Modeling and Detecting False Data Injection Attacks against Railway Traction Power Systems

Subhash Lakshminarayana, Advanced Digital Sciences Center, Illinois at Singapore
Teo Zhan Teng, GovTech, Singapore
Rui Tan, Nanyang Technological University, Singapore
David K.Y. Yau, Singapore University of Technology and Design

Modern urban railways extensively use computerized sensing and control technologies to achieve safe, reliable, and well-timed operations. However, the use of these technologies may provide a convenient leverage to cyber-attackers who have bypassed the air gaps and aim at causing safety incidents and service disruptions. In this paper, we study false data injection (FDI) attacks against railways' traction power systems (TPSes). Specifically, we analyze two types of FDI attacks on the train-borne voltage, current, and position sensor measurements – which we call *efficiency attack* and *safety attack* – that (i) maximize the system's total power consumption and (ii) mislead trains' local voltages to exceed given safety-critical thresholds, respectively. To counteract, we develop a global attack detection (GAD) system that serializes a *bad data detector* and a novel *secondary attack detector* designed based on unique TPS characteristics. With intact position data of trains, our detection system can effectively detect the FDI attacks on trains' voltage and current measurements even if the attacker has full and accurate knowledge of the TPS, attack detection, and real-time system state. In particular, the GAD system features an adaptive mechanism that ensures low false positive and negative rates in detecting the attacks under noisy system measurements. Extensive simulations driven by realistic running profiles of trains verify that a TPS setup is vulnerable to the FDI attacks, but these attacks can be detected effectively by the proposed GAD while ensuring a low false positive rate.

1. INTRODUCTION

In modern cities, safe, reliable, and well-timed operations of urban railways are critical. A modern railway is a highly complex cyber-physical system (CPS) consisting of diverse subsystems including train motion control, traction powering, signaling, etc, where deeply embedded information and communication technologies (ICTs) are used to operate each train and connect trains to an operation center. The extensive use of ICT may provide a convenient leverage to attackers, however, who may aim to hurt passengers' safety or cause widespread service disruptions. To date, the cybersecurity of modern railways has relied on air gaps that isolate their ICT systems from public networks. However, recent high-profile intrusions such as Stuxnet [Karnouskos 2011] and Dragonfly [Symantec 2014] have successfully breached the air gaps of critical CPS infrastructures and resulted in physical damage. For instance, the Stuxnet worm damaged nuclear centrifuges by injecting false control commands and forging normal system states. Its design and architecture are not domain-specific – they can be readily customized against other types of CPS including transportation [Karnouskos 2011]. Insider attacks represent another major threat to air-gapped systems; their severe consequences have likewise been well documented [Security 2011]. It is thus critical to

This work was supported in part by the National Research Foundation (NRF), Prime Minister's Office, Singapore, under its National Cybersecurity R&D Programme (Award No. NRF2014NCR-NCR001-31) and administered by the National Cybersecurity R&D Directorate and in part by a Start-up Grant at NTU.
Author's addresses: S. Lakshminarayana, Advanced Digital Sciences Center, Illinois at Singapore (e-mail: subhash.l@adsc.com.sg); Z.T. Teo, GovTech Singapore (e-mail: teozt@hotmail.com); R. Tan, School of Computer Science and Engineering, Nanyang Technological University, Singapore (e-mail: tanrui@ntu.edu.sg); D.K.Y. Yau, Singapore University of Technology and Design (e-mail: david_yau@sutd.edu.sg). The work was conducted when Z.T. Teo was with the Advanced Digital Sciences Center, Illinois at Singapore.
 2378-962X/2017/01-ART1 $15.00
DOI: 0000001.0000001

understand the cybersecurity risks of modern railways as a mission-critical CPS, and develop effective security countermeasures in their ICT design.

In this paper, we study the cybersecurity of direct current (dc) traction power systems (TPSes) that are widely deployed in urban electrical railways. The criticality of TPS is evidenced by prior severe incidents caused by TPS malfunctions. The 2014 Moscow metro derailment that led to 24 dead and 160 injured was caused by sudden braking of the train in question, when its traction voltage dropped abruptly [Gabrielle 2014]. In Singapore, a system-wide metro service disruption, triggered by TPS faults, affected almost half a million commuters during rush hours on July 7, 2015 [SMRT 2015]. Moreover, the computerized sensing and control in an automated TPS could be prime targets for cyber-attackers, who can sabotage the control and steer the system into inefficient and unsafe states.

Motivated by Stuxnet worm-type attacks that forge physical system states, in this paper we study a general class of integrity attacks called *false data injection* (FDI), which tampers with train-borne sensor measurements required by TPS control. In a TPS, the electricity power supplied by substations is delivered by a network of overhead lines and third rails to the trains. According to its operation mode, a train's power consumption can be highly dynamic. In traction mode, it draws power from the TPS, causing a drop in the train's local voltage; in braking mode, it regenerates electricity from kinetic energy and injects this electricity back to the TPS,[1] causing a rise in the voltage. To prevent the voltage from exceeding safety-critical thresholds, trains apply *overcurrent control* and *squeeze control* [Okada et al. 2004] to throttle their power draw and injection, respectively. As these controls are performed based on train-borne voltage and current sensor measurements, FDI attacks on the measurements may mislead the train into erroneous power control decisions, which may in turn produce damaging and even catastrophic physical impacts on the train and the TPS. Recent results show that the measurements can be compromised in practice by precisely controlled electromagnetic interference in analog sensors [Kune et al. 2013], hardware trojans in chips [Karri et al. 2010], and malware infections in sensor firmwares [McDaniel and McLaughlin 2009; Depuru et al. 2011; Mike 2009]. Hence, FDI attack is a clear and present threat that requires immediate attention.

In this paper, we aim to answer the following two fundamental research questions:

(1) How to characterize the impact of FDI attacks on TPS system efficiency and safety? Analysis of the impact based on an essential TPS model will provide basic understanding for developing countermeasures. However, the analysis is difficult, due to complex system dynamics arising from the trains' motion. In particular, a moving train does not only act as "load" and "generation" alternately over time, but it also alters the power network's topology and electrical parameters continually. Moreover, because different TPS components (trains, substations, etc.) become physically interconnected through a common underlying power network, effects of an erroneous power control on a train during attack may propagate to the neighboring TPS components. The analysis must address these intricate and unique characteristics of TPSes.

(2) How to develop effective approaches for detecting the FDI attacks? Our thesis is that, because measurements from different trains are inherently correlated through interconnection over the same power network, for attack resilience we can apply a global detection that cross-checks the measurements collected from all trains based on an *a priori* global TPS model. However, in contrast to alternating current (ac) power grids that have well-established centralized monitoring and sensor data cross-check safeguards for reliable holistic control [Rahman et al. 2014; Liu et al. 2009], TPS

[1] In electrical railways, trains are often equipped with regenerative brakes that generate electricity in deceleration [Fletcher 1991].

is mainly concerned with individual trains' local operation (i.e., the overcurrent and squeeze controls), and therefore it is not traditionally subject to any global sensor data checks across trains. Thus, an existing dc TPS operation center seldom scrutinizes the sensor measurements, beyond their display and presentation for human operators. In this paper, we demonstrate the importance of these global, but hitherto ignored, sensor data cross checks in the TPS domain against FDI attacks.

In answering the above two research questions, our main contributions in this paper are as follows:

First, based on essential models of power substations, power flows, and train overcurrent and squeeze controls in a TPS, we formulate two types of FDI attacks that we call *efficiency attack* and *safety attack*. These attacks (i) maximize the total instantaneous power consumption of the TPS and (ii) mislead victim trains' local voltages to exceed given safety-critical thresholds, respectively. Efficiency attacks will increase the train's traction power consumption, resulting in an increase in railways' operation expenses.[2] Efficiency attacks will also potentially increase the carbon footprint of the transportation sector, which is an important consideration for railway operators [Transport for London 2008]. On the other hand, safety attacks may trip circuit breakers, causing dangerous power loss and brake malfunction.

The efficiency attack formulation models an aggressive attacker who aims at maximizing the attack impact and provides insights into understanding the performance degradation limit caused by FDI attacks. Numerical results for a TPS section with two substations and two trains show that the efficiency attack can result in an instantaneous efficiency loss of about 20%, whereas the safety attack on a single train can indeed lead to significant safety breaches. These results substantiate the potency of FDI attacks on train-borne sensor measurements.

Second, we propose to apply a global *bad data detection* (BDD) method, similar to that widely used in ac power grids [Liu et al. 2009], to detect FDI attacks in a dc TPS. Despite a known vulnerability of the BDD – it can be bypassed by an attacker who knows enough details of its design – our numerical results show that, in order for an FDI attack to be stealthy against the BDD, it will have to settle for a significantly reduced damage on the system efficiency. Moreover, we observe that, given intact position data of trains, solutions of the BDD bypass condition will become discrete. Based on this observation, we develop a novel *secondary attack detection* (SAD) algorithm that can effectively detect the onset of an FDI attack on trains' voltage and current measurements after it has bypassed the BDD. Hence, the BDD and the SAD form in tandem a global attack detector (GAD) under the Kerckhoffs's assumption (i.e., the attacker has full and accurate knowledge of the system model, attack detection, and real-time system state), provided that the integrity of trains' position information can be verified. Building on this result, we design an approach to mitigating the impact of an attack after its detection.

Third, we report extensive simulations, driven by realistic profiles of trains in operation, to evaluate our solutions. For a TPS consisting of four trains each running over a distance of ten kilometers for 800 seconds, our results show that, without the global BDD, FDI attacks can increase the total system energy consumption by 28.3% and breach the system's safety condition. After applying the BDD, the system's total energy consumption increases by no more than 6.2% under the efficiency attack, and safety attacks become no longer successful. Moreover, the proposed SAD algorithm

[2]Energy costs of running urban rail pose a significant financial burden to transport companies, constituting about 20% of their operational expenses [Osi 2015]. Of this, about 80% of energy is consumed for traction (e.g. train's motion, braking, electric losses) [Gonzlez-Gil et al. 2014].

achieves a detection probability of 96% in detecting the onsets of the FDI attacks that have successfully bypassed the BDD.

Finally, we investigate the false positives (FPs) and missed detections (MDs) of the proposed detectors in the presence of sensor measurement noises. Simulation results illustrate that although the GAD yields low FP and MD rates during most of the simulation time, it gives a relatively high FP rates for a few short time durations when one or more trains change their status of motion (e.g., from tractioning mode to braking mode). To maintain a low FP rate all the time, we propose an adaptation mechanism based on an attack detection window for the GAD. We call the improved attack detection system GAD-W. Simulation results show that with appropriately chosen detector parameters, the GAD-W detector achieves an average FP rate of 9×10^{-4} and an MD rate of 7×10^{-4} over the entire simulation time.

This work focuses on attacks against urban metros (e.g., Tokyo, Singapore, and Berlin) that adopt dc systems. Thus, our analysis is based on the dc TPS model. On the other hand, long-distance railways usually adopt an alternating current (ac) TPS, due to higher efficiency in transmitting ac power over long distances [Rai 2016; Yadav 2013]. Although a detailed investigation of cybersecurity issues in ac traction power systems is beyond the scope of our paper, we conjecture that the vulnerabilities of the two systems are similar. This is because ac and dc TPSes mainly differ in their design of electrical components (e.g., substation and train motor) [Yadav 2013], while the ICT infrastructures in these two kinds of systems are similar. Thus, the attack surfaces of the cyber infrastructures in both cases are the same. Nevertheless, the attack impact analysis and the detector design may differ in details, which are left for future work.

The balance of the paper is organized as follows. Section 2 reviews related work. Section 3 describes our TPS model. Section 4 formulates the efficiency and safety attacks. Section 5 analyzes the effectiveness of the BDD and presents the proposed SAD algorithm that complements the BDD. Section 6 analyzes the impact of sensor measurement noises. Section 7 presents simulation results. Section 8 concludes.

2. RELATED WORK

Power flow analysis and optimization for TPS have received increasing research interest. Power flow analysis is a basic tool for TPS planning and operation. Prior work has analyzed dc power flows [Cai et al. 1995; Arboleya et al. 2016; Pires et al. 2007] and addressed the interactions between the dc TPS and a supporting ac power grid [Abrahamsson 2012], [Arboleya et al. 2012]. We adopt existing electrical models for different TPS components [Cai et al. 1995], [Arboleya et al. 2016], [Pires et al. 2007] in this work. These models provide sufficient accuracy generally [Arboleya et al. 2016], and they are tractable for analysis. Based on power flow analysis, recent research has tried to improve the energy efficiency of railways by leveraging trains' power regeneration [David 2015]. Techniques such as synchronizing the trains' speed profiles [Miyatake and Ko 2010; Shuai et al. 2014; Shuai et al. 2015] and real-time substation voltage control [Raghunathan et al. 2014] have been shown to provide efficient reuse of the regenerated power. To the best of our knowledge, none of the existing studies on TPS control have addressed it from a cybersecurity perspective. The security problem is imperative, since TPS is a form of critical infrastructure that renders it an attractive target for attacks.

Different types of CPS can have vastly different properties and characteristics, and their security concerns and admissible detection and mitigation strategies can be totally different. Typically, their cybersecurity analysis must be carried out in a domain specific manner, with customized considerations given to main details and semantics of specific systems. Cárdenas et al. [Cárdenas et al. 2011] investigate the impacts of integrity and denial-of-service attacks on the process control system, which has mul-

tiple sensors and control loops, of a chemical reactor. Amin et al. [Amin et al. 2013] perform security threat assessment of supervisory control and data acquisition systems for water supply. Other efforts [Liu et al. 2009; Jinsub and Lang 2013] have analyzed FDI attacks against ac utility power grids. They show that an attacker capable of tampering with grid sensor measurements or topology information can carefully construct attacks to bypass detection by certain existing fault data detectors. Recent studies have investigated the impact of such stealthy attacks on grid power flows [Yanling et al. 2011; Teixeira et al. 2012; Rahman et al. 2014]. They show that maliciously biased estimates of the system state can cause grid operators to make erroneous decisions that will lead to degraded performance or safety breaches. This paper is the first to analyze the efficiency and safety of TPS under FDI attacks. We provide new and non-trivial domain-specific modeling and analysis to capture the targeted application's unique features and key properties. In particular, TPS involves real-time and complex interactions between two highly dynamical physical systems, namely a mechanical system of the trains' motion and an electrical system that governs the trains' power consumption and regeneration during this motion. Attackers could exploit the interactions to strengthen their attacks.

3. TRACTION POWER SYSTEM MODEL

In this section, we present a model of a dc TPS at a certain time instant. The TPS is modeled as a power network consisting of N nodes. Denote by $\mathcal{N} = \{1, 2, \ldots, N\}$ the set of nodes and $\mathcal{L}$ the set of resistive branches connecting the nodes. The substations and the trains are connected to different nodes. The sets of nodes for the substations, the tractioning trains, and the regenerating trains are denoted by $\mathcal{N}_{\text{sub}}$, $\mathcal{N}_{\text{tra}}$, and $\mathcal{N}_{\text{reg}}$, respectively. We define $\mathcal{N}_{\text{trains}} = \mathcal{N}_{\text{tra}} \cup \mathcal{N}_{\text{reg}}$. The positions of the nodes $1, \ldots, N$ are denoted by a set $\mathbf{s} = \{s_1, s_2, \ldots, s_N\}$, where s_1 is fixed at zero and s_i is the distance from node i to node 1. Fig. 1 illustrates a TPS section with two substations at nodes 1 and 4, as well as two trains at nodes 2 and 3. In this example, the train at node 2 is tractioning and the train at node 3 is braking and regenerating. Therefore, $\mathcal{N} = \{1, 2, 3, 4\}, \mathcal{L} = \{(1, 2), (2, 3), (3, 4)\}, \mathcal{N}_{\text{sub}} = \{1, 4\}, \mathcal{N}_{\text{tra}} = \{2\}, \mathcal{N}_{\text{reg}} = \{3\}$. The electrical models for the power network, substations, and trains are described as follows.

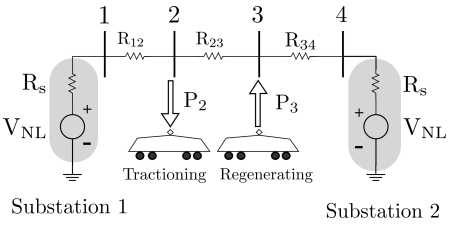


Fig. 1: Illustration of a TPS section.

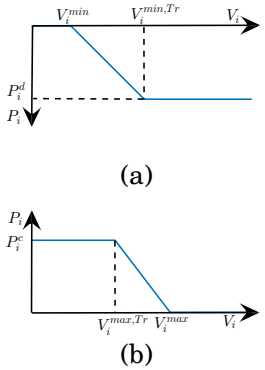


Fig. 2: (a) Overcurrent control. (b) Squeeze control.

Power network: Let V_i and I_i denote the voltage and current at node i, respectively, and $\mathbf{v}$ and $\mathbf{i}$ the vectors of the nodal voltages and currents. For safe operations, all

nodal voltages must be within a safety limit, i.e.,

$$V_i^{\min} \leq V_i \leq V_i^{\max}, \qquad \forall i \in \mathcal{N}, \tag{1}$$

where $V_i^{\min}$ and $V_i^{\max}$ are the safety thresholds for node i. By convention, we assume that the current injected into the TPS is positive. The resistance of the branch connecting the nodes i and j is denoted by $R_{i,j}(\mathbf{s})$ and its conductance by $G_{i,j}(\mathbf{s})$, where $G_{i,j}(\mathbf{s}) = 1/R_{i,j}(\mathbf{s})$. Note that branch resistance (and conductance) depends on the positions of the trains, i.e., $\mathbf{s}$. For instance, in Fig. 1, $R_{i,i+1} = \gamma(s_{i+1} - s_i)$, where γ is a constant depending on the electrical wire characteristics. From Kirchhoff's circuit laws, we have

$$\mathbf{Y}(\mathbf{s})\mathbf{v} = \mathbf{i}, \tag{2}$$

where $\mathbf{Y}(\mathbf{s}) \in \mathbb{R}^{N \times N}$ is the nodal conductance matrix and the (i,j)th element of $\mathbf{Y}(\mathbf{s})$, denoted by $Y_{i,j}(\mathbf{s})$, is given by

$$Y_{i,i}(\mathbf{s}) = \sum_{j:(i,j)\in\mathcal{L}} G_{i,j}(\mathbf{s}), \qquad Y_{i,j}(\mathbf{s}) = \begin{cases} -G_{i,j}(\mathbf{s}), & \text{if } j \neq i \text{ and } (i,j) \in \mathcal{L}, \\ 0, & \text{if } j \neq i \text{ and } (i,j) \notin \mathcal{L}. \end{cases}$$

Substations: We consider inverting substations capable of both supplying and absorbing power. They are modeled as DC voltage sources governed by

$$V_i = V_{\text{NL}} - R_s I_i, \qquad i \in \mathcal{N}_{\text{sub}}, \tag{3}$$

where V_{NL} and R_s are the no-load voltage and the internal resistance of the substation. When a substation supplies power, $I_i > 0$; when it absorbs power, $I_i < 0$ and the absorbed power is injected back to the supporting ac power grid. This dc substation model has been widely adopted in TPS analysis [Arboleya et al. 2016], [Raghunathan et al. 2014].

Trains: Let P_i denote the power absorbed or injected by a tractioning train or a regenerating train at node i. We have

$$P_i = V_i I_i. \tag{4}$$

For safety, the trains adopt the following two local power controls [Okada et al. 2004].
Overcurrent control: A tractioning train absorbs power from the power network, resulting in a drop in the train's nodal voltage. To prevent the nodal voltage from falling below the safety threshold $V_i^{\min}$, the overcurrent control is activated whenever the nodal voltage at the train node i drops below a triggering voltage $V_i^{\min,\text{Tr}}$. Let P_i^d denote the power demand of a tractioning train at node i. The overcurrent control will command the train to absorb P_i units of power, where P_i is given by

$$P_i = \begin{cases} 0, & \text{if } V_i \leq V_i^{\min}; \\ P_i^d \left(\frac{V_i - V_i^{\min}}{V_i^{\min,\text{Tr}} - V_i^{\min}} \right), & \text{if } V_i^{\min} \leq V_i \leq V_i^{\min,\text{Tr}}; \\ P_i^d, & \text{if } V_i \geq V_i^{\min,\text{Tr}}. \end{cases} \tag{5}$$

This control law is illustrated in Fig. 2 (a). Specifically, if the nodal voltage at the train is greater than the triggering voltage, the train draws a power equal to its demand. Otherwise, the train curtails its power consumption according to the V_i. If the voltage drops below $V_i^{\min}$, the train does not draw power to prevent safety incidents.
Squeeze control: A regenerating train injects power into the power network, resulting in an increase in the train's nodal voltage. To prevent the voltage from exceeding the safety limit $V_i^{\max}$, the squeeze control is activated whenever the nodal voltage at the

train node i exceeds a certain triggering voltage level $V_{i,\max}^{\mathrm{Tr}}$. Let P_i^c denote the regeneration capacity of the train. The squeeze control will command the train to inject P_i units of power into the TPS, where P_i is given by

$$P_i = \begin{cases} P_i^c, & \text{if } V_i \leq V_i^{\max,\mathrm{Tr}}; \\ P_i^c \left(\frac{V_i^{\max} - V_i}{V_i^{\max} - V_i^{\max,\mathrm{Tr}}} \right), & \text{if } V_i^{\max,\mathrm{Tr}} \leq V_i \leq V_i^{\max}; \\ 0, & \text{if } V_i \geq V_i^{\max}. \end{cases} \tag{6}$$

This control law is illustrated in Fig. 2 (b). Specifically, if the nodal voltage at the train is lower than the triggering voltage, the train injects all the regenerated power. Otherwise, the train curtails the power injection according to the V_i by burning the remaining power in a rheostatic braking system [Okada et al. 2004]. If the voltage drops below $V_i^{\max}$, the train does not inject power into the TPS to prevent safety incidents.

The train's power demand P_i^d and regeneration capacity P_i^c depend on the train's running profile and real-time state. They can be provided by the train's motion control system. We note that the electrical models described in this section address the steady-state voltages and currents. They ignore the power transients of the trains due to their internal feedback control systems that implement the overcurrent/squeeze control decisions. However, it is safe to ignore these transients because they can settle quickly, before the next overcurrent/squeeze control action [Talukdar and Koo 1977]. This steady-state analysis approach has been widely adopted in TPS power flow analysis [Cai et al. 1995; Pires et al. 2007; Arboleya et al. 2016].

4. FALSE DATA INJECTION ATTACKS AGAINST TPS

In this section, we study how an attacker can mislead the TPS into an inefficient or unsafe operating state. We focus on FDI attacks that tamper with the measurements of train-borne voltage and current sensors. Such an attack will cause the TPS to make wrong decisions of power absorption/injection, since a train's overcurrent and squeeze controls depend on the sensor measurements. We further consider attacks of two different objectives: (i) increase the system's total instantaneous power consumption, and (ii) cause breaches of the safety conditions in (1). We call these two types of attacks *efficiency attack* and *safety attack*, respectively. In this section, we first describe our threat model. Then, we analyze the attacker's approach of computing effective efficiency and safety attacks. Lastly, we present numerical results to illustrate the two kinds of attacks.

4.1. Threat Model

Real-world attackers against critical CPSes are often smart, resourceful, and highly strategic. Their strategies can be guided by detailed knowledge of their targets, which can be obtained in practice by malicious insiders, long-term data exfiltration [Symantec 2014], or social engineering against employees, contractors, or vendors of the operators in question [Karnouskos 2011]. In this paper, we follow Kerckhoffs's principle to consider an attacker who has accurate knowledge of the targeted system and read access to the system state. Knowledge of the system includes the electrical models and parameters given in Section 3, as well as the system's method of attack detection. The system state includes present power demands, regeneration capacities, and voltage, current, and position measurements of all the trains. This information can be leaked through a compromised operation center, as in recent high-profile attacks [Karnouskos 2011; Symantec 2014]. We assume that the attacker has write access to voltage, current, and position measurements of nodes in the set $\mathcal{N}_{\mathrm{a}}$, where $\mathcal{N}_{\mathrm{a}} \subseteq \mathcal{N}$, so that he can corrupt these measurements. Recent studies have demonstrated that such unautho-

rized write access can be obtained for analog sensors, traditional electro-mechanical meters, and modern smart meters [McDaniel and McLaughlin 2009; Depuru et al. 2011; Kune et al. 2013]. Analog sensors are vulnerable to precisely controlled electromagnetic interference [Kune et al. 2013]; measurement devices can be affected by *hardware trojans* [Karri et al. 2010] and infected with malwares [McDaniel and McLaughlin 2009; Mike 2009].

Under the said Kerckhoffs's assumption on the attacker's knowledge, we will analyze his strategies of achieving successful efficiency and safety attacks. Conversely, we will also develop countermeasures by a defender to detect these attacks and mitigate their impacts. Our threat model is strong, but the conservative analysis is necessary because any underestimation of the attacker's capability may have catastrophic consequences, including extremely costly infrastructure damage and loss of human lives.

We note that, alternatively, the attacker can launch FDI attacks against the *decisions* of the local controls (i.e., the P_i values for the trains). To detect such attacks, each train can compare the P_i value in question with that computed based on the train's voltage and current measurements and the *a priori* overcurrent and squeeze control laws. In the rest of this paper, we focus on the analysis and detection of FDI attacks on the voltage and current measurements only. This problem is comparatively much more challenging since information compromised right at the sources will preclude its use for any subsequent sanity checks.

Finally, we note that other potential attacks such as the denial-of-service (DoS) attacks that block sensor reading reporting can be easily detected, since in TPSes, sensors periodically report readings. Upon detection, the operator can initiate mitigation steps (e.g., stop the trains) to prevent any safety incidents. Thus, in this paper, we focus on the more challenging FDI attacks, as its detection generally needs a deep understanding on the power flows and train/substation operations.

4.2. FDI Attack Construction

In this section, we analyze how to compute an effective *attack vector*, as a vector of false voltage and current measurements to be injected into the sensing systems of the trains in $\mathcal{N}_a$. Note that, in this section we ignore position measurements in the attack vector, because they will not affect the trains' overcurrent and squeeze controls. In the rest of the paper, we will use x' to denote the compromised version of a sensor measurement x. In the following analysis, we first derive conditions for the attack vector to mislead the train into absorbing or injecting a certain amount of power. With the calculated power absorptions/injections of the trains, we can determine the system's total power consumption and hence its safety status. Thus, we can formulate the attacker's problem of finding an attack vector to achieve his goal of maximizing the total power consumption, under conditions that we will state presently for enforcing certain amounts of power absorption/injection.

The following conditions are sufficient to enforce that a train at node $i \in \mathcal{N}_a$ will absorb or inject P_i units of power:

$$V_i' \begin{cases} \geq V_i^{\min,\mathrm{Tr}}, & \text{if } P_i = P_i^d, \\ = V_i^{\min} + \frac{P_i(V_i^{\min,\mathrm{Tr}} - V_i^{\min})}{P_i^{\min}}, & \text{if } P_i^d \leq P_i \leq 0, \ \forall i \in \mathcal{N}_\mathrm{a} \cap \mathcal{N}_\mathrm{tra}; \\ \leq V_i^{\min}, & \text{if } P_i = 0, \end{cases} \tag{7}$$

$$V_i' \begin{cases} \geq V_i^{\max}, & \text{if } P_i = 0, \\ = V_i^{\max} - \frac{P_i(V_i^{\max} - V_i^{\max,\mathrm{Tr}})}{P_i^c}, & \text{if } 0 \leq P_i \leq P_i^c, \ \forall i \in \mathcal{N}_\mathrm{a} \cap \mathcal{N}_\mathrm{reg}; \\ \leq V_i^{\max,\mathrm{Tr}}, & \text{if } P_i = P_i^c, \end{cases} \tag{8}$$

$$V_i' I_i' = P_i, \qquad i \in \mathcal{N}_{\mathrm{a}}; \tag{9}$$

$$P_i \geq P_i^d, \qquad i \in \mathcal{N}_{\mathrm{tra}}; \tag{10}$$

$$P_i \leq P_i^c, \qquad i \in \mathcal{N}_{\mathrm{reg}}. \tag{11}$$

The conditions in (7) and (8) are obtained by inverting the overcurrent and squeeze control laws given in Section 3, and replacing the true voltage V_i by the compromised measurement V_i'. As a result, based on V_i', the train will follow the overcurrent/squeeze control law to regulate its power absorption/injection to the attacker's desired value P_i. This control process is often achieved in a closed loop, with the measurements V_i' and I_i' acting as feedback and the desired value P_i as setpoint. Under the condition (9), the actual power absorption/injection under the aforementioned closed-loop control will converge to P_i. Moreover, the condition (9) can hide the attack for trains that can directly measure the power consumption. The conditions (10) and (11) ensure the feasibility of inducing the train to absorb/inject P_i units of power. Specifically, the attacker's desired P_i should not exceed a regenerating train's regeneration capacity. The condition (10), where both P_i and P_i^d are negative, prevents the mechanism from violating the overcurrent control. In summary, if the compromised measurements V_i' and I_i' satisfy the conditions in (7) to (11), the train will control its power absorption/injection to P_i. With this understanding, the attacker can carefully plan the attack vector to achieve his goal. Without the conditions in (7) to (11), the attacker cannot predict the impact of his attack and therefore cannot implement his desired strategy.

Each sensor in the TPS may apply data quality checks on its measurements. For instance, the measurements at the present time instant should not differ significantly from those predicted based on the measurements at the previous time instant. Intuitively, if the compromised measurement is bounded around the true measurement, the data quality checks, designed to be insensitive to natural random noises of measurement, will not raise an alarm. Thus, we assume that the compromised measurements need to satisfy:

$$\mathbf{v} - \Delta\mathbf{v} \preceq \mathbf{v}' \preceq \mathbf{v} + \Delta\mathbf{v}, \tag{12}$$

$$\mathbf{i} - \Delta\mathbf{i} \preceq \mathbf{i}' \preceq \mathbf{i} + \Delta\mathbf{i}, \tag{13}$$

where $\Delta\mathbf{v} = [\Delta V_1, \ldots, \Delta V_N]^T$ and $\Delta\mathbf{i} = [\Delta I_1, \ldots, \Delta I_N]^T$ are the maximum errors allowed by the data quality checks (in Section 7, we illustrate how to set the values of $\Delta\mathbf{v}$ and $\Delta\mathbf{i}$ based on practical considerations); $\mathbf{x} \preceq \mathbf{y}$ means that each element of $\mathbf{x}$ is no greater than the corresponding element in $\mathbf{y}$. We note that, if $i \notin \mathcal{N}_{\mathrm{a}}$, $\Delta V_i = 0$. In practice, the attacker can obtain the settings of $\Delta\mathbf{v}$ and $\Delta\mathbf{i}$ by launching a data exfiltration attack [Symantec 2014]. In the absence of such knowledge, the attacker must choose stringent values for these quantities such that the attack is not detected by the data-quality checks.

Based on the above conditions for the compromised measurements, we now formulate the efficiency and safety attacks.

4.2.1. Efficiency Attack. An efficiency attack causes an increase or decrease in the total instantaneous power injected or absorbed by the substations. In particular, we consider an aggressive attacker who aims to maximize or minimize such injected or absorbed power. Formally, the attacker solves the following constrained optimization problem to compute the attack vector $\{V_i', I_i' | \forall i \in \mathcal{N}_a\}$:

$$\max_{\{V_i', I_i' |\ \forall i \in \mathcal{N}_a\}} \sum_{i \in \mathcal{N}_{\mathrm{sub}}} V_i I_i \tag{14}$$

$$s.t. \qquad \text{constraints in (2) to (13).}$$

The above formulation captures the physical laws governing the power network and the substations (i.e., (2) to (4)), as well as how the attack vector induces the trains to make erroneous power control decisions (i.e., (7) to (11)). Specifically, for any $\{V_i', I_i' \ | \forall i \in \mathcal{N}_a\}$ satisfying (7) to (11), the attacker can predict the trains' power absorptions/injections $\{P_i = V_i' I_i' \ | \forall i \in \mathcal{N}_{\text{trains}}\}$. He then uses the physical laws in (2), (3), and (4) to determine the actual voltages and currents of the substations (i.e., $\{V_i, I_i \ | \forall i \in \mathcal{N}_{\text{sub}}\}$) and predict the system's total power consumption $\sum_{i \in \mathcal{N}_{\text{sub}}} V_i I_i$.

Solving the constrained optimization problem in (14) can be computationally expensive, mainly because the constraints in (7) and (8) are non-smooth and non-differentiable. Existing constrained non-linear optimization solvers (e.g., the `fmincon` function of MATLAB) often require the objective and constraint functions to be smooth. To use these existing solvers, the attacker can adopt a *divide-and-conquer* approach that splits the problem (14) into multiple subproblems in which a piece of (7) or (8) is selected as a constraint for a train. By comparing the optimization results of all the subproblems, the attacker can obtain a global optimal solution to the problem in (14). Because each train has three choices in (7) or (8), this approach will generate a total of $3^{|\mathcal{N}_a|}$ subproblems, where $|\mathcal{N}_a|$ is the number of trains under FDI attacks. As the subproblems are mutually independent, the attacker can solve the subproblems in parallel, to reduce computation time. We note that the ability to solve the problem in (14) in real time can be important to the attacker. This is because, to accumulate large energy loss, the attacker needs to keep at the FDI attacks by solving (14) continually, based on the latest system state given by s, P_i^d, and P_i^c. The attacker will need to procure sufficient computing resources for achieving the real-time objective. A resource-constrained attacker can inject a suboptimal attack that does not require extensive computations like solving (14). In Section 4.3, we present a numerical example to show that such an attack can still cause a considerable performance degradation.

4.2.2. Safety Attack. For safety attacks, we model the space of attack vectors that can cause the voltages at a subset of the TPS nodes, denoted by $\mathcal{N}_{\text{unsafe}}$, to cross the safety limits in (1). The attack space is defined by all the constraint conditions in the optimization problem (14), and $V_i \notin [V_{i,\min}, V_{i,\max}], i \in \mathcal{N}_{\text{unsafe}}$. As long as the attacker can find an attack vector satisfying the above constraints, he will be able to achieve the safety breaches.

We now discuss a heuristic approach that the attacker can use to aggressively increase the extent of the safety breaches. Specifically, the attacker maximizes the total power injected into the TPS by the regenerating trains, i.e., $\sum_{i \in \mathcal{N}_{\text{reg}}} V_i I_i$, subject to all the constraints of the optimization problem in (14). The intuition is that injecting more power into the TPS will result in higher catenary voltages. This constrained optimization problem can also be solved by the aforementioned divide-and-conquer approach.

The TPS under FDI attacks can be analyzed using the same set of equations as in Section 3 (i.e., (1)-(6)), except that the train's overcurrent and squeeze control decisions are now computed based on the attacker's injections V_i' (in (5) and (6)). Based on this analysis, in Section 4.3 we present numerical example to show the impact of efficiency and safety attacks. We also present time-domain simulation results in Section 7 considering realistic running profiles of trains.

4.3. Numerical Examples

We now present numerical examples to illustrate the efficiency and safety attacks. These examples are based on the TPS shown in Fig. 1, in which both trains are decelerating and regenerating. The system model parameters are given in Table IV. We consider a time instant at which the system state in the absence of attack is given by the first part of Table II, where the total instantaneous power absorbed by the sub-

Parameters	V_{NL}	γ	R_s	$V_i^{max,Tr}$	V_i^{max}
Value	750V	30mΩ/km	29.56mΩ	850V	900V

Table I: TPS model parameters.

stations and injected back into the supporting ac power grid is 3.601 MW. In these examples, we assume that the attacker can only compromise the voltage and current measurements of the train at node 2.

4.3.1. Efficiency Attack. The attacker solves the constrained optimization problem in (14) and tampers with V_2 and I_2 accordingly. We set $\Delta V_i = 50$V and $\Delta I_i = 200$A, $\forall i \in \mathcal{N}_a$. The compromised measurements and the true state of the system under attack are given in the second part of Table II. We can see that the compromised voltage measurement at node 2 is greater than the true value. Consequently, the train injects less power into the TPS because of the squeeze control, resulting in less power absorption by the substations. Specifically, the total power absorption is 2.888 MW, a 20% reduction compared with the case of no attack. Thus, the power efficiency of the system is degraded.

We also consider a suboptimal attack in which the attacker compromises the voltage of the train at node 2 by 20 V (hence $V_i' = V_i + 20V$). Under this attack, the total power absorption is 3.25 MW, a 9.5%, reduction compared with the case of no attack. This shows that the attacker can still cause a considerable degradation in the system efficiency by injecting a suboptimal attack. In practice, the attacker can tune his attack strategy to balance between the attack impact and the computational complexity of computing the attacks.

4.3.2. Safety Attack. The attacker uses the heuristic approach in Section 4.2.2 to compute the safety attack. The compromised measurements and the true system state are given in the third part of Table II. The compromised voltage measurement at node 2 is lower than its true value. Thus, the train at node 2 injects more power into the TPS because of the squeeze control, causing the actual voltage at node 2 to exceed the safety limit. We can see that it is possible for an attacker to tamper with the measurements of a single train and already achieve a safety attack. In this example, since both the trains are regenerating, the catenary voltages are closer to the safety limit. This makes it easier for the attacker to achieve the safety attack. Thus, for an attacker with limited write access to the trains' measurements (i.e., a small set $\mathcal{N}_a$), he can continuously monitor the system and wait for feasible moments for launching safety attacks.

5. GLOBAL ATTACK DETECTION

As discussed in Section 1, dc TPSes mainly rely on trains' local controls (i.e., overcurrent and squeeze controls) to avoid unsafe states. The TPS does not otherwise cross-check sensor data from different trains to ensure the data's global consistency. However, such global monitoring is clearly advantageous, because anomalies in the data relationships can help flag the occurrence of an FDI attack. Furthermore, not only can we cross-check sensor measurements from different trains, we can also check these measurements against an *a priori* global TPS model to ensure agreement. An attacker that wishes to remain stealthy under global monitoring thus becomes more constrained, and his actions may become less effective. In this section, we present the design of a global monitor for detecting FDI attacks under the Kerckhoffs's assumption, which we will subsequently refer to as the global attack detector (GAD).

Fig. 3 overviews our global attack detection approach. In the approach, the trains' voltage, current, and position measurements are sent to a central *TPS monitor* peri-

Node		1	2	3	4	Efficiency Loss
TPS State (Without Attack)	s_i	0	0.9	1.2	2	—
	P_i^c	-	5.5	1.8	-	
	V_i	815.6	875.5	867.7	815	
	I_i	-2218.8	3079.2	1338	-2198.4	
	P_i	-1.81	2.696	1.161	-1.792	
Efficiency Attack (Optimal)	V_i'	-	888.6	-	-	20 %
	I_i'	-	1409.6	-	-	
	V_i	801.1	847.7	850	805.2	
	I_i	-1728.2	1477.6	2117.6	-1867.1	
	P_i	-1.384	1.253	1.8	-1.503	
Efficiency Attack (Suboptimal)	V_i'	-	881.9	-	-	9.5 %
	I_i'	-	1409.6	-	-	
	V_i	808.5	861.9	859.1	810.1	
	I_i	-1979.2	2301.1	1715	-2036.5	
	P_i	-1.606	1.98	1.47	-1.65	
Safety Attack	V_i'	-	862.9	-	-	—
	I_i'	-	4731.6	-	-	
	V_i	828.9	**901**	884.2	824.1	
	I_i	-2669.1	4531.6	643.2	-2505.7	
	P_i	-2.212	4.083	0.569	-2.065	

Table II: System state and compromised measurements under efficiency and safety attacks. Distance is measured in kilometers, voltage in volts, current in amperes, and power in megawatts.

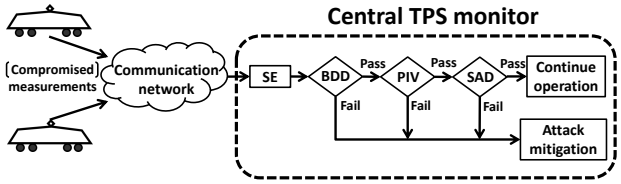


Fig. 3: Global attack detection. SE: State estimation; BDD: Bad data detection; PIV: Position integrity verification; SAD: Secondary attack detection.

odically. As Fig. 3 illustrates, the TPS monitor applies state estimation (SE), bad data detection (BDD), position integrity verification (PIV), and secondary attack detection (SAD) in sequence to detect attacks. In ac utility power grids, similar SE and related BDD are widely used for detecting faulty data or reducing the impact of noisy sensor measurements [Wood and Wollenberg 1996]. In Section 5.1, we propose a new BDD design that is specific to the application domain of dc TPS. By checking the consistency among measurements based on prior knowledge of the TPS, the BDD can detect a range of FDI attacks. However, the detection is not complete – an attacker under the

Kerckhoffs's assumption will be able to bypass it using his knowledge of the system. In Section 7, we provide numerical results to illustrate the impacts of these stealthy efficiency and safety attacks. To counter the stealthy attacks, in Section 5.3 we further propose a novel SAD algorithm to supplement the BDD, under an additional assumption that the trains' position data is intact, which is ensured by the PIV.

5.1. TPS Bad Data Detection and Its Vulnerability

Recall that in Section 4.2, the trains apply local controls based on their own voltage and current measurements only. Hence, the trains' position information does not matter. Under global detection, however, compromise of the trains' position information becomes relevant, since it may enable the attacker to mislead the TPS monitor into deriving an incorrect TPS model that is consistent with the compromised voltage and current measurements. Tampering with the position data can thus help the attacker evade detection. Although in practice it is extremely difficult for the attacker to hide the compromise of train position data because multiple sources of this data are often available (see Section 5.3 for the details), in this section, for generality, we account for possible compromise of the position data.

We use $\tilde{x}$ to represent a possibly compromised measurement x, i.e., $\tilde{x} = x$ in the absence of attack and $\tilde{x} = x'$ in the presence of attack. The state of the TPS is the vector of the nodal voltages, i.e., $\mathbf{v}$. The set of measurements includes nodal positions $\tilde{\mathbf{s}} = [\tilde{s}_1, \ldots, \tilde{s}_N]^T \in \mathbb{R}^{N\times 1}$, and nodal voltage and current readings $\tilde{\mathbf{z}} = [\tilde{\mathbf{v}}, \tilde{\mathbf{i}}]^T \in \mathbb{R}^{2N\times 1}$. In the absence of attack, the measurement vector $\mathbf{z}$ is related to the system state $\mathbf{v}$ as $\mathbf{z} = \mathbf{H}(\mathbf{s})\mathbf{v} + \mathbf{n}$, where $\mathbf{H}(\mathbf{s}) = [\mathbb{I}_N; \mathbf{Y}(\mathbf{s})]$ is a *measurement matrix* depending on the positions $\mathbf{s}$, $\mathbb{I}_N$ is an N-dimensional identity matrix, and $\mathbf{n} \in \mathbb{R}^{2N\times 1}$ is a random measurement noise vector. We assume that $\mathbf{n}$ follows a multivariate Gaussian distribution. The maximum likelihood (ML) estimate of $\mathbf{v}$, denoted by $\hat{\mathbf{v}}$, is given by [Wood and Wollenberg 1996, Chap. 12]

$$\hat{\mathbf{v}} = (\mathbf{H}(\tilde{\mathbf{s}})^T \mathbf{\Sigma}^{-1} \mathbf{H}(\tilde{\mathbf{s}}))^{-1} \mathbf{H}(\tilde{\mathbf{s}})^T \mathbf{\Sigma}^{-1} \tilde{\mathbf{z}},$$

where $\mathbf{\Sigma}$ is the covariance matrix of $\mathbf{n}$. The SE's BDD raises an alarm if

$$(\tilde{\mathbf{z}} - \mathbf{H}(\tilde{\mathbf{s}})\hat{\mathbf{v}})^T \mathbf{\Sigma}^{-1} (\tilde{\mathbf{z}} - \mathbf{H}(\tilde{\mathbf{s}})\hat{\mathbf{v}}) > \tau,$$

where τ is a constant threshold that can be determined to meet a given false alarm rate under random measurement noise[3]. The BDD is originally designed to detect faulty sensor data caused by natural malfunction of sensors. Thus, it is effective in detecting a range of FDI attacks that are not specifically designed to bypass it. However, the attacker that we consider in this paper, following the Kerckhoffs's principle, will be able to design FDI attacks with the objective of bypassing the BDD. In the following, we formulate these stealthy safety and efficiency attacks.

From an existing result [Liu et al. 2009], if the compromised measurement vector $\mathbf{z}'$ is in the column space of the compromised measurement matrix $\mathbf{H}(\mathbf{s}')$, $\mathbf{z}'$ can bypass the BDD. Applying this result to the TPS, we have the following lemma.

LEMMA 5.1. *Any compromised measurements that satisfy*

$$\mathbf{Y}(\mathbf{s}')\mathbf{v}' = \mathbf{i}' \tag{15}$$

can bypass the BDD.

PROOF. Lemma 5.1 holds since any compromised measurement vector $\mathbf{z}'$ that satisfies (15) lies in the column space of $\mathbf{H}(s')$, i.e., $\mathbf{z}' = [\mathbf{v}', \mathbf{i}']^T = [\mathbb{I}_N; \mathbf{Y}(\mathbf{s}')]\mathbf{v}' = \mathbf{H}(\mathbf{s}')\mathbf{v}'$. □

[3]A detailed description of how to set the BDD threshold is given in Appendix A.

In addition to (15), the TPS monitor may use two other sensor data checks. First, to meet the constraint in (15), the attacker may need to compromise the voltage and current measurements at the substations. The TPS monitor may check the substation measurements, i.e., V_i and I_i, $\forall i \in \mathcal{N}_{\text{sub}}$, against the substation model in (3). To be stealthy to this check, the attacker can impose an additional constraint of

$$V_i' = V_{\text{NL}} - R_s I_i', \quad \forall i \in \mathcal{N}_{\text{sub}}. \tag{16}$$

Second, the TPS monitor can also apply data quality checks similar to those in (12) and (13) to check the trains' position measurements. Thus, if the attacker can compromise the position measurements, he needs to satisfy

$$\mathbf{s} - \Delta\mathbf{s} \preceq \mathbf{s}' \preceq \mathbf{s} + \Delta\mathbf{s}, \tag{17}$$

where $\Delta\mathbf{s} = [\Delta s_1, \ldots, \Delta s_N]^T$ are the maximum allowed errors for position measurements and $\Delta s_i = 0$ if $i \notin \mathcal{N}_{\text{a}}$.

Therefore, the efficiency attacks that are stealthy to the BDD can be computed by solving the constrained optimization problem (14) with the additional constraints (15), (16), and (17). Similarly, the attack space for BDD-stealthy safety attacks is characterized by the constraints of the optimization problem (14), $V_i \notin [V_{i,\min}, V_{i,\max}], i \in \mathcal{N}_{\text{unsafe}}$, and the additional constraints (15), (16), and (17). Naturally, BDD reduces the attack space since the attacker now needs to satisfy additional constraints to remain undetected. In the simulation results presented in Section 7, we show that, under a realistic TPS setting, the BDD significantly reduces the impact of attacks.

5.2. Numerical Examples

We now present numerical examples to illustrate the efficiency and safety attacks that can bypass the BDD as analyzed in Section 5.1. The TPS model and parameters are identical to those in Section 4.3. The true system state and the compromised measurements are given in Table III. We set $\Delta s_i = 0.6\,\text{km}$, $\forall i \in \mathcal{N}_a$. To illustrate a powerful attacker, we assume that the attacker can corrupt the voltage and current measurements of all the four nodes in Fig. 1, as well as the positions of both the trains.

5.2.1. Efficiency Attack. Under the efficiency attack, the total power injected back to the supporting power grid by the substations is 3.431 MW, which is a reduction of about 4.7% compared with no attacks. This reduction is much less than the 20% caused by the efficiency attack in Section 4.3, which was achieved by compromising the voltage and current measurements of node 2 only in the absence of BDD. This result illustrates the ability of the BDD in limiting the impact of efficiency attacks.

5.2.2. Safety Attack. We observe that by compromising the nodal measurements and the trains' position information, the attacker can increase the voltage at node 2 to $901.4\,\text{V}$ while bypassing the BDD. Furthermore, if the attacker can gain write access to any one train (i.e., $|\mathcal{N}_a| = 1$), he cannot launch a successful safety attack. This is in contrast to the example in Section 4.3, where the attacker could launch a successful safety attack by compromising the measurements of a single train only.

In summary, the above examples suggest that the global monitoring and BDD can significantly limit the impact of stealthy FDI attacks on the TPS even if the attacker can compromise the measurements of multiple trains. To accomplish a safety attack, the attacker needs to compromise more trains compared with no BDD.

5.3. Secondary Attack Detection (SAD)

In this section, we propose a novel secondary attack detection (SAD) algorithm that can effectively detect the *onset* of an FDI attack that has bypassed the BDD. A requirement for the SAD is that the trains' position data communicated to the TPS monitor is

Node		1	2	3	4
Efficiency Attack	s_i'	0	1	1	2
	V_i'	812	874.9	874.9	812
	I_i'	-2096.7	3159	1034.5	-2096.8
	V_i	813.2	871	861.7	811.6
	I_i	-2138.8	3173.2	1050.4	-2084.8
	P_i	-1.739	2.764	0.905	-1.692
Safety Attack	s_i'	0	0.43	1.8	2
	V_i'	835	872.3	847.3	830.9
	I_i'	-2876.8	3487.8	2124.5	-2735.4
	V_i	829.1	**901.4**	895.1	830.1
	I_i	-2676.8	3375.3	2010.9	-2709.4
	P_i	-2.219	3.043	1.8	-2.249

Table III: System state and compromised measurements under efficiency and safety attacks that have bypassed the BDD. Distance is measured in kilometers, voltage in volts, current in amperes, power in megawatts.

intact. It is feasible for the TPS monitor to verify the integrity of the position data. For example, real-world railway systems invariably provide multiple sources of train position information including train-borne wheel sensors and GPS, track-side Balise [Alstom 2001], etc. By cross-checking position measurements from the multiple sources, we can readily identify FDI attacks on the position data unless the attacker succeeds in compromising all the data sources, which is highly challenging since these sensors use technologies that are significantly different from each other. For example, GPS is a satellite-based system, Balise uses electronic beacon or transponder placed between the rails, etc. Such cross checks constitute the PIV illustrated in Fig. 3. Given that TPS is a safety-critical system, the operator should enforce the highest consistency requirement on the position measurements from different sensors, i.e., if any inconsistency is found among different position sensors' readings, the PIV should raise a fault/attack alarm. If FDI attacks on the position data are identified, the TPS should immediately apply attack mitigation such as the approach discussed in Section 5.4.

Note that the analysis in the previous sections is for a particular time instant, and the attacker can use the techniques in Sections 4 and 5.1 to launch attacks continually over time. Once the SAD detects an attack's onset, the system can activate the attack mitigation approach in Section 5.4 to render subsequent FDIs ineffective. Thus, in this section we focus on analyzing the property of the system and designing the SAD accordingly for the onset time instant only of an attack.

5.3.1. A Discrete Solution Property. The requirement of intact position data and the design of the SAD algorithm are based on a key observation as follows. If the attacker can compromise the trains' position data, the three equality conditions (9), (15), and (16) that the attacker must obey form an underdetermined problem with $3N$ variables and $2N$ equations. Since the other conditions that the attacker needs to follow (i.e., (7), (8), (10) to (13), and (17)) are inequalities, the attacker's problem of finding stealthy FDI attack vectors most likely has infinitely many solutions that are continuous. However, if the trains' position data is intact, the three equality conditions (9), (15), and (16) with s' replaced by the known s, will form a determined problem with $2N$ variables and $2N$ equations. As a result, the attacker's problem has a finite number of discrete

solutions (which we will prove shortly) and the attacker must choose one of them that is different from the true measurement vector. In what follows, we first show that with intact position data, there are only a finite number of discrete solutions that satisfy the BDD-passing conditions. We then describe the SAD algorithm.

The BDD bypass conditions given by (15) and (16) can be compactly represented as

$$\mathbf{V}\left(\mathbf{Y}(\mathbf{s})+\mathbf{G}\right)\mathbf{v}=\mathbf{c}, \tag{18}$$

where the $(i,j)^{\text{th}}$ elements of the matrices $\mathbf{V}\in\mathbb{R}^{N\times N}$ and $\mathbf{G}\in\mathbb{R}^{N\times N}$ are given by

$$\mathbf{V}(i,j)=\begin{cases}V_i, \text{if } i=j \text{ and } i\in\mathcal{N}_{\text{trains}},\\ 1, \text{otherwise},\end{cases}, \qquad \mathbf{G}(i,j)=\begin{cases}R_s^{-1}, \text{if } i=j \text{ and } i\in\mathcal{N}_{\text{sub}},\\ 0, \text{otherwise},\end{cases},$$

and the i^{th} element of the vector $\mathbf{c}\in\mathbb{R}^{N\times 1}$ is given by

$$\mathbf{c}(i)=\begin{cases}P(i), \text{ if } i\in\mathcal{N}_{\text{trains}},\\ \frac{V_{\text{NL}}}{R_s}, \text{ if } i\in\mathcal{N}_{\text{sub}}.\end{cases}$$

Equation (18) is a system consisting of N polynomial equations with N variables. Such a system of equations is referred to as a *square polynomial* system, and the Bezout's theorem provides an upper bound on the number of solutions for such systems. The Bezout's theorem is as follows.

THEOREM 5.2. *(Bezout's Theorem) [Sottile 2011] For a square polynomial system, the bound on the number of complex solutions is at most the product of the degrees of the polynomials.*

The existence of the upper bound proves that the system of polynomials in (18) has a finite number of discrete solutions. For the BDD bypass condition in (18), we have a polynomial constraint corresponding to each train in the system ($\mathcal{N}_{\text{tra}}\cup\mathcal{N}_{\text{reg}}$), and each one is a second degree polynomial. Thus, the upper bound according to Bezout's theorem would be $2^{|\mathcal{N}_{\text{tra}}\cup\mathcal{N}_{\text{reg}}|}$. However, in practice, we found that several solutions to the square polynomial system were complex, which we can discard (since the voltages in a dc system cannot be complex).

We now provide a numerical example to illustrate this property. In this example, we use the TPS shown in Fig. 1 with the settings listed in Table IV and $P_2=P_3=-0.3$ MW. The two curves in Fig. 4 correspond to the two equality conditions that V_2' and V_3' need to satisfy to bypass the BDD. Their intersections are the solutions to the attacker's problem of finding stealthy attack vectors. We can see that the solutions are discrete.

$\mathrm{V_{NL}}$	750 V
γ	0.03 Ω km
$\mathrm{R_s}$	0.02956 Ω
$\mathrm{V}_i^{\max,\mathrm{Tr}}$	850 V
$\mathrm{V}_i^{\max}$	900 V

Table IV: TPS parameters.

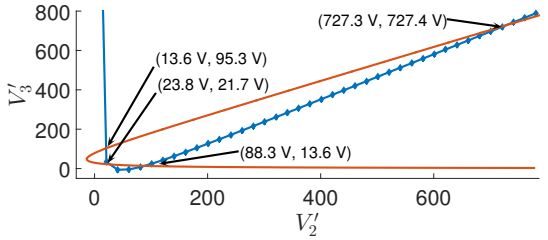


Fig. 4: A numerical example illustrating discrete solution property.

5.3.2. SAD Algorithm. Based on the discrete solution property, we design the SAD algorithm as follows.

Inputs: Trains' true positions $\mathbf{s}$, possibly compromised measurement vector $\tilde{\mathbf{z}}$, intact nodal voltage vector $\mathbf{v}_{\text{pr}}$ at the previous time instant
Output: Attack onset detection result

1. Using $\tilde{\mathbf{z}}$, compute $P_i = \tilde{V}_i \tilde{I}_i,\ i \in \mathcal{N}_{\text{trains}}$.
2. Solve the following constrained optimization problem

$$J^* = \min_{\substack{\mathbf{v}_a, \mathbf{v}_b \\ \mathbf{v}_a \neq \mathbf{v}_b}} \ ||\mathbf{v}_a - \mathbf{v}_b||_p \tag{19a}$$

$$s.t. \quad \mathbf{V}_a \left(\mathbf{Y}(\mathbf{s}) + \mathbf{G}\right) \mathbf{v}_a = \mathbf{c}, \tag{19b}$$

$$\mathbf{V}_b \left(\mathbf{Y}(\mathbf{s}) + \mathbf{G}\right) \mathbf{v}_b = \mathbf{c}, \tag{19c}$$

   where $||\mathbf{x}||_p$ represents the p-norm of a vector $\mathbf{x}$
3. Extract $\widetilde{\mathbf{v}}$ from $\widetilde{\mathbf{z}}$. If $||\widetilde{\mathbf{v}} - \mathbf{v}_{\text{pr}}||_p \leq J^*$, report no attack; Otherwise, report onset of attack.
4. If there only exists two discrete points in the solution set, $J^* = \alpha J^*$

In Step 1 of the algorithm, given the possibly compromised measurement vector $\tilde{\mathbf{z}}$, the TPS monitor computes the actual power absorption or injection of each train. We note that this follows from (4). Based on the trains' true positions $\mathbf{s}$ and powers, in Step 2, the TPS monitor solves the constrained optimization problem (19). The constraints in (19b) and (19c) are compact representations of the BDD bypass condition as explained in (18), for two distinct solutions $\mathbf{v}_a$ and $\mathbf{v}_b$. By the observation that the BDD bypass condition given the trains' true positions has discrete solutions, $\mathbf{v}_a$ and $\mathbf{v}_b$ that solve the optimization problem (19) are two distinct solutions that are closest to each other (among all such pairs of solution vectors) and the J^* given by (19) is the minimum distance.

In Step 3, the TPS monitor compares the J^* with the p-norm distance between the possibly compromised voltage measurement vector and the intact nodal voltage vector $\mathbf{v}_{\text{pr}}$ at the previous time instant, to determine the possible onset of an attack. This step is based on that if the attacker launches a BDD-stealthy attack without tampering with the trains' position information, the p-norm distance between the compromised voltage vector and the voltage vector in the absence of attack must be no less than J^*. As the voltage vector in the absence of attack is unknown to the TPS monitor, a *practical approach* is to use the $\mathbf{v}_{\text{pr}}$ that is not compromised before the onset of the attack. Since the TPS monitor can run the SAD periodically and frequently (e.g., every second), the TPS state will not change significantly over one monitoring time internal. In Section 7, extensive simulations demonstrate the effectiveness of this practical approach by comparing it with an *oracle approach* that uses the voltage vector at the present time instant in the absence of attack in Step 3. If and when the onset of an attack is detected, the TPS switches to an attack mitigation mode, as discussed in Section 5.4, to prevent safety breaches. In step 4, we scale the value of J^* by a parameter $\alpha \in [0, 1]$ in case there are only two discrete solutions that satisfy the BDD-passing condition. Step 4 is introduced to reduce the MDs of the SAD in the presence of sensor measurement noises (the rationale behind the introduction of this parameter will be explained in Section 6).

5.4. Attack Mitigation

We outline an approach to mitigating the impact of an attack that has been detected by the TPS monitor by the BDD, PIV, or SAD. On detecting the onset of the attack, the system switches to an *attack mitigation mode* in which the TPS monitor issues power absorption/injection commands to the trains to replace their local overcurrent/squeeze

controls. Specifically, the TPS monitor computes the P_i for each train based on the trains' power demands, regeneration capacities, and positions by solving the electrical models and trains' local control laws presented in Section 3. Note that the trains can report their power demands and regeneration capacities to the TPS monitor. The TPS monitor can also estimate them based on trains' running profiles that are often fixed during the planning phase. If FDI attacks on trains' position information have been detected, the system can estimate the trains' positions based on their running profiles. Each train applies the P_i received from the TPS monitor. The core idea of this mitigation approach is to run the TPS temporarily based on models rather than compromised sensor measurements. Emergent running profiles that stop the trains safely should be applied immediately once the system enters the attack mitigation mode.

6. IMPACT OF SENSOR MEASUREMENT NOISE

In this section, we examine the performance of the GAD in the presence of sensor measurement noises. In this section, we consider the additive Gaussian noise model described in Section 5.1. Sensor measurement noise provides the attacker an opportunity to hide its attack by masquerading false measurements as legitimate noisy measurements, leading to MDs and FPs. MDs may result in the loss of system efficiency or safety breaches. On the other hand, FPs result in the system operator initiating unnecessary mitigation steps that may degrade performance. We now formally define FPs and MDs for the BDD, SAD, and GAD respectively.

We consider two hypotheses: H_0 denotes that the system is not under attack, and H_1 denotes that the system is under attack. We let Z_{BDD} and Ξ_{BDD} represent the indicator variables for the occurrences of FP and MD, respectively, in the BDD, and Z_{SAD} and Ξ_{SAD} represent the corresponding quantities for the SAD. They can be mathematically stated as

$$Z_{\text{BDD}} = 1_{\{(\tilde{\mathbf{z}}-\mathbf{H}(\tilde{\mathbf{s}})\hat{\mathbf{v}})^T\mathbf{\Sigma}^{-1}(\tilde{\mathbf{z}}-\mathbf{H}(\tilde{\mathbf{s}})\hat{\mathbf{v}})>\tau \ \big| H_0\}}, \qquad Z_{\text{SAD}} = 1_{\{||\widetilde{\mathbf{v}}-\mathbf{v}_{\text{pr}}||_p>J^* \ \big| H_0\}},$$
$$\Xi_{\text{BDD}} = 1_{\{(\tilde{\mathbf{z}}-\mathbf{H}(\tilde{\mathbf{s}})\hat{\mathbf{v}})^T\mathbf{\Sigma}^{-1}(\tilde{\mathbf{z}}-\mathbf{H}(\tilde{\mathbf{s}})\hat{\mathbf{v}})\leq\tau \ \big| H_1\}}, \qquad \Xi_{\text{SAD}} = 1_{\{||\widetilde{\mathbf{v}}-\mathbf{v}_{\text{pr}}||_p\leq J^* \ \big| H_1\}},$$

where $1_{\mathcal{A}}$ is an indicator function given by $1_{\mathcal{A}} = 1$ if $\mathcal{A}$ is true, or 0 otherwise. Similarly, we define Z_{GAD} and Ξ_{GAD} for the GAD. Since the GAD serializes the BDD and SAD, it will raise an alarm if one of the following two events occurs: i) the BDD raises an alarm; or ii) if the measurements pass the BDD but the SAD raises an alarm. Thus Z_{GAD} can be expressed in terms of Z_{BDD} and Z_{SAD} as

$$Z_{\text{GAD}} = Z_{\text{BDD}} \ \vee (\neg Z_{\text{BDD}} \wedge \ Z_{\text{SAD}}).$$

Similarly, Ξ_{GAD} can be expressed in terms of Ξ_{BDD} and Ξ_{BDD} as

$$\Xi_{\text{GAD}} = \Xi_{\text{BDD}} \ \wedge \ \Xi_{\text{SAD}}. \tag{20}$$

Next, we use a numeric example to illustrate FPs and MDs in the cases of BDD and SAD, respectively, for a representative TPS network.

— **BDD FPs and MDs:** The BDD's FPs and MDs are caused by fluctuations in the residual value $(\tilde{\mathbf{z}} - \mathbf{H}(\tilde{\mathbf{s}})\hat{\mathbf{v}})^T\mathbf{\Sigma}^{-1}(\tilde{\mathbf{z}} - \mathbf{H}(\tilde{\mathbf{s}})\hat{\mathbf{v}})$, which are in turn caused by the measurement noises. This is illustrated in Fig. 5 for a TPS model identical to that in Section 5.3, considering 1000 realizations of measurement noise sampled from an i.i.d. zero-mean Gaussian distribution with a standard deviation set to 0.3% of the full-scale voltage [Smitt 2016] and current sensor readings, respectively. (The full-scale voltage and current readings are 900 V and $2,500$ A, respectively.) To generate TPS measurements under H_1, we inject an additive attack of 20 V to the voltage measurement of node 2. It can be seen that in the absence of measurement noise, the value

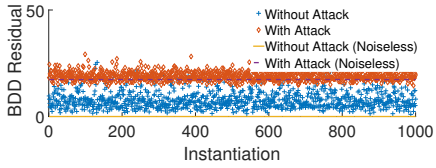


Fig. 5: BDD residual under hypothesis H_0 and H_1 with and without sensor measurement noise.

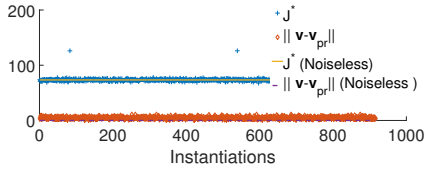


Fig. 6: J^* and $||\mathbf{v} - \mathbf{v}_{\text{pr}}||$ under hypothesis H_0 with and without sensor measurement noise.

of the residual is 0 under H_0, and non-zero under H_1. Therefore, any occurrence of a non-zero residual indicates the presence of an attack in a noiseless environment. However, in the presence of sensor measurement noise, the value of the residual fluctuates under different noise instantiations. Thus, differentiating measurements under attack from those under natural measurement noise becomes challenging.

— **FPs and MDs of SAD under the Oracle Approach:** The SAD's FPs and MDs are due to fluctuations in the value of J^* and $||\mathbf{v} - \mathbf{v}_{\text{pr}}||$ under different realizations of the measurement noise, as illustrated in Fig. 6, for hypothesis H_0. (Recall from Algorithm 1 that in this case, noisy voltage and current measurements are used as inputs to the SAD algorithm.) Note from Fig. 6 that in the absence of measurement noise, both these quantities have a fixed value, in contrast to the case of noisy measurements. Thus, in the presence of noise, an FP is declared whenever there is no attack on the system and $||\mathbf{v} - \mathbf{v}_{\text{pr}}|| \geq J^*$, and MD is declared whenever the system is under attack and $||\mathbf{v} - \mathbf{v}_{\text{pr}}|| < J^*$.

— **FPs and MDs of SAD under the Practical Approach:** Another factor that contributes to the occurrence of FPs and MDs in the practical approach is the following. Since $\mathbf{v}_{\text{pr}}$ is estimated based on the historical measurements, whenever there is a sudden change in the system state between successive time slots, the difference $||\mathbf{v} - \mathbf{v}_{\text{pr}}||$ can become large and result in FPs. In the simulations presented in Section 7.3, we observe that when one or more trains in the TPS change status from tractioning mode to breaking mode, there is a large change in the TPS system state.

To assess the performance of the proposed detectors, we examine their receiver operating characteristics (ROC) curve, obtained by varying the BDD's detection threshold τ, and the adaptive parameter of the SAD algorithm α. Each value of the parameter τ (and α) yields certain FP and MD rates, which are the x and y-axes of the ROC curve, respectively. We consider three levels of the measurement noise by varying its standard deviation from 0.1% to 0.3% of the full-scale current and voltage sensor readings. We consider two different attacks: (i) In Fig. 7a and Fig. 7b, we plot ROC curves for attacks designed without imposing the BDD-passing condition (we refer to it subsequently as *random attack*). In particular, we inject an additive attack of 20 V to the voltage measurement of node 2. (ii) In Fig. 7c and Fig. 7d, we plot the ROC curves for BDD-stealthy attacks.

The ROC curves under the two attacks exhibit different characteristics, which can be explained as follows. As evident from Fig. 7a and Fig. 7c, the BDD is effective in detecting random attacks but ineffective in detecting the BDD-passing attacks (in fact, the detection rate of BDD-stealthy attacks is 0). This behavior can be explained by the nature of BDD's design. Further, by comparing Fig. 7a and Fig. 7b, we can conclude that the SAD only marginally improves the detection rate of random attacks compared with the stand-alone BDD detector. However, when we compare Fig. 7c and Fig. 7d, we observe that for the BDD-stealthy attacks, the presence of the SAD significantly

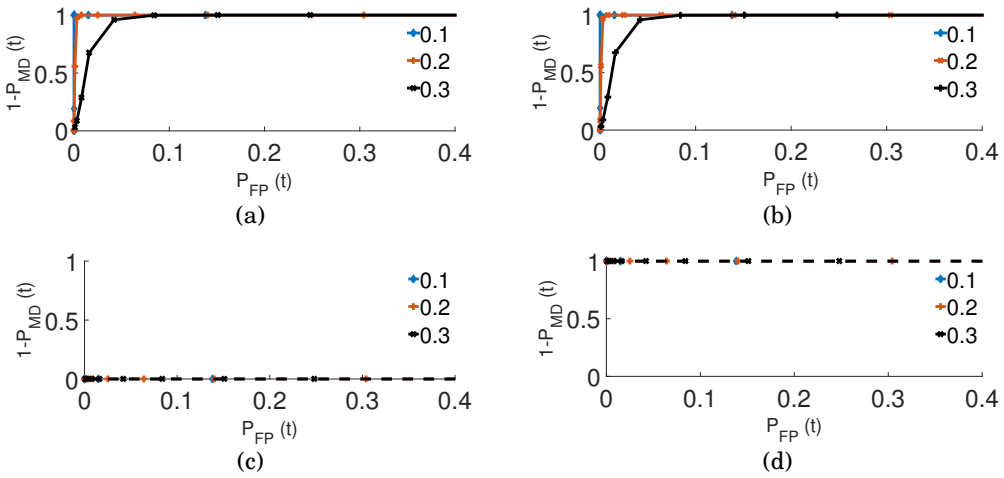


Fig. 7: (a) ROC curve of the BDD with random attack. (b) ROC curve of the GAD with random attack. (c) ROC curve of the BDD with BDD-stealthy attack. (d) ROC curve of the GAD with BDD-stealthy attack.

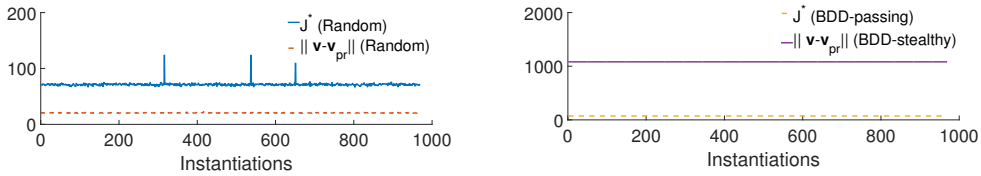


Fig. 8: Analysis of SAD under random and BDD-stealthy attacks.

improves the detection rate. Specifically, the GAD detection rate is 1 (no MDs). This shows the effectiveness of the SAD in detecting BDD-passing attacks.

To understand the performance of SAD in the two cases of random attack and BDD-passing attack, we plot the values of J^* and $||\mathbf{v} - \mathbf{v}_{\mathrm{pr}}||$ in Fig. 8. It can be seen that for random attacks that have been missed by the BDD, the value of $||\mathbf{v} - \mathbf{v}_{\mathrm{pr}}||$ is consistently lower than J^* for all the noise instantiations, which results in MDs. In contrast, for BDD-stealthy attacks, the value of $||\mathbf{v} - \mathbf{v}_{\mathrm{pr}}||$ is greater than J^*. This is because in the case of random attack, the attacker only manipulates the measurements from a few sensors (and thus the difference $||\mathbf{v} - \mathbf{v}_{\mathrm{pr}}||$ is not high). However, for the BDD-stealthy attacks, the attacker must manipulate the system measurements in a coordinated manner. In particular, for the system considered in the above simulations, we observe that the attacker must manipulate the current and voltage measurements of all the nodes. Consequently, the difference $||\mathbf{v} - \mathbf{v}_{\mathrm{pr}}||$ is high.

The FP and MD rates in the above examples are illustrated for a fixed TPS topology and parameters. The above discussions give basic understanding on the impact of random measurement noises on the performance of the attack detectors. However, as the trains change their positions and the status of motion, the TPS parameters change and consequently the FP and MD rates may vary. For instance, the practical GAD detector can have a high FP rate when one or more trains in the TPS changes its status of motion. In order to ensure that the proposed detectors have acceptable performance in these scenarios, in Section 7.3, we present an adaptation mechanism for the GAD

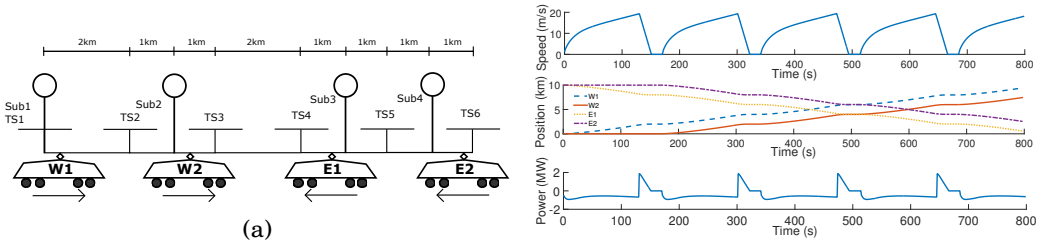


Fig. 9: (a) System set-up for simulations. Sub - Substations, TS - Train stations, W - Trains departing from the west, E - Trains departing from the east. (b) Train speed (top plot), position (middle plot), power demand, and regeneration capacity (bottom plot) over time. Power demand is negative and regeneration capacity is positive.

detector, which we call GAD with attack detection window (GAD-W). Extensive simulation results show that the GAD-W detector yields consistently low FP and MD rates for the varying TPS configurations.

We end this section by explaining the introduction of the scaling parameter α in Step 4 of Algorithm 1. Note that in the absence of measurement noise, for a TPS system under attack, the value of J^* is equal to $||\mathbf{v} - \mathbf{v}_{\mathbf{pr}}||$ whenever there are only two discrete solutions that satisfy the BDD-passing condition (since the attacker must choose the solution that is different from the true measurements as his attack vector). In such a scenario, the fluctuations in the value of J^* due to sensor measurement noises can often drive its value to greater than $||\mathbf{v} - \mathbf{v}_{\mathbf{pr}}||$, leading to a high MD rate. In order to avoid this, we scale down the value of J^* by a parameter $\alpha \in [0, 1]$. We note that this problem is unique to the case when there are only two discrete solutions that satisfy the BDD-passing condition, and hence no scaling of J^* is needed in the other cases.

7. SIMULATIONS

Our analyses in the previous sections address a particular time instant only. In this section, we conduct time-domain simulations with realistic running profiles of trains to illustrate the impact of FDI attacks. We also show the effectiveness of the BDD in reducing the impact of the attacks, and that of the SAD in detecting those attacks that are BDD-stealthy.

7.1. Simulation Settings and Methodology

As Fig. 9a illustrates, we simulate a TPS consisting of four trains (labeled W1, W2, E1, and E2), four substations (labeled Sub1 to Sub4), and six train stations (labeled TS1 to TS6). The parameters of the TPS are identical to those in Table IV. The positions of the substations and the train stations are shown in Fig. 9a. The trains W1 and W2 start their journeys from TS1 and travel from west to east, whereas the trains E1 and E2 start their journeys from TS6 and travel from east to west. The trains W1 and E1 depart at time zero and the trains W2 and E2 depart at the 170th second. At each of the train stations, the trains stop for a duration of 20 seconds. Each train follows the same speed profile as shown in the top part of Fig. 9b. The second plot of Fig. 9b shows the trains' positions over time. Each train switches between traction and braking modes during the simulation, and its power demand and regeneration capacity over time are shown in the bottom plot of Fig. 9b. This plot is derived based on mechanical energy consumption of the train under the specified running profile, and with an efficiency ratio of 70% for the traction mode [Shuai et al. 2015] and 40% for the braking mode

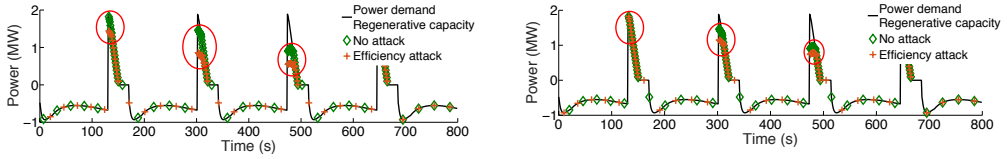
[Acikbas and Soylemez 2007] of converting kinetic energy into electrical energy. We simulate the TPS for 800 seconds at a time granularity of one second.

To simulate attacks, the attacker injects an attack vector computed using the methods given in Sections 4 and 5 every second. In the absence of BDD, the attacker compromises the voltage and current measurements of all the train nodes. In the presence of BDD, the attacker tampers with the voltage and current measurements of all the train and substations nodes as well as the position information of the train nodes. The position information of substations cannot be compromised since their locations are fixed and known *a priori*. The maximum errors that the attacker can introduce to the voltage, current, and position measurements, as described in (12), (13), and (17), are set as $\Delta V_i = 50\,\mathrm{V}$, $\Delta I_i = 200\,\mathrm{A}$, for $i \in \mathcal{N}_a$, unless otherwise specified. The choice of these parameters is made taking into account two practical considerations: (i) the measurement noise level (whose standard deviation is considered to be $\approx 0.5\%$ of the full-scale voltage and current sensor readings [Smitt 2016]) (ii) the change in voltage and current measurements of the TPS between any two successive simulation instants (which be observed to be in the considered range based on extensive simulations). Note that if the variation of voltage and current is within this range, they pass the data-quality checks.

The simulations are carried out in MATLAB. The constrained optimization problems are solved using the `fmincon` function of MATLAB with the `MultiStart` algorithm. In the absence of attack, to compute the system state, we use the `fmincon` with a constant objective function and the electrical models and trains' local control laws presented in Section 3 as the constraints. We also use the function to compute the safety attack vectors under the heuristic approach and the optimal efficiency attack vectors. If at any time instant, the `fmincon` function returns an attack vector that is the same as the true system state, the attacker does not launch an attack, since the attack will not have any impact. Step 2 of the SAD algorithm is also implemented using the `fmincon` function.

Although our analysis in this paper is general and applicable to a TPS network of arbitrary size and topology, for simulations we consider a small-scale TPS in Fig. 9a. The rationale is two fold. First, the attacker may find it difficult to coordinate his attacks on a large number of geographically distributed trains. Computing resources may present another barrier for large-scale attacks. A more credible scenario is for the attacker to focus on one or a few trains in a TPS section. Second, since real-world TPS networks are mostly radial [Abrahamsson 2012], the impact of a focused and localized attack will not propagate over long distances. In view of these factors, we use the small-scale TPS to represent well a TPS section in a large system.

Moreover, to simplify our simulations, we do not consider overcurrent control. Specifically, we set the triggering threshold $V^{\mathrm{Tr}}_{i,\min}$ to a low value, so that overcurrent control will not be activated. As a result, the trains' speed profiles will not change because the trains need not curtail their power consumption. At any time instant, therefore, a train's power consumption is equal to its power demand during acceleration. Because of this simplification, we do not simulate attacks on tractioning trains, which would alter the tractioning trains' power consumption and change their running profiles. Although we can simulate overcurrent control and attacks on tractioning trains by extending our simulator to admit changeable running profiles, the simulations reported in this paper already provide interesting understanding and insights into the impact of attacks and the effectiveness of countermeasures.



(a)Effect of efficiency attacks on Train E1 in the absence of BDD.

(b)Effect of efficiency attacks on Train E1 in the presence of BDD.

Fig. 10: Effect of efficiency attacks on Train E1. Circled regions highlight the time slots where the two curves (with and without attack) diverge. Note that the curve with attack follows the curve without attack more closely in the presence of BDD.

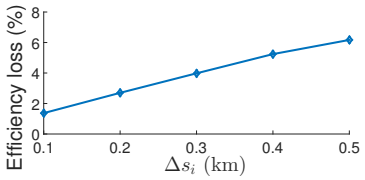
7.2. Simulation Results

7.2.1. Efficiency Attacks. The first set of simulations evaluates the impact of efficiency attacks on the TPS without BDD. Fig. 10a shows the power absorbed/injected by the train E1 in the presence and absence of attacks. We can see that the efficiency attacks cause the regeneration trains to inject less power into the power network (please see the encircled regions, e.g., from 302th to 315th second for the train E1). To calculate the loss in system efficiency, we ignore the time instants when all the trains are in traction mode, since we do not simulate attacks on the tractioning trains as discussed in Section 7.1. As a result, the efficiency attacks cause a reduction of 28.3% in the total energy adsorbed by the substations compared with the case of no attacks, during the time periods when there is at least one regenerating train under attack.

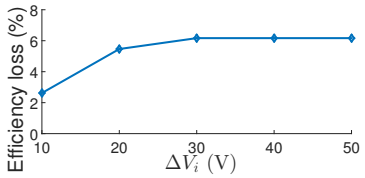
The second set of simulations evaluates the impact of efficiency attacks on the TPS with BDD. Similar to Fig. 10a, Fig. 10b shows the power absorbed/injected in the absence and presence of attacks. It can be seen that in Fig. 10 (b), the curve for the power absorbed/injected by trains in the presence of attacks follows that for the absence of attacks more closely, in comparison to the respective curves in Fig. 10 (a). (Please see the encircled parts of the two figures.) Thus, although the efficiency attack can still induce the regenerating trains to inject less power to the power network, it causes a reduction of 6.2% only in the total energy adsorbed by the substations, during the time periods when there is at least one regenerating train under attack. This is in contrast to the 28.3% for the TPS without BDD.

We also examine the effect of efficiency attacks on the TPS with BDD under different settings of Δs_i and ΔV_i in Fig. 11a and Fig. 11b, respectively. From these figures, we can see that at smaller settings of Δs_i and ΔV_i, the efficiency loss caused by the FDI attack diminishes. For instance, the efficiency loss is as low as 1.37% when $\Delta s_i = 0.1$ km. In practice, the TPS monitor can estimate the present train position based on the train's speed and its position at the previous time instant when it was known that there were no attacks. The present position reading can be compared with the estimated position using (17). The setting of Δs_i should consider natural errors of train positioning systems and the estimation error. Existing train positioning systems such as GPS and Balise can achieve an accuracy of five to ten meters [The Economic Times – Railways 2012], [Hartwig et al. 2006]. Thus, it is reasonable to assume that the combined effect of the train positioning system error and the estimation error is less than 0.1 km. Our results show that by properly tuning the BDD's attack detection parameters (e.g., Δs_i and ΔV_i), the efficiency loss caused by FDI attacks can be significantly reduced.

7.2.2. Safety Attacks. We conduct two sets of simulations to evaluate the impact of safety attacks on the TPS: the first one without BDD and the second with BDD. Under



(a)Impact of Δs_i ($\Delta V_i = 50\,\text{V}$, $\Delta I_i = 200\,\text{A}$)



(b)Impact of ΔV_i ($\Delta s_i = 0.5\,\text{km}$, $\Delta I_i = 200\,\text{A}$)

Fig. 11: Effect of efficiency attacks on the TPS with BDD under different settings of Δs_i and ΔV_i.

Table V: Time duration while the TPS experiences safety breaches under different settings of Δs_i in the presence of BDD.

Δs_i (km)	No BDD	0.5	0.4	0.3	0.2	0.1
Time duration with safety breaches (second)	8	4	1	0	0	0

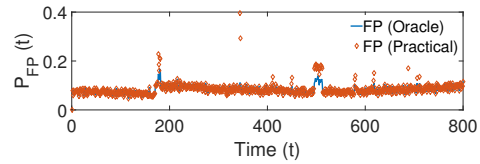
safety attacks, the regenerating trains inject more power into the power network than that under no attacks, resulting in increased voltages. We say that the TPS experiences a safety breach when at least one node in the TPS experiences a safety breach.

Table V summarizes the time durations of safety breaches under the two sets of simulations. We consider BDD with different settings of Δs_i. It can be observed that without BDD, the TPS experiences safety breaches for a total of eight seconds. The prolonged overvoltage may cause safety incidents. However, with BDD we see that, when Δs_i is in the range of $0.1\,\text{km}$ to $0.3\,\text{km}$, the attack causes no safety breaches during the simulation. As discussed previously, the setting $\Delta s_i = 0.1\,\text{km}$ is appropriate in practice. Hence, this set of results shows that by appropriately setting the BDD parameters, safety breaches can be nearly eliminated.

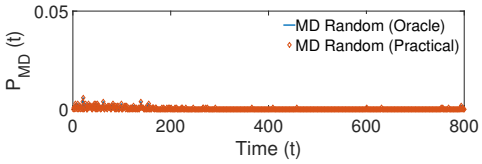
7.2.3. SAD Algorithm. The last set of simulations evaluates the effectiveness of SAD in detecting attacks that have bypassed the BDD. In this set of simulations, we set $\alpha = 1$ (since the scaling is not necessary in the absence of measurement noise). Furthermore, we use $p = 2$ in our evaluations.[4] For each time instant, among the discrete solutions to the BDD bypass condition discussed in Section 5.3, the attacker tactically chooses the one closest to the true system state in the sense of p-norm distance. We compare our *practical approach* where the $\mathbf{v}_{\text{pr}}$ is the nodal voltage vector at the previous time instant (cf. Algorithm 1), with an *oracle approach* where the $\mathbf{v}_{\text{pr}}$ is the nodal voltage vector at the present time instant in the absence of attack. For the oracle approach, we observe that the $||\tilde{\mathbf{v}} - \mathbf{v}_{\text{pr}}||_p$ is consistently higher than the J^* for the entire simulation. This suggests that the oracle approach can detect the onset of a BDD-stealthy attack launched at any time instant. For the practical approach, we observe that the $||\tilde{\mathbf{v}} - \mathbf{v}_{\text{pr}}||_p$ is higher than the J^* for 96% of the simulation time. For the remaining 4% of simulation time, the practical approach will miss the attack onset because of a significant change of $\mathbf{v}$ from the previous time instant to the present. This shows that the practical approach can detect the attack onset with a high detection probability.

We note that as the size of the TPS increases (in terms of the number of trains and substations under consideration), the number of constraints for the SAD algorithm as well as the solutions to the BDD-passing constraints will increase. Implementing

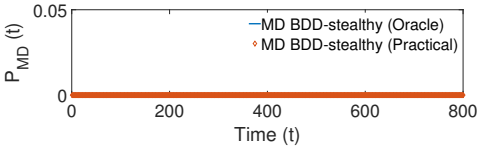
[4]Simulation results conducted with $p = 2$ and $p = \infty$ yielded similar performance of the SAD algorithm (in terms of the attack detection rate).



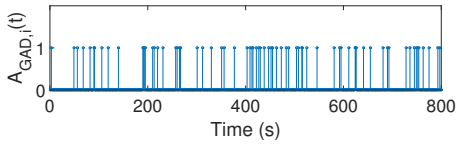
(a)FP rates of the GAD as a function of time.



(b)MD rates of the GAD as a function of time for random attacks.



(c)MD rates of the GAD as a function of time for BDD-stealthy attacks.



(d)GAD alarms over the simulation interval under no attacks.

Fig. 12: FPs and MDs of the GAD. $\tau = 16$ and $\alpha = 0.9$.

the SAD algorithm may become computationally complex. However, as we pointed out earlier, it is often sufficient to consider only a small section of TPS for security analysis. Thus, in practical application, the computational overhead of the SAD algorithm will be acceptable.

7.3. Simulation Results With Random Sensor Measurement Noises

In this subsection, we examine the empirical FP and MD rates of the GAD at different time instants of the 800 second simulation interval. To compute these quantities, we run N simulation runs. We let $Z_{\text{GAD},i}(t)$ and $\Xi_{\text{GAD},i}(t)$ denote the indicator variables representing FPs and MDs at a time instant $t \in \{1, \ldots, 800\}$ during the simulation run $i \in \{1, \ldots, N\}$. The empirical FP and MD rates at time t are then computed as $P_{\text{FP}}(t) = \frac{1}{N}\sum_{i=1}^{N} Z_{\text{GAD},i}(t)$, $P_{\text{MD}}(t) = \frac{1}{N}\sum_{i=1}^{N} \Xi_{\text{GAD},i}(t)$. In our simulations, we set $N = 1000$ and the noise level to 0.3% of the full-scale voltage and current sensor readings. The BDD detection threshold τ is set to 16, and $\alpha = 0.9$ for the SAD. The value of α was tuned numerically by observing observing the values of $||\tilde{\mathbf{v}} - \mathbf{v}_{\text{pr}}||_p$ and J^* in the scenario when the BDD-passing constraint has only two solutions. The chosen value of α is sufficient to eliminate MDs.

Fig. 12a shows the FP rate of the GAD, and Figs. 12b and 12c show the MD rates of the GAD for random and BDD-stealthy attacks. For random attacks, we inject an additive attack of 20 V to the voltage measurement of the leftmost train (in Fig. 9a). We make the following observations. First, we observe that the FP and MD rates fluctuate over time, since the TPS topology and parameters change. (Recall that the TPS topology and parameters depend on the position and the power drawn/injected by the trains.) Second, we observe that under the considered settings, both the oracle and practical GAD detectors yield very low MD rates at all time instants. Thus we conclude that by appropriately tuning the parameters of the BDD and SAD detectors (τ and α), the MD rate of the GAD can be reduced to a very low value. Third, we observe that while the FP rate is low for most of the simulation interval, there are a few time instants at which the FP rate is relatively high, particularly for the practical GAD detector (e.g., from $t = 497$ to $t = 511$, the FP rate ≈ 0.2). Furthermore, we observe that these time instants correspond to when one or more trains change their motion status

from tractioning to braking mode, thus resulting in a drastic change in the system state. Recall that an accelerating train draws power from the network resulting in a voltage drop whereas as a braking train injects power resulting in a voltage raise. In these cases, the difference $||\mathbf{v} - \mathbf{v}_{\text{pr}}||$ can be high for the practical GAD detector since $\mathbf{v}_{\text{pr}}$ is estimated only based on the historical values.

However, in practice, an extremely low FP rate is desired, since otherwise the system operator would have to frequently initiate unnecessary mitigation that may be disruptive. Thus, in what follows, we propose an adaptive version of the GAD, which we call GAD with attack detection window (GAD-W). GAD-W will give an extremely low FP rate in the presence of sensor measurement noises.

7.3.1. GAD with Attack Detection Window. The GAD-W detector applies an AND rule to fuse the detection results in an *attack detection window*, i.e., instead of declaring the presence of an attack based on a single alarm, GAD-W waits for consecutive alarms over several time slots before declaring it. In the following, we first formally state the GAD-W detector and then provide the intuition behind its design. Denote by $A_{\text{GAD}}(t) \in \{0, 1\}$ the detection result of the GAD at time t and by $W \in \mathbb{N}$ the window size. The GAD-W detector raises an alarm only if there is an alarm at all the time instants within the attack window, i.e.,

$$A_{\text{GAD-W}}(t) = A_{\text{GAD}}(t) \wedge A_{\text{GAD}}(t+1) \wedge \cdots \wedge A_{\text{GAD}}(t+W-1). \tag{21}$$

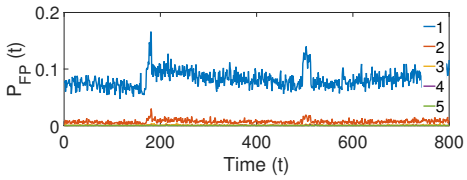
The rationale is that in the absence of attacks, the occurrence of GAD alarms can be due to two factors: (i) the fluctuations of BDD residual induced by the measurement noise, or (ii) a drastic change in the system state between consecutive time slots. In the above two cases, the BDD and SAD will raise an alarm, respectively. The first case is a randomly occurring event (due to noise) and the second is a sparsely occurring event. Thus, the probability of having consecutive GAD alarms over a time window is low. Fig. 12d confirms this hypothesis, in which we plot the GAD alarms for one instantiation of the 800 second simulation interval in the absence attacks. It can be seen that the occurrence of alarms is sparse. Thus, the AND fusion rule in an attack detection is effective.

A larger window size W can lower the probability of consecutive alarms within the detection window, resulting in a lower FP rate. However increasing the window size may lead to higher MD rates when an attack is present. Moreover, it also introduces longer delay in detecting the attacks. Thus, the setting of the optimal window size should balance between the FP and MD rates. In what follows, we present simulation results to show the variations of FP and MD rates for different window sizes, which will guide the setting of the window size.

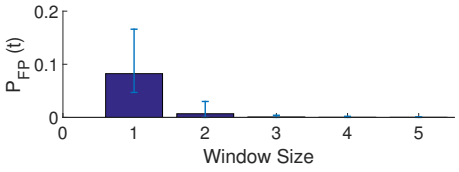
Fig. 13 and Fig. 14 show the FP and MD rates for GAD-W detector under both random and BDD-stealthy attacks. We observe that as the window size increases, the FP rate decreases, whereas the MD rate increases, for the random attacks. We observe that for a window size of 3, the average FP rate is 9×10^{-4}. The average MD rate for the random attack is 7×10^{-4}. Such extremely low of FP and MD rates are acceptable under practical scenarios. Finally, we observe that the MD rates for the BDD-stealthy attacks are very low both under the oracle and practical GAD detectors. This is because the SAD detector is specifically designed to detect BDD-stealthy attacks.

8. CONCLUSIONS

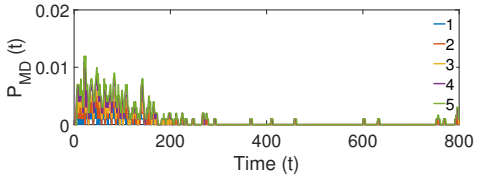
In this paper, we studied FDI attacks on train-borne sensor measurements used in railway TPSes. To the best of our knowledge, ours is the first effort that has studied TPSes from a cybersecurity perspective. To account for the safety-criticality of TPS, we adopted the Kerckhoffs's principle and addressed two fundamental problems of



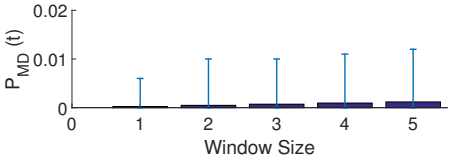
(a)FP rate over time for different window sizes.



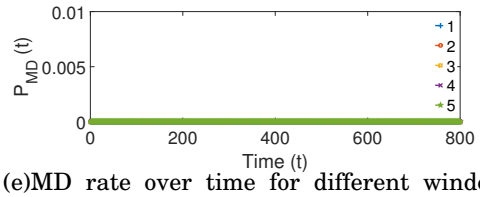
(b)FP rate under various attack detection window sizes. Error bars represent maximum and minimum values.



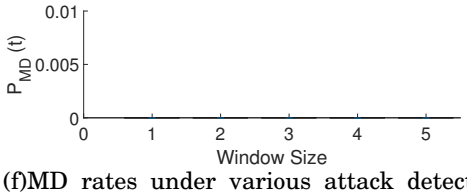
(c)MD rate over time for different window sizes under random attack.



(d)MD rate under various attack detection window sizes under random attack. Error bars represent maximum and minimum values.



(e)MD rate over time for different window sizes under BDD-stealthy attack.



(f)MD rates under various attack detection window sizes under BDD-stealthy attack. Error bars represent maximum and minimum values.

Fig. 13: FP and MD rates for oracle GAD detector under random attacks.

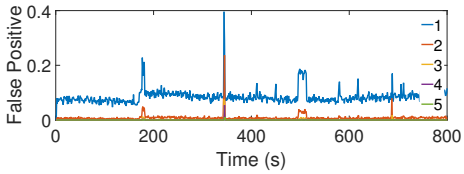
importance, namely, characterization of the impact of FDI attacks on TPSes, and development of detection techniques for these attacks. We formulated and analyzed the efficiency and safety attacks that aim to minimize the system energy efficiency and breach system safety conditions, respectively. To detect these attacks, we proposed a global detection system that serializes the proposed BDD and SAD algorithms, both of which may be implemented at a central TPS monitor. Furthermore, we proposed an adaptive GAD-W detector that achieves a very low FP rate in the presence of noisy sensor measurements. Our simulation results verified the susceptibility of the TPS setup to the FDI attacks, but these attacks can be detected effectively by the proposed global detection system.

REFERENCES

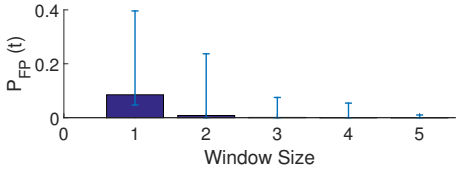
2015. Osiris & Urban Rail - Comprehensive Approach to Making the Save. *Mobility - The European Collective Transport Magazine* (2015). http://bit.ly/2pryv7E.

2016. ELECTRIC TRACTION POWER SUPPLIES. *Railway Technical Web Pages* (2016). http://www.railway-technical.com/etracp.shtml.

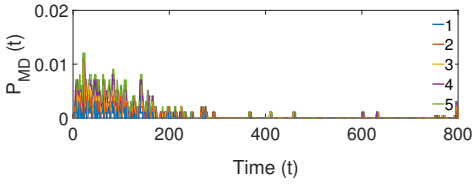
L. Abrahamsson. 2012. *Optimal Railroad Power Supply System Operation and Design*. PhD Thesis, KTH Sweden.



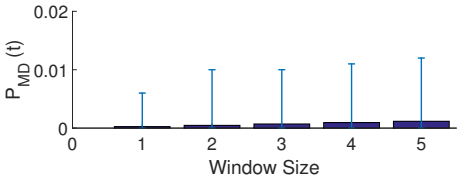
(a)FP rate over time for different window sizes.



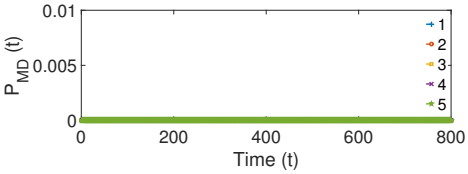
(b)FP rate under various attack detection window sizes. Error bars represent maximum and minimum values.



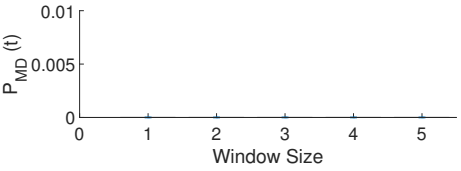
(c)MD rate over time for different window sizes under random attack.



(d)MD rate under various attack detection window sizes under random attack. Error bars represent maximum and minimum values.



(e)MD rate over time for different window sizes under BDD-stealthy attack.



(f)MD rates under various attack detection window sizes under BDD-stealthy attacks. Error bars represent maximum and minimum values.

Fig. 14: FP and MD rates for practical GAD detector with BDD-stealthy attacks.

S. Acikbas and M.T. Soylemez. 2007. Parameters affecting braking energy recuperation rate in DC rail transit. In *ASME/IEEE Joint Rail Conf. & Internal Combustion Engine Division Spring Technical Conf.*

Alstom. 2001. ERTMS/ETCS On-Board ALSTOM Solution. (2001). https://bit.ly/1OOb38f.

S. Amin, X. Litrico, S. Sastry, and A.M. Bayen. 2013. Cyber security of water scada systemspart I: Analysis and experimentation of stealthy deception attacks. *IEEE Trans. Control Syst. Technol.* 21, 5 (Sept. 2013), 1963–1970.

P. Arboleya, G. Diaz, and M. Coto. 2012. Unified AC/DC Power Flow for Traction Systems: A New Concept. *IEEE Trans. Veh. Technol* 61, 6 (July 2012), 2421–2430.

P. Arboleya, B. Mohamed, C. Gonzlez-Morn, and I. El-Sayed. 2016. BFS Algorithm for Voltage-Constrained Meshed DC Traction Networks With Nonsmooth Voltage-Dependent Loads and Generators. *IEEE Trans. Power Syst.* 31 (2016), 1526–1536.

Y. Cai, M.R. Irving, and S.H. Case. 1995. Iterative techniques for the solution of complex DC-rail-traction systems including regenerative braking. *IEE Proc. Generation, Transmission and Distribution* 142, 5 (1995).

A.A. Cárdenas, S. Amin, Z. Lin, Y. Huang, C. Huang, and S. Sastry. 2011. Attacks against process control systems: Risk assessment, detection, and response. In *Proc. ACM AsiaCCS*.

K.G. David. 2015. The train that powers its station. (2015). http://bbc.in/1KRROZK.

S.S.S.R. Depuru, L. Wang, and V. Devabhaktuni. 2011. Electricity theft: Overview, issues, prevention and a smart meter based approach to control theft. *Energy Policy* 39, 2 (Feb. 2011), 1007–1015.

R.G. Fletcher. 1991. Regenerative equipment for railway rolling stock. *Power Engineering Journal* 5, 3 (May 1991), 105–114.

T.F. Gabrielle. 2014. Deadly Derailment in Moscow Metro. (2014). http://bit.ly/2d5D7dy.

A. Gonzlez-Gil, R. Palacin, P. Batty, and J.P. Powell. 2014. A systems approach to reduce urban rail energy consumption. *Energy Conversion and Management* 80 (2014), 509 – 524.

K. Hartwig, M. Grimm, M. Meyer zu Hörste, and K. Lemmer. 2006. Requirements for safety relevant positioning applications in rail traffic - A demonstrator for a train borne navigation platform called "DemoOrt". (2006). http://elib.dlr.de/21252/1/wcrr.pdf.

K. Jinsub and T. Lang. 2013. On Topology Attack of a Smart Grid: Undetectable Attacks and Countermeasures. *IEEE J. Sel. Areas Commun.* 31, 7 (July 2013), 1294–1305.

S. Karnouskos. 2011. Stuxnet worm impact on industrial cyber-physical system security. In *Conf. IEEE Industrial Electronics Society*.

R. Karri, J. Rajendran, K. Rosenfeld, and M. Tehranipoor. 2010. Trustworthy Hardware: Identifying and Classifying Hardware Trojans. *Computer* 43, 10 (Oct 2010), 39–46.

D.F. Kune, J. Backes, S.S. Clark, D. Kramer, M. Reynolds, K. Fu, Y. Kim, and W. Xu. 2013. Ghost talk: Mitigating EMI signal injection attacks against analog sensors. In *IEEE Symp. Security and Privacy*.

Y. Liu, P. Ning, and M.K. Reiter. 2009. False Data Injection Attacks Against State Estimation in Electric Power Grids. In *ACM CCS*.

P. McDaniel and S. McLaughlin. 2009. Security and Privacy Challenges in the Smart Grid. *IEEE Security Privacy* 7, 3 (2009), 75–77.

D. Mike. 2009. Recoverable Advanced Metering Infrastructure. In *Proc. Black Hat Technical Security Conference*.

M. Miyatake and H. Ko. 2010. Optimization of Train Speed Profile for Minimum Energy Consumption. *IEEJ Transactions on Electrical and Electronic Engineering* 5, 3 (2010), 263–269.

Y. Okada, T. Koseki, and K. Hisatomi. 2004. Power management control in DC-electrified railways for the regenerative braking systems of electric trains. *Advances in Transport* 15 (2004), 919–929.

C.L. Pires, S.I. Nabeta, and J.R. Cardoso. 2007. ICCG method applied to solve DC traction load flow including earthing models. *IET Electric Power Applications* 1, 2 (March 2007), 193–198.

A.U. Raghunathan, T. Wada, K. Ueda, and S. Takahashi. 2014. Minimizing Energy Consumption in Railways by Voltage Control on Substations. In *Proc. Intl. Conf. Railway Engineering Design and Optimization*.

M.A. Rahman, E. Al-Shaer, and R.G. Kavasseri. 2014. A Formal Model for Verifying the Impact of Stealthy Attacks on Optimal Power Flow in Power Grids. In *Proc. ACM/IEEE ICCPS*.

Homeland Security. 2011. U.S. DHS. Insider threat to utilities. (2011). https://bit.ly/1YPFoZH.

S. Shuai, T. Tao, and C. Roberts. 2015. A Cooperative Train Control Model for Energy Saving. *IEEE Trans. Intell. Transp. Syst.* 16, 2 (April 2015), 622–631.

S. Shuai, T. Tao, L. Xiang, and G. Ziyou. 2014. Optimization of Multitrain Operations in a Subway System. *IEEE Trans. Intell. Transp. Syst.* 15, 2 (April 2014), 673–684.

Mors Smitt. 2016. Traction energy measuring solutions. (2016). http://bit.ly/2q5OUuZ.

SMRT. 2015. Press Release. (July 2015). https://bit.ly/1RxGBSk.

F. Sottile. 2011. *Real solutions to equations from geometry*. Vol. 57. American Mathematical Society Providence, RI.

Symantec. 2014. Dragonfly: Cyberespionage Attacks Against Energy Suppliers. (2014). http://symc.ly/2cowemc.

S.N. Talukdar and R.L. Koo. 1977. The analysis of electrified ground transportation networks. *IEEE Trans. Power App. Syst.* 96, 1 (1977).

A. Teixeira, H. Sandberg, G. Dan, and K.H. Johansson. 2012. Optimal power flow: Closing the loop over corrupted data. In *Proc. ACC*.

The Economic Times – Railways. 2012. Indian Railways to launch real-time train tracking via Google maps. (2012). https://bit.ly/1OIcMOe.

Transport for London. 2008. LU Carbon footprint report 2008. (2008). http://bit.ly/2pgb8xb.

A.J. Wood and B.F. Wollenberg. 1996. *Power Generation, Operation, and Control*. A Wiley-Interscience.

Anil Yadav. 2013. Traction choices: Overhead ac vs third rail dc. (2013). http://bit.ly/2orprPW.

Y. Yanling, L. Zuyi, and R. Kui. 2011. Modeling Load Redistribution Attacks in Power Systems. *IEEE Trans. Smart Grid* 2, 2 (2011).

Appendix A: BDD Threshold

In this Appendix, we present how to set the BDD threshold τ to ensure that the false positive rate is maintained at a certain level.

Recall that the expression for BDD residual is given by $r = ||\mathbf{z} - \mathbf{H}\hat{\mathbf{v}}||$, where $\mathbf{z} = \mathbf{H}\mathbf{v} + \mathbf{n}$, $\hat{\mathbf{v}} = (\mathbf{H}^T\mathbf{\Sigma}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{\Sigma}\mathbf{z}$. Substituting the expression of $\hat{\mathbf{v}}$, we obtain:

$$\begin{aligned} r &= ||\mathbf{z} - \mathbf{H}(\mathbf{H}^T\mathbf{\Sigma}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{\Sigma}\mathbf{z}|| \\ &= ||\mathbf{H}\mathbf{v} + \mathbf{n} - \mathbf{H}(\mathbf{H}^T\mathbf{\Sigma}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{\Sigma}(\mathbf{H}\mathbf{v} + \mathbf{n})|| \\ &= ||(\mathbf{I} - \mathbf{\Gamma})\mathbf{n}||, \end{aligned} \tag{22}$$

where $\mathbf{\Gamma} = \mathbf{H}(\mathbf{H}^T\mathbf{W}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{\Sigma}$. From (22), r follows a chi-square distribution, since the noise $\mathbf{n}$ is Gaussian. To maintain a certain FP rate α, the BDD threshold can be set by solving $\mathbb{P}(r \geq \tau) = \alpha$.