ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/347390481

Transactions on Transportation Electrification Cyber-Attack Detection for Electric Vehicles Using Physics-Guided Machine Learning

Article *in* IEEE Transactions on Transportation Electrification · December 2020

DOI: 10.1109/TTE.2020.3044524

CITATIONS
21

READS
520

2 authors:

Lulu Guo
Tongji University
74 PUBLICATIONS **1,541** CITATIONS

SEE PROFILE

Bowen Yang
University of Georgia
23 PUBLICATIONS **280** CITATIONS

SEE PROFILE

All content following this page was uploaded by Lulu Guo on 16 December 2020.

The user has requested enhancement of the downloaded file.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TTE.2020.3044524, IEEE Transactions on Transportation Electrification

Cyber-Attack Detection for Electric Vehicles Using Physics-Guided Machine Learning

Lulu Guo, Jin Ye, *Senior Member, IEEE,* Bowen Yang

***Abstract*—With the continuous development of intellectualization and network interconnection, the cyber-physical security of power electronics systems has become increasingly prominent. Particularly, powertrain systems composed of one or more electric drives in modern electric vehicles (EVs) are becoming more vulnerable to cyber threats due to the connection to external networks in the intelligent traffic environment. In this paper, we design a physics-guided machine learning to detect cyber-attacks on EVs considering varying driving scenarios, which, to our knowledge, has not been attempted before. To reflect the transient physical characteristics of EVs, we collect both device-level (e.g., current and voltage in the motor drive) and vehicle-level signals. Then, innovative data features concerning critical system performance and physical dynamics of the vehicle are proposed, with which we leverage data-driven methodology with high-fidelity physical power electronics and vehicular models. Based on the physics-guided data features, a machine-learning-based classifier is developed and validated in an OPAL-RT hardware-in-the-loop (HIL) simulation testbed and demonstrates high accuracy under various driving scenarios.**

***Index Terms*—Electric vehicles, Cyber-physical security, Cyber-attack detection, Physics-guided machine learning, Powertrain system in electric vehicles.**

I. Introduction

THE Internet of Things (IoT) has led the way to smarter and ultimately more efficient cyber-physical systems, for instance, motor drives connected to a computer network controlling industrial processes [1]. However, despite the advantages of IoT-enabled applications, power electronics systems are becoming more vulnerable to cyber-attacks. An example of power electronics systems in IoT applications is the electric vehicle (EVs) composed of one or more electric drive systems (EDSs). In general, the EDS is connected to the in-vehicle controller area network (CAN). Compared to a traditional vehicle, connected cars present a greatly expanded attack surface due to the higher-level connectivity to the external network. This connectivity allows remote access to the attacker [2]–[7]. Once a car is affected by cyber-attacks, it may lead to a widespread influence other than the specific compromised vehicle. Following the publication of [8], [9], which reported how a Jeep Cherokee was remotely hacked and stopped on a highway, Chrysler issued a recall for 1.4 million vehicles that may be affected by a hackable software in 2015.

Manuscript received XXX, 2020; revised XXX, 2020; accepted XXX, 2020; online XXX, 2020. This work was supported in part by the National Science Foundation under Grant ECCS-1946057. (*Corresponding author: Jin Ye.)*

The authors are with the Intelligent Power Electronics and Electric Machine Laboratory, University of Georgia, Athens, GA 30602, USA (e-mail: lulu.guo@uga.edu, bowen.yang@uga.edu, jin.ye@uga.edu).

Aware of the cybersecurity problem, the automotive industry has attempted to develop security standards, such as the Society of Automotive Engineers (SAE) J3061 [10], International Organization for Standardization (ISO) 26262 [11], and committee draft of the "ISO-SAE Road Vehicles - Cybersecurity Engineering" standard [12]. In academia, many research works have been published in recent years to present the various threats, mitigation, and other possible countermeasures for vehicles [3], [13], [14]. With the advent of IoT and connectivity in the in-vehicle network, cybersecurity issues are coming into each component in an EV, including battery management, motor drives, braking, and steering. In [15], for the cybersecurity issues of the battery system in EVs, a framework for analysis, comparison, and test of standards is presented by identifying the key player in-vehicle cybersecurity. In addition, potential cyber-attacks on battery lifetime are also evaluated in [16], [17]. Similarly, considering cybersecurity vulnerabilities of the inter-vehicle network of EVs, cyber-attacks on electric drives can seriously impact motor current signature and cause performance degradation [18], [19].

To address the issue of cybersecurity in power electronic systems, authors in [1] demonstrated the potential challenges related to IoT-controlled power electronic systems. Hardware- and software-based intrusion-detection making use of the physical layer were considered preferred for the cybersecurity solution. On the one hand, information security, e.g., secure hardware, secure communication techniques, firewall, secure software update, etc., are the primary methods to prevent malicious attacks [20]. On the other hand, software-based intrusion detection that aims to design a reliable real-time monitoring system has been widely concerned, particularly for cyber-physical power systems [21]–[26].

In general, methods of cyber-attack detection can be classified into physics- and data-based strategies. The main work of physics-based detection [27] is to obtain the residual between the predicted system outputs and the measures, which is considered a proxy for the presence of attacks, e.g., research works in [28]–[30]. The prediction model can be developed from physical relationships like Newton's laws, fluid dynamics, electromagnetic laws, or an auto-regressive model [31], [32]. The detection methods are typically designed based on a linear dynamic system with sensor and perturbation noise, for instance, literature [33], [34]. Unlike physics-based solutions, data-driven based algorithms are model-free; thus, neither system parameters nor models are needed. Due to the advancements in computing technologies and the high potentials in identifying complicated cyber and physical attacks, data-driven approaches are gaining increasing attention in recent years [35]–[38].

2332-7782 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: University of Georgia. Downloaded on December 16,2020 at 22:18:35 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TTE.2020.3044524, IEEE Transactions on Transportation Electrification

Up to date, there are many data-driven methods for the security issues, such as geometrically designed residual filter, generalized likelihood ratio, leverage score, influential point selection, support vector machine (SVM), machine learning, deep learning, etc. In general, data-driven methods can be viewed as using trained models to detect abnormal system behaviors based on the observation data collected from the system, which is usually based on the idea that under normal conditions, the observation data would be constant with minor variations due to measurement inaccuracies and system noises. Commonly, labeling information is needed for supervised learning, and one can train a classifier to identify attacks according to the class labels (normal versus tampered) [39]–[41].

While the aforementioned physics and data-driven approaches provide necessary technical foundations, the cybersecurity of EVs has yet to be explored and poses several challenges: (1) The powertrain in an EV is a complicated and integrated system that includes many subsystems (e.g., battery system, energy management system (EMS), EDSs, etc.), in which nonlinear constraints and a large number of signal measurements need to be considered when identifying the presence of cyber-attacks. Then, physics-based approaches in current literature are not applicable because a simple linear dynamic system model cannot describe the vehicle control system. (2) Existing data-driven methods, however, have not yet addressed the physics characteristics of electric drives. The real-world driving conditions of vehicles change massively, even in normal circumstances. In contrast, in other applications, e.g., power grids, the sensor data in regular situations vary within a certain range (see Figs. 3-4 and the corresponding discussion in Section III). The specific feature of varying working conditions in vehicles may lead to training failures when designing a learning-based detector.

To address the two challenges, integrating physics and data-driven methods must be explored to detect cyber-attacks on electric drives and electric vehicles. Physics-guided machine learning emerges as a novel framework to leverage neural networks with physics-based models. In [42], the authors presented a physics-guided neural network by leveraging the output of a physics-based model along with observational features to generate predictions using a neural network architecture, and the effectiveness is illustrated in the problem of lake temperature modeling. In [43], a physics-guided recurrent neural network model is proposed, which combines the recurrent neural network and physics-based models to improve prediction accuracy. In [44], the authors concluded that physics knowledge and artificial intelligence could cooperate to build robust disruption avoidance systems for relevant fusion devices. In the cybersecurity domain, to detect the false data injection attacks in power grids, a physics-guided deep learning approach was proposed in [45]. The physics-guided deep learning outputs the estimated states by taking real-time measurements as inputs to neural networks and then reconstructs measurements considering power system physics. However, to our knowledge, physics-guided machine learning has not been attempted before to enhance the cyber-physical security of vehicles.

In response to the challenges of vehicle cybersecurity and inspired by the concept of physics-guided machine learning in power grids, we attempt to use both the physics model and machine learning to detect the cyber-attacks. This research will present how to utilize the knowledge about vehicular longitudinal dynamics and motor drive's model to improve the detection accuracy of machine learning. First, in addition to the device-level signals, e.g., current and voltage in the motor drive, we use vehicle-level signals reflecting transient vehicular states to design the cyber-attack detection. Second, innovative data features concerning critical system performance and physical characteristics of the vehicle are proposed, with which data-driven methodology with high-fidelity physical power electronics and vehicular models can be developed. Third, based on the proposed physics-guided data features, a machine learning-based classifier is developed and validated to improve training accuracy. Real-time monitoring results in an OPAL-RT hardware-in-the-loop (HIL) simulation testbed under the selected driving cycle that is excluded in the training process demonstrate that the proposed physics-guided cyber-attack detection can deal with varying driving scenarios. Finally, a case study to distinguish between malicious cyber-attacks and physical faults is also discussed.

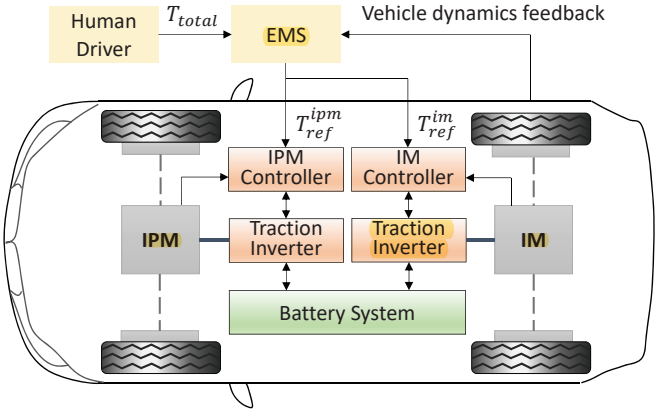


Fig. 1: System diagram of the powertrain system [46], [47].

II. Description of the Powertrain System and Cyber-Attack Modeling

A. Description of the Powertrain System

Fig. 1 presents a typical dual-motor driven EVs, which is one of the most used powertrain structures in automotive industry [46], [47], for instance, Tesla Model S on the market. This dual motor-driven topology provides a necessity for power distribution between the two electric drives. In this dual-motor driven EVs, the driver model is developed using a proportional-integral-derivative (PID) controller, which is designed to track the given velocity profile. The control variable of the PID is the total required torque, T_{total}. Then, the EMS transfers the driver's power/torque requirements to each motor's torque reference. As the focus of the paper is to identify cyber-attacks on the electric drives, the EMS is designed by a simple methodology that is widely used in the automotive industry -

2332-7782 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: University of Georgia. Downloaded on December 16,2020 at 22:18:35 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TTE.2020.3044524, IEEE Transactions on Transportation Electrification

distribute the torque on an equal basis to each machine. All of the signals are transmitted by the high-speed CAN buses, Local Interconnect Network, and FlexRay communication. For cyber-physical security study, we assume that the attacker can illegally get access to the in-vehicle communication buses, arbitrarily modify the sensor measurements, and hijack the device-level electric drives. The monitoring system obtains both the system-level signals (e.g., vehicle speed, torque reference, etc.) and device-level signals (e.g., voltage, current in the electric machine - interior permanent magnetic synchronous machine (IPMSM) and induction machine (IM)). Once the monitoring system identifies the threat, the driver will be alerted to cyber-attacks at an early stage.

In recent years, IPMSM has been widely used as the traction motor of EVs at the device level due to the high-power density and smooth torque production. As is shown in Fig. 2, the configuration of the IPMSM-based EDS includes both cyber and physical parts. In this configuration, we use a flux controller and the maximum torque per ampere (MTPA) to generate the pulse width modulation (PWM) signals. Based on the torque reference T_{ref}^{ipm} provided by the EMS, the current vector $[i_d, i_q]^T$ is derived to fulfill the torque requirement, which can be optimized by the MTPA algorithm. The PI control is used to regulate the tracking error of the output torque, d-axis, and q-axis current. Detailed procedures of the algorithm are described as in [18], [19].

B. Cyber-Attack Modeling in the IPMSM

1) Signals that are potentially compromised by a malicious attacker: Due to the communication between the local devices to the higher controllers in the vehicle control unit, a malicious attacker might also compromise the device-level motor drive. Some realistic examples of the most recent cyber-attacks on electric drive systems have demonstrated the vulnerability to cyber-attacks of electric drives. In March 2019, hackers in Tencent Keen Security Lab attacked Tesla's autopilot and manipulated the control of the vehicle that is powered by electric drive systems [48]. In August 2019, security researchers found a zero-day vulnerability in building HVAC, which are primarily electric drive systems [49]. Realistically, manufacturers may implant Trojans or malware in the controllers for data integrity attacks. Therefore, we consider a general electric drive system shown in Fig. 2, which are vulnerable to a variety of cyber-attacks.

Once the malicious attacker obtains access to the internal communication of the control system, each point of the control system may be attacked. However, in most cases, the sensor feedback signals, control sequences, and reference signals are most vulnerable to cyber-attacks, as their values are updated in very high frequency. Other signals like controller parameters are less vulnerable because they are relatively more static than the above three signals. Moreover, some of the controller parameters are stored in ROMs, which are difficult to be modified by network cyber-attacks. Therefore, in this paper, we consider three typical categories of cyber-attacks based on the different attack locations: sensor (location A: three-phase current from the IPM drive), controller (location C: the output of the PI controller), and the communication channel with the EMS (location B: torque reference from the higher-level controller).

Although we use a limited number of attack cases, these attack scenarios can reflect a typical dynamic response. In real applications, more cases should be defined for practicability. Because the paper focuses on developing a detection algorithm, the data collected from the defined cyber-attack scenarios is considered enough to validate the overall performance of the proposed method. Without loss of generality, we conduct cyber-attacks on i_a, T_{ref}^{ipm}, and v_d. Here, i_a represents the a-phase current; T_{ref}^{ipm} is the torque reference of the IPMSM; v_d is the voltage of the d-axis.

2) Attack modeling: Generally speaking, cyber-attacks can be categorized into three types from a control system's perspective: denial of service attacks, replay attacks, and false data injection attacks [50]. Denial of service attacks typically attempt to make the resources unavailable; replay attacks repeat and send the set of past information to the network; and data injection attacks can directly falsify measurements or inject incorrect instructions to the system, which can be expressed in many forms, e.g., scaling and addictive attacks [51], high-frequency harmonics and periodic pulse injection [18]. Among these attacks, denial of service attacks are relatively easy to identify because it remains constant with attack duration, which can be identified through the obtained waveforms. For example, once the other states and their references are changing while the collected signal data remains constant within a period, then the corresponding signal is possibly affected by a denial of service attack. In the paper, we consider the other two cyber-attacks: replay and false data injection attacks.

In the false data injection attacks, we suppose the time horizon under attack is $[t_{atk}, t_{atk} + \mathcal{T}_{atk}]$, then the signal affected by a false data injection can be expressed as

$$\bar{y} = \begin{cases} \nu \cdot y(t), & \text{if } t \in [t_{atk}, t_{atk} + \mathcal{T}_{atk}] \\ y(t), & \text{else} \end{cases} \tag{1}$$

where y represents the target signal, and ν is attack coefficient. Specific definitions of the false data injection attacks are given in Table I. In replay attacks, y is firstly recorded ahead of time by the attacker, and during the attack interval, the sensor measurements are obtained by repeating the recorded values, as $\bar{y} \in \mathbf{Y}$, where $\mathbf{Y}$ is the set of past information. For clear expression, we denote the start time and the duration of data record as t_{atk}^{rec} and $\mathcal{T}_{atk}^{rec}$, respectively. Then, the time horizon under attack is $[t_{atk}^{rec} + \mathcal{T}_{atk}^{rec}, t_{atk}^{rec} + 2\mathcal{T}_{atk}^{rec}]$. Without loss of generality, we set $\mathcal{T}_{atk}^{rec} = \mathcal{T}_{atk} = 5$s. Finally, replay attacks with different targets and time settings are summarized in Table II. Notice that the attack coefficients ν, $\mathcal{T}_{atk}$, and t_{atk}, have a significant influence on the results despite the same attack formula. It is expected that the higher level of modification to the signal measurements, the higher damage it would cause. For a fair comparison between attack types and cases, all of the intensity of attacks is the same; the maximum magnitude is limited to $\pm 25\%$ of the original value.

2332-7782 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: University of Georgia. Downloaded on December 16,2020 at 22:18:35 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TTE.2020.3044524, IEEE Transactions on Transportation Electrification

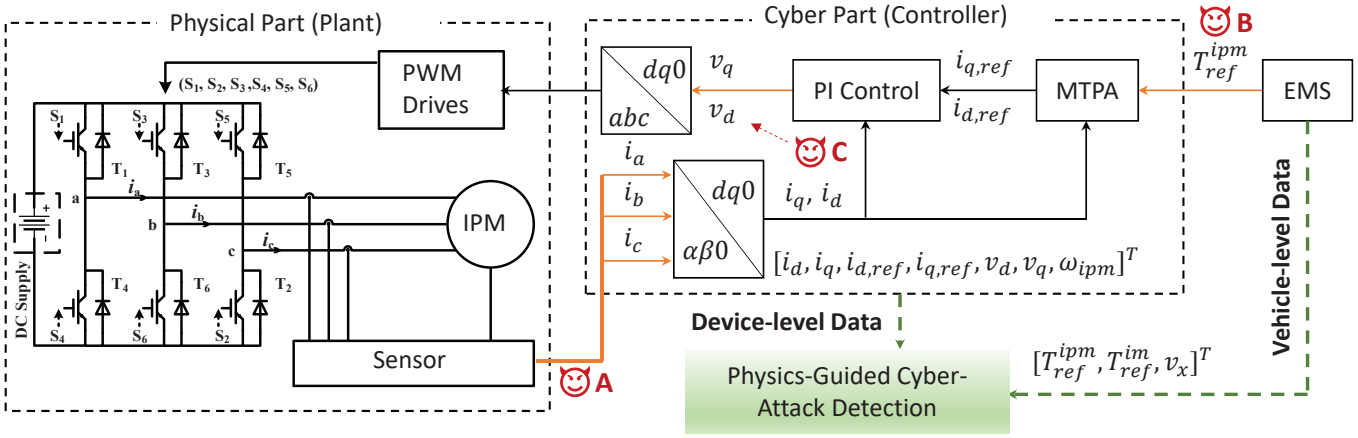


Fig. 2: Schematic diagram of the IPMSM drive, and signals that are potentially affected by a malicious attacker.

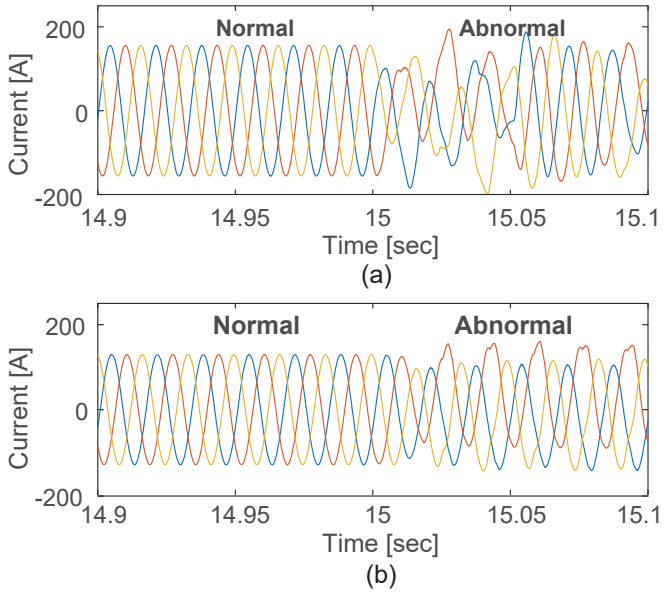


Fig. 3: Waveform in an IEEE 37-node distributed power grid with distributed energy resources. (a) shows the results of false data injection attack; (b) shows a PWM open circuit fault in the distributed energy resource.

TABLE I: Definitions of False Data Injection Attacks.

No.	Target	ν	t_{atk}	No.	Target	ν	t_{atk}
1	i_a	1.2	152	10	v_d	-0.75	98
2	i_a	1.2	204	11	v_d	-0.75	21
3	i_a	1.2	108	12	v_d	-0.75	204
4	i_a	-0.75	360	13	T_{ref}^{ipm}	-1.2	71
5	i_a	-0.75	77	14	T_{ref}^{ipm}	-1.2	92
6	i_a	-0.75	163	15	T_{ref}^{ipm}	-1.2	104
7	v_d	-1.2	184	16	T_{ref}^{ipm}	-0.75	47
8	v_d	-1.2	235	17	T_{ref}^{ipm}	-0.75	28
9	v_d	-1.2	172	18	T_{ref}^{ipm}	-0.75	206

TABLE II: Definitions of Replay Attacks.

No.	Target	t_{atk}^{rec}	No.	Target	t_{atk}^{rec}
19	i_a	72	23	v_d	103
20	i_a	103	24	v_d	181
21	i_a	181	25	T_{ref}^{ipm}	72
22	v_d	72	26	T_{ref}^{ipm}	181

III. Physics-Guided Machine Learning Cyber-Attack Detection

In this section, we will propose a physics-guided machine learning cyber-attack detection, which fully utilizes prior scientific knowledge of control systems. As shown in Fig. 2, the data used in the proposed physics-guided cyber-attack detection include device-level and vehicle-level signals, expressed as

$$X_{device} = [i_d, i_q, i_{d,ref}, i_{q,ref}, v_d, v_q, \omega_{ipm}]^T \quad (2a)$$

$$X_{vehicle} = [T_{ref}^{ipm}, T_{ref}^{im}, v_x]^T \quad (2b)$$

respectively, where i_d is the d-axis current; i_q is the q-axis current; $i_{d,ref}$ is the reference of i_d; $i_{q,ref}$ is the reference of i_q; v_d and v_q represent the voltage of the d-axis and q-axis, respectively; ω_{ipm} is the electrical angular speed of the IPMSM; T_{ref}^{ipm} is the torque reference of the IPMSM; T_{ref}^{im} is the torque reference of the IM; and v_x is the vehicle speed. Then, the data vector used in the detection is set to

$$X = [X_{device}; X_{vehicle}]. \quad (3)$$

In real-time applications, the moving horizon split-window is used to generate the raw data matrix. The window size is set to N_w with a sampling time Δt. The raw data matrix X_k is formulated as $X(k) = [X(k - N_w + 1), X(k - N_w + 2), ..., X(k)]$, where k represents the current time instant.

A simple detection methodology is to use the collected raw data to train a classifier network. However, different from other research works, e.g., power grids [52], real-time driving conditions of vehicles may significantly vary with time. This feature makes it difficult for anomaly identification by using raw data and machine learning networks alone, compared to the research domain of the power grid. For example, Fig. 3 shows an electrical waveform in a 37-node power grid with distributed

2332-7782 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: University of Georgia. Downloaded on December 16,2020 at 22:18:35 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TTE.2020.3044524, IEEE Transactions on Transportation Electrification

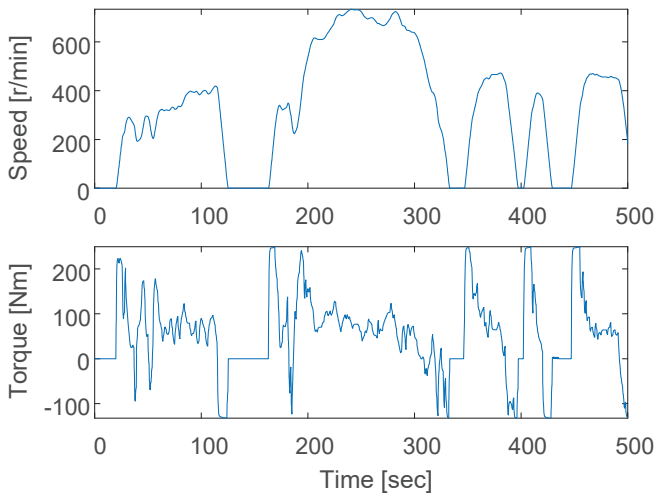


Fig. 4: Normal conditions of electric machines in urban driving.

energy resources due to fault signal propagation. The waveform in this figure shows a regular time-varying feature and indicates that the fault can directly be reflected by frequency, magnitude, and phase. In such a case, time-frequency domain data or raw data can be fed to a learning network and obtain a desirable result. Conversely, the torque and speed in Fig. 4 demonstrates high uncertainty and a broader range of variation (in normal cases). Therefore, for those cyber-attacks causing changes in data profiles, it is difficult to distinguish abnormal conditions and varying driving conditions, such as frequent start-stop driving scenarios in urban traffic. To address this issue, we propose a physics-guided cyber-attack detection method that ingests physic-guided data features with the raw data.

A. Physics-guided data features reflecting critical system performance and physics dynamics

As shown in Fig. 5, two categories of data features are developed: tracking accuracy and dynamic residuals. The tracking accuracy emphasizes the damage caused by malicious behavior in terms of tracking performance. We define the tracking accuracy as the first category of data feature as follows:

$$\widetilde{x}_{Id}(k) = |i_d(k) - i_{d,ref}(k)| \tag{4a}$$

$$\widetilde{x}_{Iq}(k) = |i_q(k) - i_{q,ref}(k)| \tag{4b}$$

$$\widetilde{x}_T(k) = \frac{|\hat{T}_{ipm}(k) - T^{ipm}_{ref}(k)|}{|T^{ipm}_{ref}(k)| + \epsilon} \tag{4c}$$

where $\widetilde{x}_T$ represents a relative error of tracking the torque reference; $\epsilon > 0$ is introduced to deal with the case of $T^{ipm}_{ref} = 0$, and without loss of generality, we set it to 10Nm in the paper; $\hat{T}_{ipm}$ represents the estimated output torque, which can be calculated by i_d and i_q [18], expressed as

$$\hat{T}_{ipm} = \frac{3}{2} p[\Phi_{pm} i_q + (L_d - L_q) i_d i_q]. \tag{5}$$

Here p is the number of pole pairs; L_d and L_q are the inductance of d-axis and q-axis, which are functions of i_d and i_q, as $L_d = \mathcal{M}_d(i_d, i_q)$, $L_q = \mathcal{M}_q(i_d, i_q)$; Φ_{pm} is the flux linkage of permanent magnet.

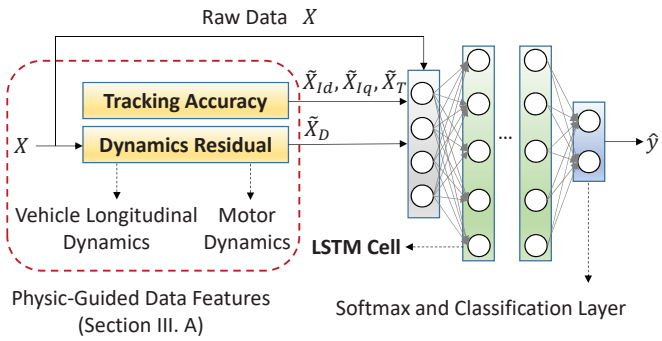


Fig. 5: Physics-based network architectures that ingest the developed physics-guided data features in Section III.A.

The second category of data features describes the residual between the predicted variables through physics modeling and the collected sensor measurements. According to the prior knowledge of the EDS and longitudinal vehicle dynamics, we can predict the response of the system. Although the model is established under simplifying assumptions, they can reflect the critical system components; hence, residuals between the predicted and obtained values remain within a specific range in normal conditions. Suppose the vector of predicted values at k^{th} time instance is $\hat{X}_p(k)$, and the corresponding signals is $X_p(k)$, which is a subset of the collected data, as $X_p(k) \subset X(k)$. Then, according to the static and transient physics-based relationship of these signals, there is

$$\hat{X}_p(k) = \mathbf{G}(X(k), X(k-1), \xi(k-1), \xi(k)) \tag{6}$$

where $\mathbf{G}$ represents the corresponding dynamics equations; ξ represents the set of system parameters. Then, the second category of data feature is defined by

$$\widetilde{X}_D(k) = \|\hat{X}_p(k) - X_p(k)\|_Q, \tag{7}$$

where Q represents the weight. Once the system is affected by cyber-attacks leading to abnormal dynamics structure, parameters, control inputs, and system outputs, there will be a considerable increase in the residual. Specifically, for the investigated vehicle and IPM system, the dynamic residuals are defined as follows.

1) Predicted variable obtained from longitudinal vehicle dynamics: The first dynamic residual is developed based on the longitudinal vehicle dynamics [53], as

$$\hat{v}_x(k) = v_x(k-1) + [\frac{T_w(k-1)}{r_w M_{veh}} + \kappa(k-1)]\Delta t \tag{8}$$

with $T_w = \hat{T}_{ipm} i_{g,ipm} + T_{im} i_{g,im}$, where T_w represents the total wheel torque; v_x represents the vehicle speed; r_w is the wheel radius; M_{veh} is the vehicle mass. In the above equation, κ denotes the longitudinal acceleration due to resistance:

$$\kappa = \frac{1}{2} c_{\text{d}} \rho A_f v^2 + g f \cos(\theta) + g \sin(\theta) \tag{9}$$

where c_{d}, ρ, A_{f} are the air resistance coefficient, air density, and face area, respectively; g is gravitational constant; f is the

2332-7782 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: University of Georgia. Downloaded on December 16,2020 at 22:18:35 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TTE.2020.3044524, IEEE Transactions on Transportation Electrification

rolling resistance coefficient; and θ is the road slope. Here, θ varies with the traveling distance and influences the modeling accuracy. Therefore, in real applications, the road slope needs to be known via an online estimator, which has been well studied, for instance, [54]–[57]. Notice that the obtained residual is not directly used for detection, but is one of the inputs of the learning network. In most attack scenarios, there is a mutation in the dynamic residuals (see Fig. 11), while the road slope is typically a continuous and slow-varying variable, especially at the millisecond level. Therefore, uncertainty due to inaccurate road slope estimation would have little influence on the results. When the torque reference T_{ref}^{ipm} is tampered by malicious attackers, it is obvious that using signal data of EDS alone is inadequate to identify the cyber-attack. In such a case, the obtained $\hat{v}_x(k)$ in (8) may be useful to reflect the abnormal dynamics in the system level.

2) Predicted variable obtained from motor dynamics: In the IPM, the predicted signals are calculated by

$$\hat{v}_d(k) = R_s i_d(k) - \omega_{ipm} L_q i_q(k) \tag{10a}$$
$$\hat{v}_q(k) = R_s i_q(k) + \omega_{ipm}(L_d i_q(k) + \Phi_{pm}) \tag{10b}$$

where R_s is the equivalent winding resistance, and ω_{ipm} is the electrical angular speed. In consequence, the vector of predicted values is formulated as

$$\hat{X}_p(k) = [\hat{v}_x(k), \hat{v}_d(k), \hat{v}_q(k)]^T \tag{11}$$

and the dynamics residual is calculated by (7), where $X_p(k) = [v_x(k), v_d(k), v_q(k)]^T$ is used.

Finally, the inputs are reformulated as $\widetilde{X} = [X^T, \widetilde{X}_{Id}, \widetilde{X}_{Iq}, \widetilde{X}_T, \widetilde{X}_D]^T$. Based on the redefined training data, a classification method is used to detect the cyber-attack with a long short-term memory network (LSTM).

B. Classification using LSTM

As an extended version of recurrent neural networks, LSTM has been used widely in many domains. By replacing nodes in the recurrent neural network with memory cells and gating mechanism [58], [59], LSTM can effectively handle long-term dependencies of data. Compared with the conventional machine learning models, LSTM can better capture the dynamic features and temporal patterns of the streaming data, which is an obvious advantage for the time-series sensor data analytics. Briefly, similarly to system dynamics, the LSTM architecture can establish the relationship between the time-series data sets. This relationship can help capture the overall system state, and hence are beneficial for distinguishing between normal and abnormal scenarios.

In essence, the LSTM model defines a transition relationship for hidden representation through an LSTM cell. The basic architecture of an LSTM memory cell is illustrated in Fig. 6, in which forget, input, output gates are used to solve the issue of vanishing gradients in an recurrent neural network. Specially, given an input x_t and the previous timestamp output h_t, at each time step, an LSTM first generates a candidate cell state $\widetilde{c}_t$ with h_t and x_t, as $\widetilde{c}_t = \tanh(W_h^c h_{t-1} + W_x^c x_t)$, and calculates the forget gate, input gate, and output gate by [43], [60]

$$f_t = \sigma(W_h^f h_{t-1} + W_x^f x_t) \tag{12a}$$
$$i_t = \sigma(W_h^i h_{t-1} + W_x^i x_t) \tag{12b}$$
$$o_t = \sigma(W_h^o h_{t-1} + W_x^o x_t) \tag{12c}$$

respectively, where W_h^f, W_x^f, W_h^i, W_x^i, W_h^o, W_x^o are weights. Then, the new cell state and the hidden representation are obtained by $c_t = f_t \otimes c_{t-1} + i_t \otimes \widetilde{c}_t$, $h_t = o_t \otimes \tanh(c_t)$, where $\otimes$ denotes the entry-wise product.

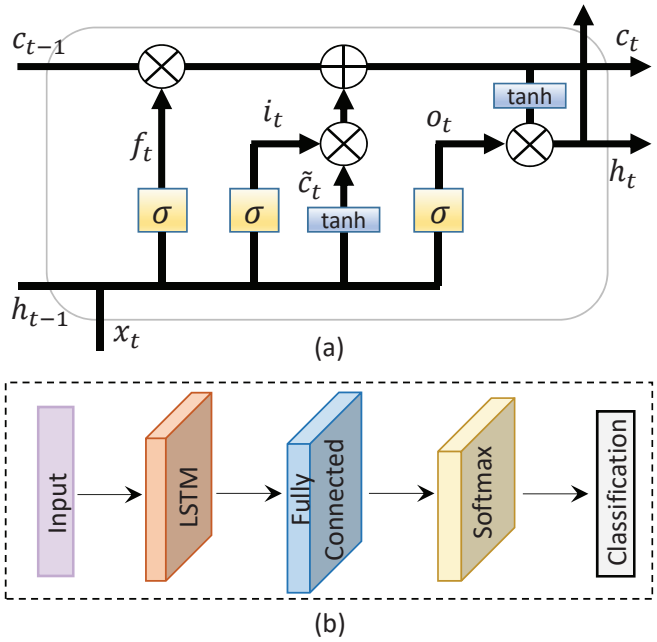


Fig. 6: (a) structure of a basic LSTM cell; (b) architecture of the proposed LSTM network.

For the problem of cyber-attack detection, we design a simple LSTM network. The diagram of the architecture is shown in Fig.6(b). To predict class labels, the network ends with a fully connected layer, a softmax layer, and a classification output layer. The softmax layer is used to normalize the output predictions of the model to be a valid probability distribution, expressed as $\hat{y}_j = e^{z_j(i)}/[\Sigma_{j=1}^{n_c} e^{z_j^i}]$, where j denotes the generic neuron of the fully connected layer; n_c is the number of target classes; i refers to the generic training example; z is the output value of the fully connected layer. Finally, a classification layer is adopted by using the cross entropy error to calculate the cost function, as

$$\min_W \; J = -\frac{1}{M}\sum_{i=1}^{M}\sum_{j=1}^{n_c} y_j^i \cdot \log(\hat{y}_j^i) \tag{13}$$

where M is the number of training examples; y represents the true label; $\hat{y}$ is the predicted value; W represents the sets to be optimized in the learning network, including W_h^f, W_x^f, W_h^i, W_x^i, W_h^o, W_x^o in each LSTM cell, and parameters in the fully connected layer. For the issue of intrusion detection, n_c is set to two: 0 and 1, which represent normal and abnormal conditions. Specifically, hyper parameters for the LSTM models are given as follows: batch size = 120, learning rate = 0.001, hidden size

2332-7782 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: University of Georgia. Downloaded on December 16,2020 at 22:18:35 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TTE.2020.3044524, IEEE Transactions on Transportation Electrification

= 32, optimizer = Adam, number of layers = 2 (one LSTM layer and one fully connected layer). Based on the above hyper parameters, the LSTM model is created by using TensorFlow.

IV. Simulation and Performance Evaluation

In this section, we present the evaluation results of the proposed cyber-attack detection methodology. A high-accurate dual-motor based EV powertrain is established on the MATLAB-Simulink platform, which can be conducted in real-time in an OPAL-RT, as is shown in Fig. 7. The platform is constructed based on the mathematical model and includes a 50kW IPMSM in the front axle and a 60kW IM in the rear axle. In the vehicle modeling, we consider several nonlinear dynamics, such as tire model, wheel slip ratio during driving, road slope, etc. The IPMSM controller and mechanical system are simulated with a sampling time of 25μs. The electric machines and power electronics devices are simulated with a sampling time of 0.5μs. Then, we conducted the simulation under standard driving cycles: New European Driving Cycle and Urban Dynamometer Driving Schedule (UDDS), which are supposed to represent the typical usage of a car in Europe and the United States, respectively. Then, we can obtain a large number of databases of the system under different circumstances. The data set is split into a training (90% of data) set and a test (10%) set in the training process.

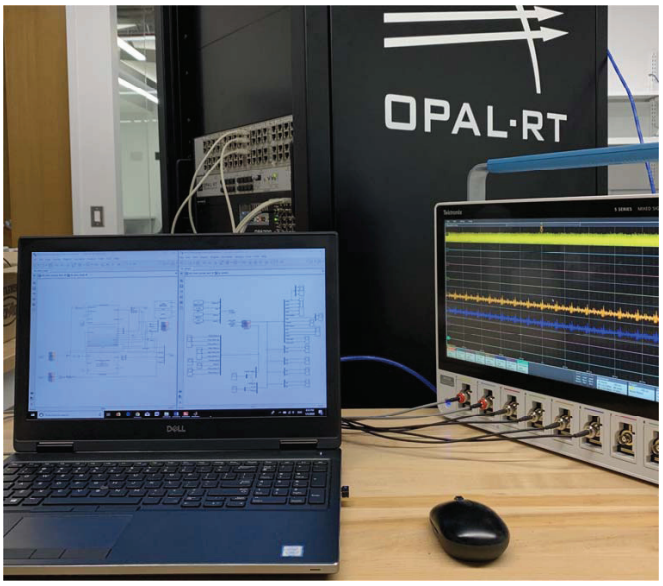


Fig. 7: OPAL-RT HIL real-time simulation testbed.

To evaluate the performance of the recurrent classifiers, several metrics are typically used in literature, for instance, accuracy, precision, recall, F1-score, etc., as detailed discussed in [61]. In this section, we use accuracy to present the overall effectiveness, which is defined as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{14}$$

where TP (true positive), TN (true negative), FP (false positive), and FN (false negative) denote the number of examples that actual attack is correctly identified as attack, actual normal is correctly identified as normal, actual attack is wrongly identified as normal, and actual normal is wrongly identified as attack, respectively.

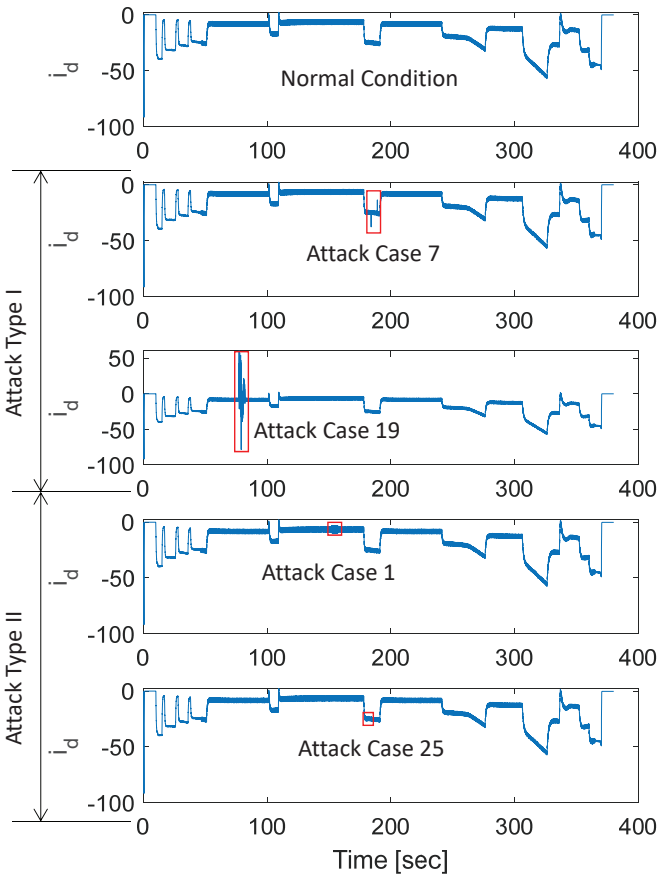


Fig. 8: The current i_d [A] in different attack cases. (I) and (II) respectively represent the result of the first and the second attack types according to their influence on the control system.

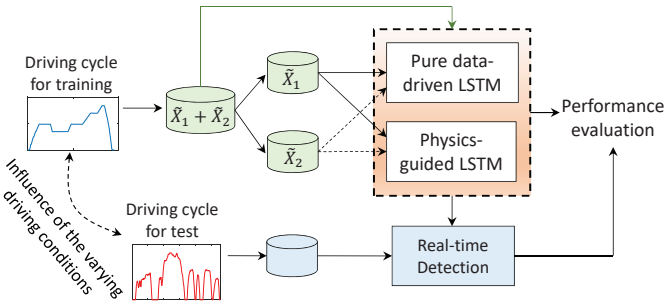


Fig. 9: The main idea of the performance evaluation of the proposed detection method.

A. Effectiveness of the introduced physic-guided features

For a detailed analysis of the introduced physics-based data features, we divide the cyber-attacks into two categories based on their effect on the signal waveforms. The first category may cause a significant ripple, impulsion, or even stability of the system, which can be obviously identified. The second type of attack can indirectly influence the IPMSM;

2332-7782 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: University of Georgia. Downloaded on December 16,2020 at 22:18:35 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TTE.2020.3044524, IEEE Transactions on Transportation Electrification

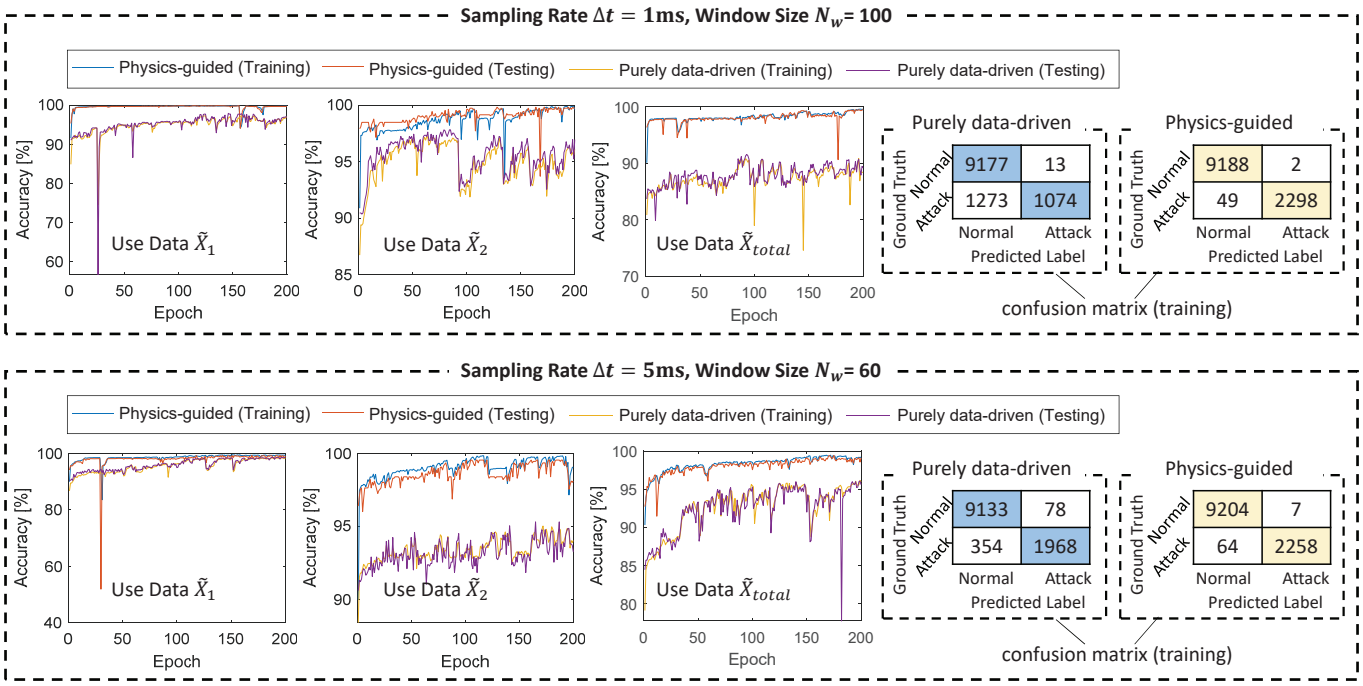


Fig. 10: Accuracy curves in training process under two conditions. The confusion matrix represents the training results of purely data-driven and physics-guided LSTM methods by using data $\widetilde{X}_{total}$.

therefore, in-apparent results are reflected in the obtained signal measurements (see Fig. 8). The corresponding training data of the two attack types are denoted as $\widetilde{X}_1$ and $\widetilde{X}_2$, respectively.

From the perspective of physical meaning, we note that: 1) Due to the stability and robustness of the controller, even though the feedback signal is modified maliciously, to a certain extent, the system can recover to its nominal condition when the cyber-attack is withdrawn. Then, for those false data integrity attacks $\nu > 0$, the influence of these attacks is relatively smaller than those attacks causing positive feedback. 2) Replay attacks on either i_a or v_d could easily lead to an imbalance between the three-phase current, which can be reflected in the signal measurements (see Fig. 8). 3) Attacks on torque reference have little influence on the lower-level EDS from the perspective of regular working states. In other words, although the reference is affected and revised by the malicious attacker, the lower-level EDS can work properly. Thus, it is impossible to distinguish between a cyber-attack and an abrupt change in torque requirement from the driver if we only use the EDS waveforms.

According to the analysis and attack results, the first training data set $\widetilde{X}_1$ (create a significant influence on the system) include Cases 4-12 and Cases 19-24; the second training data set $\widetilde{X}_2$ include Cases 1-3, Cases 13-18, and Cases 25-26. Then, by the comparison between the results of $\widetilde{X}_1$ and $\widetilde{X}_2$, we can evaluate the effectiveness of the proposed method. The main idea of the comparison is shown in Fig. 9. For fair comparison, the results under two conditions are used: $\{\Delta t = 1\text{ms}, N_w = 100\}$ and $\{\Delta t = 5\text{ms}, N_w = 60\}$. A detailed discussion of the choice of Δt and N_w is given in Section IV. C.

Fig. 10 shows the training process of the purely data-driven (only the raw data is used) and physics-guided LSTM networks. From the results, we can observe that the proposed physics-guided data features can significantly improve the convergence speed and training accuracy. More specifically, when only the second type of attacks (data $\widetilde{X}_2$) are used, the training accuracy would decrease because the impact of cyber-attacks is unnoticeable (see Case 25 - replay attacks on T_{ref}^{ipm}). In such a case, it is difficult to identify attacks with raw data only. Simultaneously, the physics-guided LSTM displays better performance, validating the effectiveness of the developed data features. The physics-guided data features obtained from vehicle and EDS dynamics present a good observation of the system's performance. For example, once the torque reference is compromised, the modified T_{ref}^{ipm}, for the device-level IPM controller, is still a normal signal and can be tracked well. In this case, although the actual output torque cannot fulfill the driver's requirement, the system is considered healthy. Therefore, $\mathbf{G}$ and $\widetilde{X}_T$ are essential to fill the gap between the actual torque response and its desired value from the higher controller.

Finally, by using both the two training data sets, denoted as $\widetilde{X}_{total} = \widetilde{X}_1 + \widetilde{X}_2$, we show the comprehensive results in Fig. 10. Detailed training and testing accuracy are summarized in Table III. We can conclude that the proposed physics-guided data features can help to improve detection performance. Besides, for better performance evaluation, a series of traditional classifiers are designed, including decision tree (DT), support vector machine (SVM), K-nearest neighbor (KNN), and naive Bayes classifier (NB). The results are shown in Table IV ($\widetilde{X}_{total}$ is used). In these traditional machine learning algorithms, both purely data-driven or physical-guided methods are used (see the first column of Table IV). The results indicate that the

2332-7782 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: University of Georgia. Downloaded on December 16,2020 at 22:18:35 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TTE.2020.3044524, IEEE Transactions on Transportation Electrification

proposed physics-guided LSTM achieves the best performance. This is because, compared to the traditional algorithms, LSTM can extract deep features from time-series data, which can better reflect the system states and identify the cyber-attacks.

B. Real-time cyber-attack detection results

To evaluate the practicability of the trained network in real-world applications, we will observe the cyber-attack detector's performance under one other driving cycle that has not been used in the training process. For this purpose, we choose a part of typical urban speed profiles - UDDS as the test driving condition, as shown in Fig. 11. In the testing driving cycle, several attack events are designed, and the well-trained physics-guided LSTM network monitors the system in real time. The detection results are shown in Fig. 11, illustrating that the proposed physics-guided LSTM can deal with varying working conditions. The reason is that physics modeling can effectively extract the vehicle features and thus can reflect abnormal circumstances.

Notice that, compared to normal scenarios, the real-time detection accuracy in attack cases is much lower, indicating an excessive sensitivity of the developed monitor. Therefore, further work is needed to adjust its sensitivity according to the varying driving conditions. Besides, as shown in the area marked by a red circle in Fig. 11, although the cyber-attacks are withdrawn, adverse effects on the control system may continue for a long time before the car is stopped, which may reduce the accuracy of detection results. Although the detection accuracy is acceptable, in-world applications, occasional misbehavior in detecting the cyber-attacks are not acceptable for human drivers, despite the low probability. One of the solutions is to revise the decision law by using an average result, as

$$y_{final,k} = 1, \text{ only if } \hat{y}_{k-2} = \hat{y}_{k-1} = \hat{y}_k = 1, \qquad (15)$$

where $y_{final,k}$ represents the final decision at time instance k; $\hat{y}_{k-2}$, $\hat{y}_{k-1}$, and $\hat{y}_k$ represent the corresponding results at the time instances $k-2$, $k-1$, and k, respectively.

C. Detection performance under different sampling rate and window size

Notice that sampling rate Δt and window size N_w may influence the performance of attack detection. Then, in this subsection, we present the detection results under different sampling rate ([1ms, 20ms]) and window size, to further characterize the model sensitivity and provide guidance for the choice of window size and sampling rate in practice. This time level of sampling rate is totally realizable in the CAN bus of a real vehicle, for instance, raw acceleration measured by the accelerometers (100Hz [62]), longitudinal vehicle (up to 5kHz [63]), 200Hz sampling rate in a wheel slip control [64], etc. For fair comparison, all of the LSTM models are trained under the same structure and parameters. To evaluate the highest accuracy of this proposed detection method, $\widetilde{X}_{total}$ and physics-guided method are used. The results are shown in Table V. The bold numbers in the table (for $\Delta t = 1$ms and $\Delta t = 5$ms) are the largest values with comprehensive consideration of training and testing accuracy, and the detailed

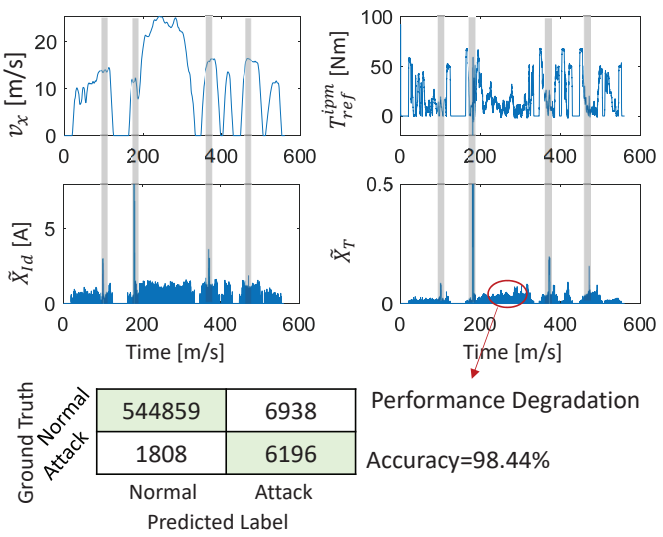


Fig. 11: Detection results in the UDDS driving cycle. The cyber-attacks are marked with gray bars, and from left to right, they represent attack cases 1, 4, 13, and 7.

training process is shown in Fig. 10. In each sampling rate, the maximum time length of moving horizon split-window is set to $N_w \Delta t = 1$s. We can see that the proposed physics-guided method is insensitive to sampling rate and window size from the results. Overall, the reasonable time length of the split window should be $N_w \Delta t \in [0.1, 0.6]$s. With the increasing window size, the accuracy would be relatively lower, especially for long sampling time.

D. Preliminary discussion on distinguishing between malicious cyber-attacks and physical faults

When it comes to cybersecurity, a broad concern is distinguishing between malicious cyber-attacks and physical faults. However, for most cyber-physical control systems, it is quite challenging to diagnose threats because most of the systems are closed-loop control. In current literature that focuses on fault diagnosis, a physical fault can be directly reflected by the distortion of the mapping from $\widetilde{u}$ to y. If malicious cyber-attacks are not considered, then the accurate control inputs $\widetilde{u} = u$, the outputs $\widetilde{y} = y$, and their physics-based relationship provide a sufficient basis to identify a physical fault. But when considering the cyber-attacks that may cause inaccurate input data in the monitoring systems, e.g., $\widetilde{u} \neq u$ and $\widetilde{y} \neq y$, this method would be unrealistic because they both can lead to unmatched behavior between the nominal physics model and the obtained data.

To observe the difference between malicious cyber-attacks and physical faults, we present an open circuit fault of the IPM traction inverter, which is one of the most common faults in motor drives. The fault event is emulated by setting the switching signal for one of the insulated-gate bipolar transistors in IPM traction inverter to zero. Then, the obtained i_d is shown in Fig. 12, from which we can see that i_d increases dramatically once the fault is activated. Comparing to the results of cyber-attacks in Fig. 8, we can observe that most of the cyber-attacks,

2332-7782 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: University of Georgia. Downloaded on December 16,2020 at 22:18:35 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TTE.2020.3044524, IEEE Transactions on Transportation Electrification

TABLE III: Detection Results [Training, Testing] (%).

Data Set	$\Delta t = 1$ms, $N_w = 100$		$\Delta t = 5$ms, $N_w = 60$	
	Physics-guided LSTM	Purely Data-driven LSTM	Physics-guided LSTM	Purely Data-driven LSTM
$\widetilde{X}_1$	[99.64, 99.67]	[96.95, 96.89]	[99.15, 99.02]	[98.33, 98.47]
$\widetilde{X}_2$	[99.59, 99.71]	[96.09, 96.94]	[99.05, 98.09]	[94.02, 93.24]
$\widetilde{X}_{total}$	[99.56, 99.53]	[88.85, 89.23]	[99.38, 99.14]	[96.25, 95.63]

TABLE IV: Results of Different Classifiers (%).

Method	Sampling Rate $\Delta t = 1$ms				Sampling Rate $\Delta t = 5$ms			
	DT	SVM	KNN	NB	DT	SVM	KNN	NB
Purely data-driven method	92.5	88.0	91.2	75.8	92.7	87.2	90.9	75.9
Physics-guided method	93.2	91.5	92.1	89.9	93.3	91.0	91.2	90.0

TABLE V: Results of physics-guided LSTM under sampling rate and window size [Training,Testing (%)] (Epoch=200).

$\Delta t = 1$ms	$N_w = 50$	$N_w = 100$	$N_w = 200$	$N_w = 400$	$N_w = 600$	$N_w = 800$	$N_w = 1000$
	[98.73, 99.06]	**[99.56, 99.53]**	[99.02, 99.14]	[99.12, 99.22]	[99.11, 98.99]	[98.36, 98.27]	[99.20, 99.06]
$\Delta t = 5$ms	$N_w = 20$	$N_w = 40$	$N_w = 60$	$N_w = 80$	$N_w = 100$	$N_w = 150$	$N_w = 200$
	[99.64, 98.52]	[98.96, 98.13]	**[99.38, 99.14]**	[98.32, 98.91]	[98.77, 98.28]	[98.88, 99.08]	[99.91, 99.06]
$\Delta t = 10$ms	$N_w = 10$	$N_w = 20$	$N_w = 30$	$N_w = 40$	$N_w = 50$	$N_w = 80$	$N_w = 100$
	[99.19, 96.64]	[99.20, 98.67]	[99.20, 98.44]	[99.02, 98.12]	[99.40, 99.14]	[98.28, 99.14]	[98.47, 97.73]
$\Delta t = 20$ms	$N_w = 5$	$N_w = 10$	$N_w = 15$	$N_w = 20$	$N_w = 30$	$N_w = 40$	$N_w = 50$
	[98.36, 97.03]	[99.18, 97.50]	[99.48, 99.22]	[99.37, 98.52]	[99.35, 99.30]	[99.30, 98.44]	[99.08, 98.75]

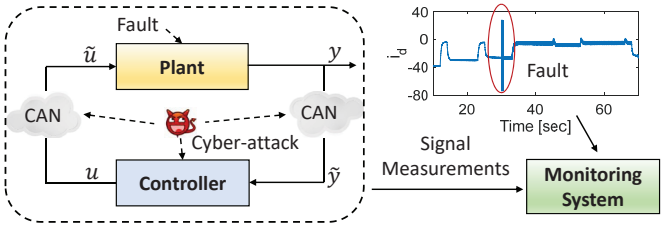


Fig. 12: Cyber-attacks and physical faults in a control system.

e.g., replay attacks, torque reference attacks, and several kinds of false data injection attacks ($\nu > 0$), have less impact on the system. This distinction may help to distinguish physical faults and cyber-attacks.

Considering the fault in the converter, which may have different fault duration and amplitude, we present several fault scenarios, as shown in Fig. 13. Results of a cyber-attack are also given for comparison (Attack Case 19). In this figure, physical faults 1 to 3 represent winding grounded short circuit fault on phase A, phase A & B, and phase A & B & C, respectively. Physical fault 4 represents an open circuit fault in upper switch of phase A. Overall speaking, the D-axis current profiles have severe distortion and oscillation for both cyber-attacks and physical faults, but the frequency of the oscillation is different.

While the oscillation and distortion patterns due to cyber-attacks are considerably random, the ones due to physical faults show specific regular variation because faults often have a fixed physical model, such as short circuit faults. In particular, we can see that despite the physical faults, the i_d still presents a fixed frequency characteristics. For one type of physical fault, the amplitude is related to the physical parameters of that fault and shows a stable and regular waveform. This feature may provide an aspect to distinguish the cyber-attacks and physical faults. Moreover, for those physical faults causing gradual performance degradation, such as increasing internal resistance, specific physic characteristics should be utilized to address the long-term abnormal behavior. This kind of persistent rule of performance degradation may also be used to distinguish faults from cyber-attacks.

Comparing the load changes and the cyber-attacks, the patterns of the current profiles are quite different, as shown in Fig. 14. The current variation due to load step change has a limited changing rate due to speed and current controllers, while the current distortion due to cyber-attacks tends to change a lot faster, as cyber-attacks often directly modify the system parameters and signals. It should be noted that this is a general observation and conclusion because we do not consider the stealthy attacks that simulate health system responses. Theoretically, based on the full knowledge of the control system, a malicious attacker can design a certain parameter changing rate to increase the difficulty of distinguishing between load transient and cyber-attacks. As discussed in the literature [65], the attacker can drive the system to an unsafe state while remaining stealthy by using the system knowledge. However, these stealthy attacks need much more prior knowledge and

2332-7782 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: University of Georgia. Downloaded on December 16,2020 at 22:18:35 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TTE.2020.3044524, IEEE Transactions on Transportation Electrification

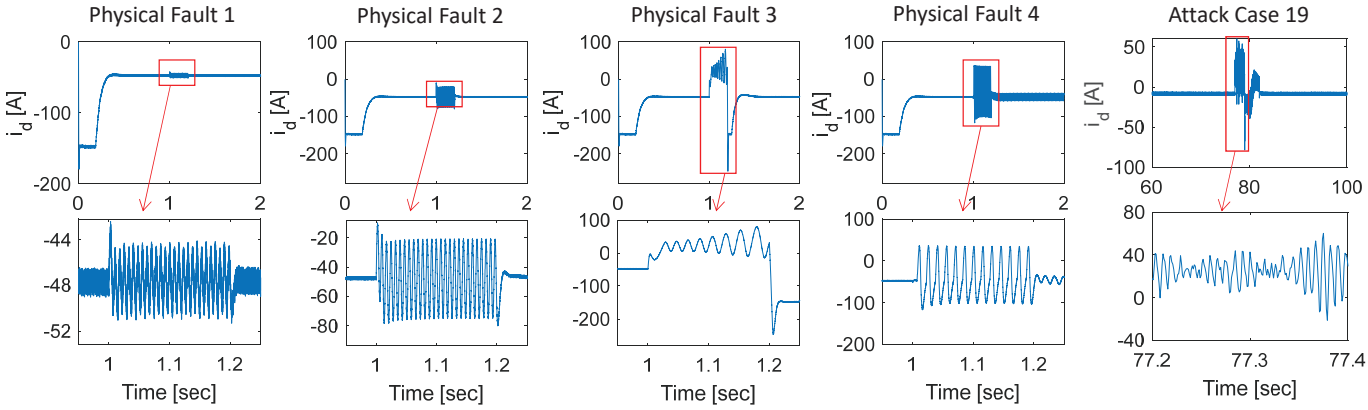


Fig. 13: Comparison between malicious cyber-attacks and physical faults.

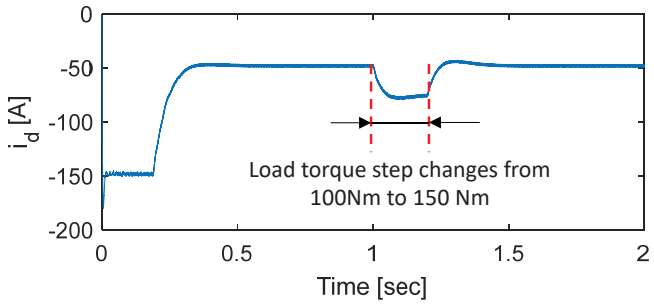


Fig. 14: A load change from 100Nm to 150Nm.

longer attack duration, leading to a higher cost of conducting a cyber-attack.

E. Discussion on the real-time implementation of the proposed method

Generally speaking, cyber-attack detection is the first step for cybersecurity. After identifying the cyber-threats, root-cause diagnosis and resilient control should be applied to recover the system from cyber-attacks at the early stage. This roadmap is based on dual-microcontroller system topology, in which both primary and secondary microcontrollers can receive sensor feedback signals and provide a control command to the system. In normal cases, the system is controlled by the primary controller, while the monitoring systems, including cyber-attack detection and diagnosis algorithms, are integrated into the secondary microcontroller. Once a cyber-attack is identified, and the compromised signal is diagnosed, a resilient control algorithm in the secondary microcontroller would be used. For example, in [66], a cyber-secure power router prototype was proposed for the cybersecurity of grid-connected power electronics. Using two digital signal processors (DSPs) in this topology, the switching between the primary and secondary microcontrollers is realized. Because the main focus of the paper is to provide a cyber-attack detection algorithm, we do not provide further detailed discussions on this cyber-secure design topology. It is worth noting that, although root-cause diagnosis and resilient control remain to be solved, at least in a short time, these works are less critical than threat detection. This is because, once a potential cyber-attack or a physical fault is detected during driving, the human driver can stop the car and request car maintenance for further detection and diagnosis.

V. Conclusion

In this paper, we have presented a physics-guided machine learning approach for cyber-attack detection to address the issue of cyber-physical security of EVs. Besides the raw data obtained from device-level and vehicle-level systems, innovative physics-guided features are also used to reflect the transient physical characteristics of the vehicle. Simulation results in an OPAL-RT HIL simulation testbed have validated the proposed detection method and demonstrated high accuracy under various driving scenarios. In real applications, the algorithm is firstly trained offline and then applied to real-time cyber-attack detection. Despite the acceptable detection accuracy in the training process, several challenges remain to be solved for practical applications. One of the challenges is how to ensure the real-time detection performance using the offline trained network, especially when considering the varying external driving environment, uncertain noises, and scenarios that are not considered in the design process. As stated above, frequent misbehavior in cyber-attack identification would make the security equipment worthless for human drivers. Therefore, in the future, we will focus on adaptive cyber-attack detection that can be adjusted with online driving data and more in-depth physical characteristics of EVs.

References

[1] J. C. Balda, A. Mantooth, R. Blum, and P. Tenti, "Cybersecurity and power electronics: Addressing the security vulnerabilities of the internet of things," *IEEE Power Electronics Magazine*, vol. 4, no. 4, pp. 37–43, Dec 2017.

[2] C. Hodge, K. Hauck, S. Gupta, and J. C. Bennett, "Vehicle cybersecurity threats and mitigation approaches," National Renewable Energy Lab.(NREL), Golden, CO (United States), Tech. Rep., 2019.

[3] G. K. Rajbahadur, A. J. Malton, A. Walenstein, and A. E. Hassan, "A survey of anomaly detection for connected vehicle cybersecurity and safety," in *2018 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2018, pp. 421–426.

[4] M. Levi, Y. Allouche, and A. Kontorovich, "Advanced analytics for connected car cybersecurity," in *2018 IEEE 87th Vehicular Technology Conference (VTC Spring)*. IEEE, 2018, pp. 1–7.

2332-7782 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: University of Georgia. Downloaded on December 16,2020 at 22:18:35 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TTE.2020.3044524, IEEE Transactions on Transportation Electrification

[5] X. Shao, C. Dong, and L. Dong, "Research on detection and evaluation technology of cybersecurity in intelligent and connected vehicle," in *2019 International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM)*. IEEE, 2019, pp. 413–416.

[6] Y. Xun, J. Liu, N. Kato, Y. Fang, and Y. Zhang, "Automobile driver fingerprinting: A new machine learning based authentication scheme," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 2, pp. 1417–1426, 2019.

[7] L. Guo, J. Ye, and L. Du, "Cyber-physical security of energy-efficient powertrain system in hybrid electric vehicles against sophisticated cyber-attacks," *IEEE Transactions on Transportation Electrification*, 2020.

[8] A. Greenburg, "Hackers remotely kill a Jeep on the highway - with me in it," Available: https://www.wired.com/2015/07/hackers-remotely-kill-jeep-highway/, Tech. Rep., Jul. 2015.

[9] C. Miller and C. Valasek, "Remote exploitation of an unaltered passenger vehicle," *Black Hat USA*, vol. 2015, p. 91, 2015.

[10] "Vehicle electrical system security committee: SAE J3061 cybersecurity guidebook for cyber-physical automotive systems," SAE, Tech. Rep., 2016.

[11] "International organization for standardization: ISO 26262 road vehicles functional safety part 1–10. technical report, international organization for standardization," ISO, Tech. Rep., 2011.

[12] C. Schmittner and G. Macher, "Automotive cybersecurity standards-relation and overview," in *International Conference on Computer Safety, Reliability, and Security*. Springer, Cham, 2019, pp. 153–165.

[13] J. Liu, S. Zhang, W. Sun, and Y. Shi, "In-vehicle network attacks and countermeasures: Challenges and future directions," *IEEE Network*, vol. 31, no. 5, pp. 50–58, 2017.

[14] S. Abbott-McCune and L. A. Shay, "Intrusion prevention system of automotive network can bus," in *2016 IEEE International Carnahan Conference on Security Technology (ICCST)*. IEEE, 2016, pp. 1–8.

[15] A. Khalid, A. Sundararajan, A. Hernandez, and A. I. Sarwat, "Facts approach to address cybersecurity issues in electric vehicle battery systems," in *2019 IEEE Technology & Engineering Management Conference (TEMSCON)*. IEEE, 2019, pp. 1–6.

[16] S. Chakraborty, M. A. Al Faruque, W. Chang, D. Goswami, M. Wolf, and Q. Zhu, "Automotive cyber–physical systems: A tutorial introduction," *IEEE Design & Test*, vol. 33, no. 4, pp. 92–108, 2016.

[17] R. Hull, "Nissan disables leaf electric car app after revelation that hackers can switch on the heater to drain the battery," *Thisismoney, Feb*, vol. 26, 2016.

[18] B. Yang, L. Guo, F. Li, J. Ye, and W.-Z. Song, "Vulnerability assessments of electric drive systems due to sensor data integrity attacks," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 5, pp. 3301–3310, 2020.

[19] B. Yang, L. Guo, F. Li, J. Ye, and W. Song, "Impact analysis of data integrity attacks on power electronics and electric drives," in *2019 IEEE Transportation Electrification Conference and Expo (ITEC)*. IEEE, 2019, pp. 1–6.

[20] A. Weimerskirch and R. Gaynier, "An overview of automotive cybersecurity: Challenges and solution approaches." in *TrustED@ CCS*, 2015, p. 53.

[21] J. Valenzuela, J. Wang, and N. Bissinger, "Real-time intrusion detection in power system operations," *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 1052–1062, 2012.

[22] A. Subasi, K. Al-Marwani, R. Alghamdi, A. Kwairanga, S. M. Qaisar, M. Al-Nory, and K. A. Rambo, "Intrusion detection in smart grid using data mining techniques," in *2018 21st Saudi Computer Society National Computer Conference (NCC)*. IEEE, 2018, pp. 1–6.

[23] B. Chen, S. Mashayekh, K. L. Butler-Purry, and D. Kundur, "Impact of cyber attacks on transient stability of smart grids with voltage support devices," in *2013 IEEE Power & Energy Society General Meeting*. IEEE, 2013, pp. 1–5.

[24] Y. Isozaki, S. Yoshizawa, Y. Fujimoto, H. Ishii, I. Ono, T. Onoda, and Y. Hayashi, "Detection of cyber attacks against voltage control in distribution power grids with PVs," *IEEE Transactions on Smart Grid*, vol. 7, no. 4, pp. 1824–1835, 2015.

[25] H. Zhang, W. Meng, J. Qi, X. Wang, and W. X. Zheng, "Distributed load sharing under false data injection attack in an inverter-based microgrid," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 2, pp. 1543–1551, 2018.

[26] Y. Cao, K. Davis, and S. Zonouz, "A framework of smart and secure power electronics driven HVAC thermal inertia in distributed power systems," in *2018 IEEE Green Technologies Conference (GreenTech)*. IEEE, 2018, pp. 127–132.

[27] J. Giraldo, D. Urbina, A. Cardenas, J. Valente, M. Faisal, N. O. T. J. Ruths, H. S. Sandberg, and R. Candell, "A survey of physics-based attack detection in cyber-physical systems," *ACM Computing Surveys (CSUR)*, vol. 51, no. 4, p. 76, 2018.

[28] 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.

[29] F. Miao, Q. Zhu, M. Pajic, and G. J. Pappas, "Coding sensor outputs for injection attacks detection," in *53rd IEEE Conference on Decision and Control*. IEEE, 2014, pp. 5776–5781.

[30] S. Sridhar and M. Govindarasu, "Model-based attack detection and mitigation for automatic generation control," *IEEE Transactions on Smart Grid*, vol. 5, no. 2, pp. 580–591, 2014.

[31] Y. Mo and B. Sinopoli, "Secure control against replay attacks," in *2009 47th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2009, pp. 911–918.

[32] S. Mishra, Y. Shoukry, N. Karamchandani, S. N. Diggavi, and P. Tabuada, "Secure state estimation against sensor attacks in the presence of noise," *IEEE Transactions on Control of Network Systems*, vol. 4, no. 1, pp. 49–59, 2017.

[33] D. Hadziosmanovic, R. Sommer, E. Zambon, and P. H. Hartel, "Through the eye of the PLC: semantic security monitoring for industrial processes," in *Proceedings of the Annual Computer Security Applications Conference (ACSAC'14)*, 2014, pp. 126–135.

[34] C. Schellenberger and P. Zhang, "Detection of covert attacks on cyber-physical systems by extending the system dynamics with an auxiliary system," in *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*. IEEE, 2017, pp. 1374–1379.

[35] D. D. Ferreira, J. M. de Seixas, A. S. Cerqueira, C. A. Duque, M. H. Bollen, and P. F. Ribeiro, "A new power quality deviation index based on principal curves," *Electric Power Systems Research*, vol. 125, pp. 8–14, 2015.

[36] H. Liu, F. Hussain, Y. Shen, S. Arif, A. Nazir, and M. Abubakar, "Complex power quality disturbances classification via curvelet transform and deep learning," *Electric Power Systems Research*, vol. 163, pp. 1–9, 2018.

[37] Y. Shi, F. Li, W. Song, X.-Y. Li, and J. Ye, "Energy audition based cyber-physical attack detection system in iot," in *Proceedings of the ACM Turing Celebration Conference-China*, 2019, pp. 1–5.

[38] O. P. Mahela, A. G. Shaik, and N. Gupta, "A critical review of detection and classification of power quality events," *Renewable and Sustainable Energy Reviews*, vol. 41, pp. 495–505, 2015.

[39] M. Esmalifalak, L. Liu, N. Nguyen, R. Zheng, and Z. Han, "Detecting stealthy false data injection using machine learning in smart grid," *IEEE Systems Journal*, vol. 11, no. 3, pp. 1644–1652, 2014.

[40] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM computing surveys (CSUR)*, vol. 41, no. 3, p. 15, 2009.

[41] M. Al-Saud, A. M. Eltamaly, M. A. Mohamed, and A. Kavousi-Fard, "An intelligent data-driven model to secure intravehicle communications based on machine learning," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 6, pp. 5112–5119, 2019.

[42] A. Karpatne, W. Watkins, J. Read, and V. Kumar, "Physics-guided neural networks (PGNN): An application in lake temperature modeling," *arXiv preprint arXiv:1710.11431*, 2017.

[43] X. Jia, J. Willard, A. Karpatne, J. S. Read, J. A. Zwart, M. Steinbach, and V. Kumar, "Physics-guided machine learning for scientific discovery: An application in simulating lake temperature profiles," *arXiv preprint arXiv:2001.11086*, 2020.

[44] A. Piccione, J. Berkery, S. Sabbagh, and Y. Andreopoulos, "Physics-guided machine learning approaches to predict the ideal stability properties of fusion plasmas," *Nuclear Fusion*, vol. 60, no. 4, p. 046033, 2020.

[45] L. Wang and Q. Zhou, "Physics-guided deep learning for time-series state estimation against false data injection attacks," in *2019 North American Power Symposium (NAPS)*. IEEE, 2019, pp. 1–6.

[46] X. Hu, Y. Li, C. Lv, and Y. Liu, "Optimal energy management and sizing of a dual motor-driven electric powertrain," *IEEE Transactions on Power Electronics*, vol. 34, no. 8, pp. 7489–7501, 2018.

[47] B. Sun, S. Gao, C. Ma, and J. Li, "System power loss optimization of electric vehicle driven by front and rear induction motors," *International Journal of Automotive Technology*, vol. 19, no. 1, pp. 121–134, 2018.

[48] T. K. S. Lab, "Experimental security research of tesla autopilot," Tech. Rep., 2019.

[49] I. Ilascu, "Hvacking: Remotely exploiting bugs in building control systems," Available: https://www.bleepingcomputer.com/news/security/hvacking-remotely-exploiting-bugs-in-building-control-systems/, Aug. 2019.

2332-7782 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: University of Georgia. Downloaded on December 16,2020 at 22:18:35 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TTE.2020.3044524, IEEE Transactions on Transportation Electrification

[50] M. S. Mahmoud, M. M. Hamdan, and U. A. Baroudi, "Modeling and control of cyber-physical systems subject to cyber attacks: a survey of recent advances and challenges," *Neurocomputing*, vol. 338, pp. 101–115, 2019.
[51] Y.-L. Huang, A. A. Cárdenas, S. Amin, Z.-S. Lin, H.-Y. Tsai, and S. Sastry, "Understanding the physical and economic consequences of attacks on control systems," *International Journal of Critical Infrastructure Protection*, vol. 2, no. 3, pp. 73–83, 2009.
[52] F. Li, R. Xie, B. Yang, L. Guo, P. Ma, J. Shi, J. Ye, and W. Song, "Detection and identification of cyber and physical attacks on distribution power grids with PVs: An online high-dimensional data-driven approach," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, 2019.
[53] L. Guzzella, A. Sciarretta *et al.*, *Vehicle propulsion systems*. Springer, Berlin, Heidelberg, 2007, vol. 1.
[54] X. Liao, Q. Huang, D. Sun, W. Liu, and W. Han, "Real-time road slope estimation based on adaptive extended kalman filter algorithm with in-vehicle data," in *2017 29th Chinese Control And Decision Conference (CCDC)*. IEEE, 2017, pp. 6889–6894.
[55] M. N. Mahyuddin, J. Na, G. Herrmann, X. Ren, and P. Barber, "Adaptive observer-based parameter estimation with application to road gradient and vehicle mass estimation," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 6, pp. 2851–2863, 2013.
[56] S. Hao, P. Luo, and J. Xi, "Estimation of vehicle mass and road slope based on steady-state kalman filter," in *2017 IEEE International Conference on Unmanned Systems (ICUS)*. IEEE, 2017, pp. 582–587.
[57] M. Klomp, Y. Gao, and F. Bruzelius, "Longitudinal velocity and road slope estimation in hybrid electric vehicles employing early detection of excessive wheel slip," *Vehicle System Dynamics*, vol. 52, no. sup1, pp. 172–188, 2014.
[58] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
[59] N. Wagner and J. M. Rondinelli, "Theory-guided machine learning in materials science," *Frontiers in Materials*, vol. 3, p. 28, 2016.
[60] F. Karim, S. Majumdar, H. Darabi, and S. Chen, "Lstm fully convolutional networks for time series classification," *IEEE access*, vol. 6, pp. 1662–1669, 2017.
[61] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information processing & management*, vol. 45, no. 4, pp. 427–437, 2009.
[62] M. Straczkiewicz, J. Urbanek, W. Fadel, C. M. Crainiceanu, and J. Harezlak, "Automatic car driving detection using raw accelerometry data," *Physiological Measurement*, vol. 37, no. 10, p. 1757, 2016.
[63] M. Tanelli, S. M. Savaresi, and C. Cantoni, "Longitudinal vehicle speed estimation for traction and braking control systems," in *2006 IEEE Conference on Computer Aided Control System Design, 2006 IEEE International Conference on Control Applications, 2006 IEEE International Symposium on Intelligent Control*. IEEE, 2006, pp. 2790–2795.
[64] M. S. Basrah, E. Siampis, E. Velenis, D. Cao, and S. Longo, "Wheel slip control with torque blending using linear and nonlinear model predictive control," *Vehicle System Dynamics*, vol. 55, no. 11, pp. 1665–1685, 2017.
[65] A. Teixeira, I. Shames, H. Sandberg, and K. H. Johansson, "A secure control framework for resource-limited adversaries," *Automatica*, vol. 51, pp. 135–148, 2015.
[66] S. Moquin, S. Kim, N. Blair, C. Farnell, J. Di, and H. A. Mantooth, "Enhanced uptime and firmware cybersecurity for grid-connected power electronics," in *2019 IEEE CyberPELS (CyberPELS)*. IEEE, 2019, pp. 1–6.

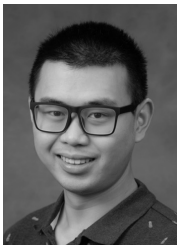
Lulu Guo received the B.S. degree in vehicle engineering and the Ph.D. degree in control engineering from Jilin University, Changchun, China, in 2014 and 2019, respectively.

He is currently a Postdoctoral Research Associate with the University of Georgia, Athens, GA, USA. His current research interests include advanced vehicle control, energy management, and vehicle cybersecurity.

Jin Ye (S'13-M'14-SM'16) received the B.S. and M.S. degrees in electrical engineering from Xi'an Jiaotong University, Xi'an, China, in 2008 and 2011, respectively, and the Ph.D. degree in electrical engineering from McMaster University, Hamilton, ON, Canada, in 2014.

She is currently an Assistant Professor of electrical engineering and the Director of the Intelligent Power Electronics and Electric Machines Laboratory, University of Georgia, Athens, GA, USA. Her current research interests include power electronics, electric machines, energy management systems, smart grids, electrified transportation, and cyber-physical systems.

Dr. Jin Ye is the General Chair of 2019 IEEE Transportation Electrification Conference and Expo (ITEC), and the Publication Chair and Women in Engineering Chair of 2019 IEEE Energy Conversion Congress and Expo (ECCE). She is an Associate Editor for IEEE TRANSACTIONS ON POWER ELECTRONICS, IEEE OPEN JOURNAL OF POWER ELECTRONICS, IEEE TRANSACTIONS ON TRANSPORTATION ELECTRIFICATION, and IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY.



Bowen Yang received the B.S. degree in electrical engineering from Huazhong University of Science and Technology, Wuhan, China, in 2018. He is currently pursuing the Ph.D. degree with the University of Georgia, Athens, GA, USA.

He is also a Research Assistant with the University of Georgia, USA. His current research interests include advanced control for power electronics and electric machines, energy management system, and cyber-physical security for intelligent electric drives.

2332-7782 (c) 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: University of Georgia. Downloaded on December 16,2020 at 22:18:35 UTC from IEEE Xplore. Restrictions apply.
View publication stats