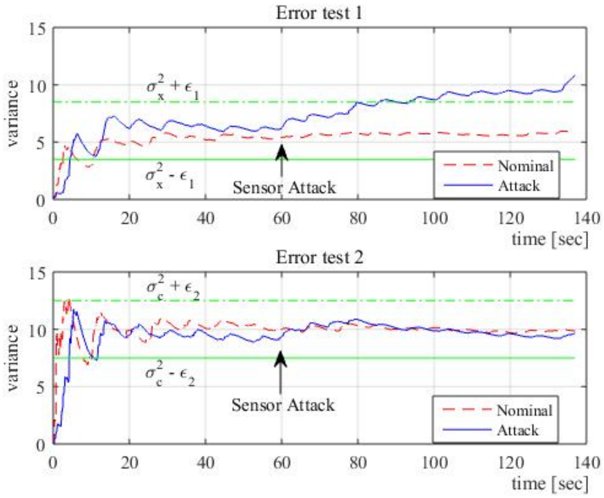

Fig. 2. Test statistics of error test 1 and 2 as a function of time

strategy that the sensor uses to fabricate the measurements is the same as (7), except that now, $n[t] \sim \mathcal{N}(0, \sigma_x^2 + \sigma_e^2)$. Note that this choice of $n[t]$ enables the attacker to evade detection if Test 2 alone is employed. The two tests (11) and (12) of dynamic watermarking are employed by the controller. The asymptotic tests (11) and (12) are converted in a straightforward manner to statistical tests for implementation by checking if for each time, the LHSs of (11) and (12) are within thresholds ϵ_1 and ϵ_2 respectively of their asymptotic values.

Fig. 2 plots the LHSs of (11) and (12) as a function of time. As can be seen, the false measurements pass test 2 but fail test 1, raising the alert that the system is under attack. The restoration of collision freedom for the automatic transportation system can be seen in [20].

VI. Conclusion

This paper addresses the problem of cybersecurity of advanced transportation systems, whether autonomous or driver assisted. This is an exemplar of the broader class of cyber-physical systems for which there is great current concern on the issue of security. We show how collisions can be caused in such systems by attacking the sensor, even though the control logic contains a collision avoidance system. To provide security against such attacks, we consider the usage of dynamic watermarking where actuators inject small private watermark signals into the system and check the reported sensor measurements for statistical consistency with the injected noise. We extend the theory of dynamic watermarking that has been developed for linear systems to nonlinear systems describing the equations of vehicular motion. We implement dynamic watermarking on a prototypical laboratory transportation system and experimentally demonstrate that it restores collision freedom.

References

[1] "Hackers remotely kill a jeep on the highway- with me in it." [Online]. Available: https://www.wired.com/2015/07/hackers-remotely-kill-jeep-highway/

[2] C. Miller and C. Valasek, "Remote exploitation of an unaltered passenger vehicle," *Black Hat, USA*, 2015.

[3] "Hackers reveal nasty new car attacks–with me behind the wheel (video)." [Online]. Available: http://www.forbes.com/sites/andygreenberg/2013/07/24/hackers-reveal-nasty-new-car-attacks-with-me-behind-the-wheel-video/#29621f635bf2

[4] "Security experts say that hacking cars is easy." [Online]. Available: http://fortune.com/2016/01/26/security-experts-hack-cars/

[5] "Your next car will be hacked. will autonomous vehicles be worth it?" [Online]. Available: https://www.theguardian.com/technology/2016/mar/13/autonomous-cars-self-driving-hack-mikko-hypponen-sxsw

[6] Y. Mo and B. Sinopoli, "Secure control against replay attacks," in *47th Annual Allerton Conference on Communication, Control, and Computing*, Sept 2009.

[7] Y. Mo, S. Weerakkody, and B. Sinopoli, "Physical authentication of control systems: Designing watermarked control inputs to detect counterfeit sensor outputs," *IEEE Control Systems*, vol. 35, no. 1, pp. 93–109, Feb 2015.

[8] S. Weerakkody, Y. Mo, and B. Sinopoli, "Detecting integrity attacks on control systems using robust physical watermarking," in *53rd IEEE Conference on Decision and Control*, Dec 2014, pp. 3757–3764.

[9] B. Satchidanandan and P. R. Kumar, "Dynamic watermarking: Active defense of networked cyber-physical systems," *Proceedings of the IEEE, to appear*.

[10] ——, "Secure control of networked cyber-physical systems," in *55th IEEE Conference on Decision and Control, 2016, to appear*.

[11] A. Cardenas, S. Amin, B. Sinopoli, A. Giani, A. Perrig, and S. Sastry, "Challenges for securing cyber physical systems," in *Workshop on future directions in cyber-physical systems security*, 2009.

[12] F. Pasqualetti, F. Dörfler, and F. Bullo, "Attack detection and identification in cyber-physical systems," *IEEE Transactions on Automatic Control*, vol. 58, no. 11, pp. 2715–2729, 2013.

[13] H. Fawzi, P. Tabuada, and S. Diggavi, "Secure estimation and control for cyber-physical systems under adversarial attacks," *IEEE Transactions on Automatic Control*, vol. 59, no. 6, pp. 1454–1467, 2014.

[14] S. Mishra, Y. Shoukry, N. Karamchandani, S. Diggavi, and P. Tabuada, "Secure state estimation: optimal guarantees against sensor attacks in the presence of noise," in *2015 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2015, pp. 2929–2933.

[15] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson, "Revealing stealthy attacks in control systems," in *Communication, Control, and Computing (Allerton), 2012 50th Annual Allerton Conference on*, Oct 2012, pp. 1806–1813.

[16] S. Gisdakis, T. Giannetsos, and P. Papadimitratos, "Shield: A data verification framework for participatory sensing systems," in *Proceedings of the 8th ACM Conference on Security & Privacy in Wireless and Mobile Networks*, ser. WiSec '15. New York, NY, USA: ACM, 2015. [Online]. Available: http://doi.acm.org/10.1145/2766498.2766503

[17] Y. Mo, R. Chabukswar, and B. Sinopoli, "Detecting integrity attacks on SCADA systems," *IEEE Transactions on Control Systems Technology*, vol. 22, no. 4, pp. 1396–1407, 2014.

[18] J. Valente and A. A. Cárdenas, "Using visual challenges to verify the integrity of security cameras," in *Proceedings of the 31st Annual Computer Security Applications Conference*, ser. ACSAC 2015. New York, NY, USA: ACM, 2015, pp. 141–150. [Online]. Available: http://doi.acm.org/10.1145/2818000.2818045

[19] Y. Shoukry, P. Martin, Y. Yona, S. Diggavi, and M. Srivastava, "Pycra: Physical challenge-response authentication for active sensors under spoofing attacks," in *Proceedings of the 22Nd ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '15. New York, NY, USA: ACM, 2015, pp. 1004–1015. [Online]. Available: http://doi.acm.org/10.1145/2810103.2813679

[20] "Secure control of an intelligent transportation system." [Online]. Available: https://youtu.be/qMSakEtkk_0

[21] C. L. Robinson, H.-J. Schutz, G. Baliga, and P. Kumar, "Architecture and algorithm for a laboratory vehicle collision avoidance system," in *2007 IEEE 22nd International Symposium on Intelligent Control*. IEEE, 2007, pp. 23–28.

[22] T. L. Lai and C. Z. Wei, "Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems," *The Annals of Statistics*, pp. 154–166, 1982.