Secure Data Transmission and Trustworthiness Judgement Approaches Against Cyber-Physical Attacks in an Integrated Data-Driven Framework

Yuchen Jiang, *Member, IEEE*, Shimeng Wu, Hongyan Yang, *Member, IEEE*, Hao Luo, *Senior Member, IEEE*, Zhiwen Chen, *Member, IEEE*, Shen Yin, *Senior Member, IEEE*, and Okyay Kaynak, *Life Fellow, IEEE*

***Abstract*—Threats of cyberattacks have penetrated from disclosing critical user information to destroying/manipulating industrial control systems. Study on data security during network transmission has raised increasing attention in the systems and control community, which is found very necessary and timely in the context of Industry 4.0. In most existing approaches, the protection of the transmitted data from eavesdropping attacks and the detection of malicious integrity attacks are usually carried out separately. In this study, an integrated data-driven framework applicable at the control level is proposed to deal with secure transmission and attack detection simultaneously. In the framework, a secure correlation-based encryption/decryption approach and a trustworthiness judgement approach are proposed. Comprehensive discussions are made regarding the analysis of the sensitivity to attacks, the introduced time delay, and the design degree-of-free. Executable algorithms are presented, corresponding to which hardware is modularized and can work standalone independent from the configuration of the monitoring and control systems or any third-party authentication agencies. Evaluation results on a simulated two-area frequency-load control power grid system are provided to show the effectiveness and performance of the proposed approaches.**

***Index Terms*—Attack detection, cyber-physical system (CPS), data-driven framework, industrial security.**

I. Introduction

A. Background and Motivation

IN THE MODERN industry, the scales of the plants and the processes have been enlarged. There can be hundreds to thousands of sensors installed in a large-scale industrial (e.g., chemical or biomedical reaction) system [1]. They are used to collect necessary real-time information about the states of the systems and the environmental condition. Due to large-scale and distributed deployment, the sensors are geographically dispersed and are sometimes remote from the monitoring and control center [2]. In such a context, network transmission is necessary. Since the sensor networks are a part of field production systems and are supposed to be accessible only by the internal field control and management sectors. Traditionally, they are physically isolated from external networks. However, in the context of Industry 4.0, deep integration of the information and communication technologies (ICTs) and the control engineering establishes novel connectivity between the field sensors to the external networks [3]–[6]. Without a dedicated security design, the data transmitted in the sensor networks are exposed to malicious external cyberattacks [7]–[9].

In the post-Snowden age, it should be understood that the confidentiality of the industrial process data is critical to not only the business secrets but also to the safe and reliable operation of the closed-loop control of the field processes and devices. It can never be an exaggeration to emphasize the fact that different from IT attacks, malicious cyber-physical attacks can destroy or manipulate the closed-loop control systems and cause physical damages directly. In fact, there have been several shocking cyber-physical system (CPS) attacks to the real-world industrial systems reported, many of which are the critical infrastructure, ranging from the power grids to the nuclear plants to the wastewater treatment systems [10]. For instance, the attackers managed to break into the internal network and injected the Stuxnet virus into the supervisory control and data acquisition (SCADA) system, which made the centrifuge at the Uranium enrichment base in Iran run out of control. This makes the Stuxnet virus the first malicious code that directly destroys industrial infrastructure in the real world. Another cyber-physical attack took place in December 2015. The Ukrainian power network was attacked

Manuscript received 15 January 2022; accepted 26 March 2022. Date of publication 14 April 2022; date of current version 18 November 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 62073104, and in part by the Young Scientist Studio of Harbin Institute of Technology. This article was recommended by Associate Editor X. Ge. *(Corresponding authors: Yuchen Jiang; Shen Yin.)*

Yuchen Jiang is with the Department of Control Science and Engineering, Harbin Institute of Technology, Harbin 150001, China, and also with the Industrial Intelligence Academician Studio, Peng Cheng Laboratory, Shenzhen 518066, China (e-mail: yc.jiang@hit.edu.cn).

Shimeng Wu and Hao Luo are with the Department of Control Science and Engineering, Harbin Institute of Technology, Harbin 150001, China.

Hongyan Yang is with the Department of Control Science and Engineering, Beijing University of Technology, Beijing 100124, China.

Zhiwen Chen is with the Department of Automation, Central South University, Changsha 410083, China, and also with the Industrial Intelligence Academician Studio, Peng Cheng Laboratory, Shenzhen 518066, China.

Shen Yin is with the Department of Mechanical and Industrial Engineering, Faculty of Engineering, Norwegian University of Science and Technology, 7491 Trondheim, Norway (e-mail: shen.yin@ntnu.no).

Okyay Kaynak is with the Department of Electrical and Electronics Engineering, Bogazici University, 34342 Istanbul, Turkey, and also with the Turkish Academy of Sciences, 06670 Ankara, Turkey.

This article has supplementary material provided by the authors and color versions of one or more figures available at https://doi.org/10.1109/TSMC.2022.3164024.

Digital Object Identifier 10.1109/TSMC.2022.3164024

TABLE I
DIFFERENT TYPES OF CYBER-PHYSICAL ATTACKS

Type of attack	System model & configuration	I/O data collection	I/O data manipulation
Replay attack	N	Y	Y
DoS^a attack	N	N	Y
FDIA^b	P	N	Y
Zero dynamics	Y	N	Y
Convert attack	Y	Y	Y
Eavesdropping	N	Y	N

^{a}DoS: Denial of service ^{b}FDIA: False data injection attack

Y: yes N: no P: partially

TABLE II
STATUS OF RESEARCH ON CPS SECURITY/DEFENSE SCHEMES

	Explicit threat model	Sensitivity to attacks	Time-delay & overhead	Executable algorithm
[28]	N	N	N	P
[20]	Y	N	N	P
[25]	N	N	N	Y
[29]	N	N	N	Y
[30]	N	Y	Y	P
[31]	N	Y	Y	Y
[5]	Y	N	Y	Y
[27]	Y	N	Y	Y
[32]	Y	Y	N	Y

Y: yes N: no P: partially

by BlackEnergy. It caused the power generation equipment to malfunction. Hundreds of thousands of households in the region suffered a blackout.

As long as network transmission is involved, there are potential risks of data breaches. According to the influence on the plants, attacks can be categorized into the eavesdropping attack and the integrity attack [11]. Eavesdropping attack only steals information, whereas an integrity attack causes abnormal changes in the transmitted data. By careful design, the changes can be made undetectable by bad data detection systems or system monitoring systems [12]. Different types of attacks are summarized in Table I, according to how much system knowledge is used, how many channels are eavesdropped on, and how many process and control variables the attackers manipulate. Secure data transmission schemes and trustworthiness judgement schemes are urgently required to ensure system safety both before and after attacks [13], [14]. In the literature, these tasks are also referred to as attack defending and attack detection [11].

B. Related Work

With the rising awareness of industrial cyber-physical security, researchers have proposed a series of solutions based on quite different theoretical foundations and assumptions [15], [16]. The existing approaches that are designed from the systems and control perspective and are applicable to the application level can be generally categorized as the following.

Regarding the defending and anti-eavesdropping strategies, cryptography may be the most popular one with thousands of years of history. The development of modern electrical computer technology opens a new chapter for classic cryptography. The boosting computing power enforces the keyspace to be extremely large to defend against the brute-force attacks. Different from cryptography, randomization-based approaches dedicate to confuse the attackers when the predictability of the deterministic rules has been leveraged [11]. Regarding the detection and trustworthiness judgement strategies, the most effective ones are inherited from the study of signal processing, statistical analysis, and machine learning [17], [18]. According to different detection objectives, watermarking-based approaches inspect the specific characteristics of the known auxiliary signals, i.e., the so-called watermark signal [19]; signature-based approaches look for specific characteristics of the known-mechanism attacker signals [11]; anomaly detection-based approaches examine the preselected features of the transmitted data or the difference between the practical system and the nominal system [20], [21].

In terms of algorithmic approaches, Wang *et al.* [22] studied the model-based framework and proposed to introduce a watermarking-based encryption mode to defense against linear deception attacks.

Ge *et al.* [23] designed finite horizon distributed attack detection estimators in the Krein space to enable distributed attack detection for time-varying systems.

For secure communication, He *et al.* [24] proposed an event-triggered strategy to achieve synchronization of master–slave neural networks with bounded synchronization errors.

Tao *et al.* [25] proposed for the discrete-event industrial CPSs a trustworthy and secured data collection scheme before and after transmission. Beg *et al.* [20] dealt with dc microgrids and proposed to identify changes in a set of inferred candidate invariants, where extensive preliminary modeling workload is required.

Ding *et al.* [26] provided a comprehensive survey of the state-of-the-art approaches for secure state estimation and control of CPSs.

In terms of implementation, Pearce *et al.* [5] developed dedicated hardware that isolates the programmable logic controllers and the physical machinery. It was implemented in an overcurrent relay as a runtime enforcer, which is based on bidirectional safety policies. Khan and Tomic [27] developed a runtime security monitor that can detect abnormalities across the application layer and communication layer, which was tested on a small-scale water distribution system.

In terms of research methodology, Table II lists a few examples of the comprehensiveness of the existing research. The evaluation indices in the top row are inspired by [6], whereas the references are updated with more recent ones. Although the tiny survey has some limitations, it reveals that there are weak points in the analysis of the sensitivity/robustness to attacks and the performance impact on the control systems caused by the security countermeasures. In this article, these aspects are addressed and discussed in details.

C. Contribution of This Work

In industrial CPSs, an inherited defect lies in that the priorities of design are safety and performance, rather than security. As long as the industrial enterprises focus only on the application-oriented functionalities (such as product quality) and disregard the cybersecurity design, the severe degree of the potential threats to life and property safety and the environmental impact will always be underestimated [6]. Based on the above facts and observations, this work is dedicated to providing simple but reliable countermeasures to the secure transmission and trustworthiness judgement tasks regarding the sensor network data. The contributions of this article beyond the existing work lie in the following aspects.

1) An integrated security design framework against eavesdropping attacks and integrity attacks is proposed, which is studied from a systems and control perspective.
2) In the integrated framework, a data-driven encryption and decryption approach against eavesdropping attacks with a high design degree of freedom (DoF) is proposed. A trustworthiness judgement scheme is proposed, which is sensitive to integrity attacks.
3) For easy deployment to the existing sensor networks, complete algorithm steps with modularized realization are introduced.

II. Problem Statement

In this study, it is assumed that the attackers have broken into the network firewall and can gain illegitimate access to the data transmission channels. On this basis, they eavesdrop and record the measurement data transmitted via the physical networks (regardless of wired or wireless transmission). They modify data transmitted in the network to launch deception attack where the compromised data are carefully designed to avoid triggering the traditional bad data detection systems and fault detection systems [33].

With the above assumptions, this work will address the following scientific questions.

1) How to encrypt the measurement data to be transmitted via the networks so that they are not easily eavesdropped on and cracked even the malicious party knows the encryption mechanism?
2) How to distinguish whether or not the received data are compromised due to external attacks and whether they are trustworthy for subsequent use, e.g., for control and monitoring purposes?
3) How to accomplish the above two tasks simultaneously in an integrated data-driven framework where only the measurement data are necessary for design without need of any priori system knowledge?

The expectations of potential solutions include: no need for third-party involvement; difficult to crack using brute force methods; and lightweight design and implementation, i.e., can be realized by an embedded system or dedicatedly designed hardware. The data transmission process does not need any additional infrastructures beyond the existing ones. From the algorithmic perspective, the offline design and online implementation solutions introduce little time delay and do not need large-scale training.

Bearing these in mind, the proposed framework and solutions are introduced in the next section.

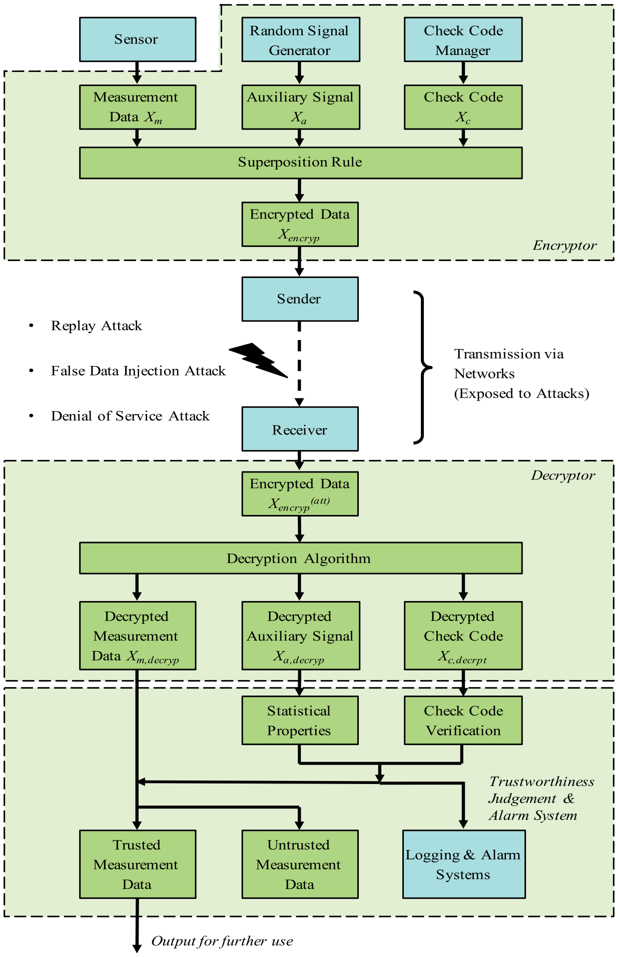


Fig. 1. Integrated framework against eavesdropping and integrity attacks.

III. Main Contribution

A. Integrated Framework

In this part, a data-driven framework is proposed for secure data transmission against eavesdropping and integrity attacks. As shown in Fig. 1, the framework is composed of an encryptor module, a decryptor module, and a trustworthiness judgement module. The sensor measurements and the check codes are superposed to the auxiliary signals (the carrier) to obtain the encrypted data before transmission. At the receiver end, decryption is performed to recover the sensor measurements. Two trustworthiness indicators are used to determine whether the receiver data are compromised or not. By using the terminology "data driven," it is intended to refer to using the measurement data X_m and the auxiliary data X_a to design the encryption/decryption scheme, without relying on any system knowledge or model of the transmitted signals. In another aspect, at the implementation stage, the online measurement data and the auxiliary data are used to drive the encryption process while the received encrypted data are used to drive the decryption process and to judge whether an attack takes place during transmission. This work is dedicated to the security design at the monitoring and control level rather than the communication and network level. In other words, the attackers have access to the transmission channels to collect and change whatever data during the transmission process, but not at the sender end or the receiver end.

There are three key design elements in the framework. The first design element is the auxiliary signal (act as the carrier). It must be generated with some degree of randomization to prevent the attackers from cracking the encrypted data. It shall also have certain characteristics that can be used to decouple from the encrypted data. The second design element is the check code. Although the terminology is borrowed from the communication techniques, the concept is generalized for trustworthiness judgement at the monitoring and control level. The check code in the proposed framework is defined as a confidential time series, which is preknown to both the sender and the receiver. Under this definition, there could be many realization approaches. Moreover, it should be noted that along with the proposed data-driven encryption/decryption approach, the check code does not have to be a function of the transmitted measurement variables. The third design element is the trustworthiness indicators. As will be shown in Section III-D, the indicators should be detectable in case of sparse attacks (attacks to one or a few channels rather than to all the channels) and be sensitive to the attacks.

In the following Sections III-B–III-D, the three key elements will be, respectively, discussed in more detail. Then, a complete implementation procedure is summarized in Section III-E.

B. Data-Driven Encryption and Decryption Approaches

In the proposed framework, the key to security design against the eavesdropping attack lies in the design of the auxiliary signals. The auxiliary signals are used as masks of the measurement data to achieve encrypted transmission. Assume that the eavesdroppers manage to break into the communication-level/network-level security defenses, at the control level (which is the focus of this work), they shall be unable to recover the measurements from the encrypted data. In below, a lightweight data-driven scheme is proposed, whose basic idea can be summarized as follows.

1) *Encryption:* Introduce two sets of auxiliary variables X_a and Y_a that are fully correlated so that there is no independent subspace. The sensor signals can be arbitrarily superposed to the auxiliary variable channels under a certain superposition rule. Without loss of generality, the first m channels of X_a are used to carry the sensor measurements X_m.
2) *Decryption:* Use the following proposed modified total principal component regression (MTPCR) model to analyze X^{enc} and Y_a. To extract X_m at the receiver end, a projection matrix M_o is designed during the offline phase as a decryptor.

TABLE III
SPACE DECOMPOSITION BASED ON MTPCR

Subspace	Interpretation
$\mathcal{S}_{\hat{\hat{\Phi}}}$	Information subspace of Φ that is fully correlated to $\hat{Y}$, and can be used to predict Y.
$\mathcal{S}_{\tilde{\hat{\Phi}}}$	Information subspace of Φ that is uncorrelated to $\hat{Y}$, and is thus useless to the prediction of Y.
$\mathcal{S}_{\tilde{\Phi}}$	Stochastic noise subspace.
$\mathcal{S}_{\hat{\hat{Y}}}$	Predictable subspace that contains most information.
$\mathcal{S}_{\tilde{\hat{Y}}}$	Predictable subspace that contains measurement noise.
$\mathcal{S}_{\tilde{Y}}$	Unpredictable subspace by Φ.

Consider the partial least squares (PLSs) decomposition

$$\begin{cases} \Phi = TP^T + \tilde{\Phi} \\ Y = TQ^T + \tilde{Y} \end{cases} \tag{1}$$

where $P \in \mathbb{R}^{m \times nlv}$ and $Q \in \mathbb{R}^{l \times nlv}$ are the loading matrices. (nlv denote the number of latent variables. m and l denote the input and the output dimension, respectively.) $T \in \mathbb{R}^{N \times nlv}$ is the score matrix. $\tilde{\Phi} \in \mathbb{R}^{N \times m}$ and $\tilde{Y} \in \mathbb{R}^{N \times l}$ are the residual matrices. Denote $\hat{\Phi} = TP^T \in \mathbb{R}^{N \times m}$ and $\hat{Y} = TQ^T \in \mathbb{R}^{N \times l}$. The above model can be obtained by solving the following optimization problem:

$$[P^*, Q^*] = \arg_{P,Q} \max \ \left\| P^T \Phi^T Y Q \right\| \tag{2}$$

$$\text{s.t.} \ <P_i, P_i> = 1, \ <Q_k, Q_k> = 1$$

$$<P_i, P_j> = 0 \tag{3}$$

where $<\cdot, \cdot>$ denotes the inner product. P_i and Q_k denote the ith and kth column of P and Q, respectively. Some commonly used numerical solution to the problem can be found in [34].

It should be noted that in the PLS model, $\mathcal{S}_{\hat{Y}}$ is corrupted by the residual term. For decryption, it is necessary to further distinguish the information subspace and the noise subspace of $\hat{Y}$. Consider the MTPCR model

$$\begin{cases} \Phi = T_y P_y^T + \tilde{\hat{\Phi}} + \tilde{\Phi} \\ Y = T_y Q_y^T + \tilde{\hat{Y}} + \tilde{Y} \end{cases} \tag{4}$$

where $P_y \in \mathbb{R}^{m \times npc}$ and $Q_y \in \mathbb{R}^{l \times npc}$ are the loading matrices. (npc denotes the number of principal components, the value of which can be adjusted at the design stage.) $T_y \in \mathbb{R}^{N \times npc}$ is the score matrix. Denote $T_y P_y^T$ by $\hat{\hat{\Phi}}$, and $T_y Q_y^T$ by $\hat{\hat{Y}}$. The target is to derive the reconstructible quality information from the process variables. To this end, some extra procedures are necessary: 1) based on the standard PLS decomposition, perform principal component analysis (PCA) on $\hat{Y}$ to obtain the useful information subspace spanned by the variables in $\hat{\hat{Y}}$ and the measurement noise subspace spanned by the variables in $\tilde{\hat{Y}}$ and 2) perform least-square decomposition between $\hat{\hat{Y}}$ and $\hat{\Phi}$ to obtain $\hat{\hat{\Phi}}$ and $\tilde{\hat{\Phi}}$. The interpretation of the subspaces of the MTPCR model is summarized in Table III, where $\mathcal{S}_{\Xi}$ denotes the space spanned by the column vectors of Ξ (Ξ can be $\hat{\hat{\Phi}}$, $\tilde{\hat{\Phi}}$, $\tilde{\Phi}$, $\hat{\hat{Y}}$, $\tilde{\hat{Y}}$, and $\tilde{Y}$).

Given the MTPCR model, it is straightforward to assign $X_m \rightarrow \tilde{\hat{\Phi}}$, $X_a \rightarrow \hat{\hat{\Phi}}$, and $Y_a \rightarrow \hat{\hat{Y}}$ for the encryption and decryption purposes. The detailed procedures are summarized in Table IV. It is a data-driven solution to the decryption matrix, which requires X^{enc} and Y_a.

Remark 1: Since Y_a is only used during the design phase and it is not transmitted during online application, there is no way for the attackers to solve for the decryption matrix based only on X^{enc}.

TABLE IV
OFFLINE DESIGN—DECRYPTOR DESIGN

Step 1. Training data generation
1) Collect sensor measurement data X_m.
2) Generate random numbers $sigGen$.
3) Set add-correlation matrix M_{AC} and correlating output matrix M_{GC}.
4) Calculate auxiliary masking data and correlating output data by
$X_a = sigGen \cdot M_{AC}$, $Y_a = X_a \cdot M_{GC}$.
5) Set check code X_c (See the next section).
6) Initialize $X^{enc} = X_a$ and superpose the sensor measurements and the check code to different channels of the masking data, according to a certain superposition rule $\mathcal{R}_s$, to obtain the encrypted data X^{enc} for transmission.
Step 2. Decryption matrix design
1) Collect X_{enc} and Y_a.
2) Do PCA to X_{enc} and obtain the score matrix T and the loading matrix P.
3) Calculate the projection matrix from T to $\hat{Y}_a$:
$Q = ((T^TT)^{-1}T^TY_a)^T$.
4) Calculate $\hat{Y}_a = TQ^T$.
5) Do PCA to $\hat{Y}_a$ and obtain the score matrix T_y and the loading matrix Q_y.
6) Calculate the projection matrix from T_y to $\hat{\hat{\Phi}}$:
$P_y = ((T_y^TT_y)^{-1}T_y^T(TP^T))^T$.
7) Calculate decryption matrix $M_o = I - P \cdot Q \cdot Q_yP_y^T$
where I is the unit matrix with an appropriate dimension.

Remark 2: The major differences between MTPCR and TPCR in [35] and [36] lie in that: 1) the target is different. MTPCR is dedicated to the reconstruction problem, whereas TPCR was designed to be variation sensitive for the fault diagnosis applications and as a result, 2) in the MTPCR model, the last regression is between $\hat{\hat{\Phi}}$ and $\hat{\hat{Y}}$ for the reconstruction purpose, and X is decomposed into three subspaces in total. In contrast, in TPCR, the regression is performed between X and $\hat{\hat{Y}}$ for fault-related subspace decomposition purpose, and X is decomposed into two subspaces in total.

Remark 3: There are other types of auxiliary signals in the existing literature. In parallel to the proposed correlation-based auxiliary signal, Wang *et al.* [22] proposed to add a Gaussian random variable with known covariance as the auxiliary signal, which was also referred to as the watermark signal. Porter *et al.* [19] proposed to used a time-varying dynamic system to generate the auxiliary signals. However, it is worth noting that whether the auxiliary signal is suitable is also related to the attack detection scheme.

C. Check Code Design

As discussed in Section III-A, there can be many realization forms of the check code. A suitable solution should ensure high-level security by following Kerckhoffs's principle: *A cryptosystem should be secure even if everything about the system, except the key, is public knowledge.* In this work, the possibility of employing the chaotic system-based approaches is explored.

Chaotic systems are deterministic systems whose state trajectories are extremely sensitive to the initial condition. A tiny error in the parameter will lead to notable differences along the process of evolving. Such characteristics and power of unpredictable make it more favorable compared with a general system (nonchaotic system), which can be reconstructed using system identification techniques by the attackers. In other words, one cannot identify the mathematics that describes the system based only on the output data (rather than advanced chosen-ciphertext attacks) and then predict how the system will unfold. Because the very simple rules naturally give rise to very complex objects, it is suitable for large-scale yet lightweight implementation at the sender's ends and the receiver's ends.

Recall that in the proposed framework, the check code is designed not to carry measurement signals (which is different from [24]); thus, there is no input to the chosen chaotic system. More specifically, the measurement data are not input to the master system in the case of master–slave synchronization. Once the chaotic system is configured, the check code is fully dependent on the initial states and the timestamps. From the perspective of deployment to large-scale sensor networks, the complexity and workload of configuration can be much reduced, by choosing the same chaotic sequence generator with unique initial conditions known only to the corresponding communication ends.

D. Detectability and Sensitivity Analysis

Let $X_a = [X_{a,1}^{N\times m} \quad X_{a,2}^{N\times c} \quad X_{a,3}^{N\times v}] \in \mathbb{R}^{N\times a}$, $X_m \in \mathbb{R}^{N\times m}$, and $X_c \in \mathbb{R}^{N\times c}$ denote the matrices of the auxiliary signals, the measurement signals, and the check code, respectively. Consider a superposition rule $R_{s,1}$ that leads to the following relationship (referred to as single transmission):

$$\begin{aligned} X^{\text{enc}} &= X_a + \left[X_m \quad \mathbf{0}^{N\times(c+v)}\right] + \left[\mathbf{0}^{N\times m} \quad X_c \quad \mathbf{0}^{N\times v}\right] && (5)\\ &= \big[\underbrace{X_{a,1} + X_m}_{\mathcal{S}_1} \; \underbrace{X_{a,2} + X_c}_{\mathcal{S}_2} \; \underbrace{X_{a,3}}_{\mathcal{S}_3}\big]. && (6)\end{aligned}$$

The decrypted check code, measurements, and auxiliary signals are calculated by

$$X_c^{\text{dec}} = \left(X^{\text{enc}} \cdot M_o\right)_{ccl} \tag{7}$$

$$X_m^{\text{dec}} = \left(X^{\text{enc}} \cdot M_o\right)_{mcl} \tag{8}$$

$$X_a^{\text{dec}} = X^{\text{enc}} \cdot M_y \tag{9}$$

where the subscripts ccl and mcl denote the operators that extract the corresponding columns/channels of the check codes and measurements from the matrix. In the attack-free scenario, the T^2 statistics of the measurements and the auxiliary signals are calculated by

$$J_{T^2,m}^0 = x_m^{0\,T}\left(\frac{X_m^TX_m}{n-1}\right)^{-1} \cdot x_m^0 = x_m^{0\,T}\Sigma_m^{-1}x_m^0 \tag{10}$$

$$J_{T^2,a}^0 = x_a^{0\,T}\left(\frac{X_a^TX_a}{n-1}\right)^{-1} \cdot x_a^0 = x_a^{0\,T}\Sigma_a^{-1}x_a^0. \tag{11}$$

Consider an attack $\alpha = [\alpha_1^{1\times m} \quad \alpha_2^{1\times c} \quad \alpha_3^{1\times v}] \in \mathbb{R}^{1\times a}$ to the encrypted data during transmission. Then, the decrypted check signal and the test statistics become

$$x_c^\alpha = \left((x^{\text{enc}} + \alpha)M_o\right)_{ccl} \tag{12}$$

$$\begin{aligned} J^{\alpha}_{T^2,m} &= \left((x^{\text{enc}}+\alpha)M_o\right)^T_{mcl}\Sigma_m^{-1}\left((x^{\text{enc}}+\alpha)M_o\right)_{mcl} \\ &= \left(x_m^0+(\alpha M_o)_{mcl}\right)^T\Sigma_m^{-1}\left(x_m^0+(\alpha M_o)_{mcl}\right) \end{aligned} \quad (13)$$

$$J^{\alpha}_{T^2,a} = \left(x_a^0+\alpha M_y\right)^T\Sigma_a^{-1}\left(x_a^0+\alpha M_y\right). \quad (14)$$

The sensitivity to the attack α can be quantified by

$$\Delta_m^{\alpha} = \left(J^{\alpha}_{T^2,m} - J^{0}_{T^2,m}\right)/J^{0}_{T^2,m} \quad (15)$$

$$\Delta_a^{\alpha} = \left(J^{\alpha}_{T^2,a} - J^{0}_{T^2,a}\right)/J^{0}_{T^2,a} \quad (16)$$

$$\Delta_c^{\alpha} = \left(\left\|x_c^{\alpha}-x_c^0\right\|\right)\Big/\left\|x_c^0\right\| \quad (17)$$

which are determined by the attack-free terms x_m^0, x_a^0, x_c^0, Σ_m, and Σ_a, and the attack-related terms $(\alpha M_o)_{mcl}$, $(\alpha M_o)_{ccl}$, and αM_y. Since $M_y = I - M_o$, it holds that $\lim_{M_o \to I}\Delta_a^{\alpha} = 0$. Also, $\lim_{(\alpha M_o)_{mcl}\to 0}\Delta_m^{\alpha} = 0$ and $\lim_{(\alpha M_o)_{ccl}\to 0}\Delta_c^{\alpha} = 0$.

Recall that X_a is fully correlated to Y and that X_m and X_c are uncorrelated to Y. Thus, the subspaces spanned by the occupied channels, i.e., $\mathcal{S}_1$ and $\mathcal{S}_2$, are only partially correlated to Y. In this sense, the prediction of Y by these two subspaces is dominant by the uncorrelated parts and the contributions $\hat{Y}_a|_{\mathcal{S}_1,\mathcal{S}_2}$ are minor, where the denotion $\hat{\Omega}|_{\mathcal{S}_i}$ indicates the partial prediction of Ω that is contributed by the subspace $\mathcal{S}_i$. As a result, it can be derived that $T_y|_{\mathcal{S}_1,\mathcal{S}_2} \to 0$, $P_y|_{\mathcal{S}_1,\mathcal{S}_2} \to 0$, and $M_o|_{\mathcal{S}_1,\mathcal{S}_2} \to I$. In contrast, the subspace spanned by the vacant channels, i.e., $\mathcal{S}_3$, is still fully correlated to Y, and the elements in $M_o|_{\mathcal{S}_3}$ are generally nonzero. Specifically, by introducing the notation $\alpha = [\alpha_{1,1}\cdots\alpha_{1,m}, \alpha_{2,1}\cdots\alpha_{2,c}, \alpha_{3,1}\cdots\alpha_{3,v}]$, then

$$(\alpha M_o)_{ccl} = \begin{cases} \alpha\cdot(M_o)_{ccl}, & M_o|_{\mathcal{S}_1,\mathcal{S}_2} \neq I \\ \left[\alpha_{2,v+1}\cdots\alpha_{2,c}\ \ \alpha_3\right], & M_o|_{\mathcal{S}_1,\mathcal{S}_2} = I, v < c \\ \left[\alpha_{3,v-c+1}\cdots\alpha_{3,v}\right], & M_o|_{\mathcal{S}_1,\mathcal{S}_2} = I, v \geq c. \end{cases} \quad (18)$$

It is straightforward from the above equations that following the current transmission strategy, the attacks to the measurement channels are undetectable by the check code, and if $v \geq c$, the attacks to the check code channels are also undetectable. From

$$(\alpha M_o)_{mcl} = \begin{cases} \alpha\cdot(M_o)_{mcl}, & M_o|_{\mathcal{S}_1,\mathcal{S}_2} \neq I \\ \alpha_1, & M_o|_{\mathcal{S}_1,\mathcal{S}_2} = I \end{cases} \quad (19)$$

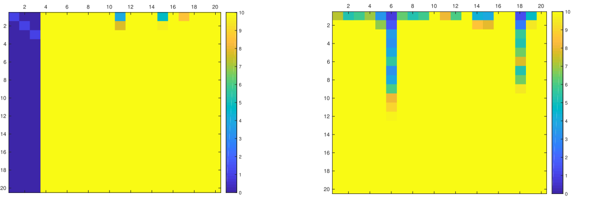
and let $\Sigma^{-1} = \Gamma\cdot\Gamma^T$, it can be further derived that

$$\Delta_m^{\alpha} = \frac{\left\|\left(x_m^0+\alpha_1\right)^T\Gamma\right\|^2 - \left\|x_m^{0\,T}\Gamma\right\|^2}{\left\|x_m^{0\,T}\Gamma\right\|^2} \leq \frac{\left\|\alpha_1^T\Gamma\right\|^2}{\left\|x_m^{0\,T}\Gamma\right\|^2}. \quad (20)$$

In case only a limited number of channels are compromised or the attack is sparse (nonconsecutive), the sensitivity is low.

To deal with this, it is needed to ensure $M_o|_{\mathcal{S}_1,\mathcal{S}_2} \neq I$, and by tracing back from the above analysis, to increase the correlation of the information subspaces $\mathcal{S}_1, \mathcal{S}_2$ with Y. To this end, it is proposed to conduct dual transmission.

Let $X_a = [X_{a,1}^{N\times m}\ \ X_{a,2}^{N\times m}\ \ X_{a,3}^{N\times c}\ \ X_{a,4}^{N\times c}\ \ X_{a,5}^{N\times v}] \in \mathbb{R}^{N\times a}$, $X_m \in \mathbb{R}^{N\times m}$, $X_c \in \mathbb{R}^{N\times c}$. Consider a superposition rule $\mathcal{R}_{s,2}$ that leads to the following relationship (referred to as dual transmission):



(a) (b)

Fig. 2. Elements in the matrix M_o. (a) Using single transmission. (b) Using dual transmission.

$$X^{\text{enc}} = \big[\underbrace{X_{a,1}+X_m, X_{a,2}+X_m}_{\mathcal{S}_1'}\ \ \underbrace{X_{a,3}+X_c, X_{a,4}+X_c}_{\mathcal{S}_2'}\ \ \underbrace{X_{a,5}}_{\mathcal{S}_3'}\big]. \quad (21)$$

Since $(X_{a,1}-X_{a,2})$ and $(X_{a,3}-X_{a,4})$ are fully correlated with Y, it holds that $T_y|_{\mathcal{S}_1',\mathcal{S}_2'} \nrightarrow 0$, $P_y|_{\mathcal{S}_1',\mathcal{S}_2'} \nrightarrow 0$, and $M_o|_{\mathcal{S}_1',\mathcal{S}_2'} \nrightarrow I$. In this strategy, attacks to all the channels can be detected with good sensitivity.

An illustrative comparison is provided in Fig. 2. It can be seen that by using single transmission, the elements in the columns corresponding to the information channels (the first three columns) are very close to zero. As a result, if abnormalities occur in these channels, they are less likely to be detected due to small weights. In contrast, when using dual transmission, all the elements are nonzero and the weighing of variations in different channels is more reasonable. Therefore, dual transmission according to a superposition rule like $\mathcal{R}_{s,2}$ is recommended.

Remark 4: If the attacker adds an identical bias term ξ to the channels that carry the sensor measurements (or the check codes), which can be written by $x^{enc,\xi} = [x_{a,1}+x_m+\xi, x_{a,2}+x_m+\xi, x_{a,3}+x_c, x_{a,4}+x_c, x_{a,5}]$, the attack will be treated as the measurements ($x_m' \leftarrow x_m+\xi$) or the check codes ($x_c' \leftarrow x_c+\xi$). Nevertheless, this can be easily prevented by introducing an invertible nonlinear function $\mathcal{G}(\cdot)$ before transmission and recovering by $\mathcal{G}^{-1}(\cdot)$ at the receiver end. Following this, $x^{dec,\Xi} = [\mathcal{G}^{-1}(\mathcal{G}(x_{a,1}+x_m)+\xi), \mathcal{G}^{-1}(\mathcal{G}(x_{a,2}+x_m)+\xi), x_{a,3}+x_c, x_{a,4}+x_c, x_{a,5}]$ where the attacks are detectable. Invertibility is to ensure that there is zero precision loss caused by the nonlinear projection at the receiver's end compared with the sender's end. This is not considered as a conservatism because there are unlimited number of invertible nonlinear functions. In condition that some degree of prevision loss is acceptable, the assumption can be relaxed. In such case, autoencoders can also be used.

Remark 5: If not all the channels are occupied, the vacant channels are fully correlated to Y_a and thus may be utilized by the attackers (in case they have such knowledge about the defense system) to solve for the decryption matrix. To deal with this, it is proposed to superposed uncorrelated disturbance signals to these channels.

Remark 6: As an internally and naturally occurring type of abnormality, packet dropout has typical features to be easily distinguished from the external attacks. While it will lead to degradation in the trustworthiness judgement performance if the packet dropout rate is very high, such phenomenon can be

TABLE V
ALGORITHM 1. IMPLEMENTATION PROCEDURES

OFFLINE DESIGN PHASE

Step 1. Solving for M_o according to *Table IV*

The *Step 6* therein is implemented according to the rule $\mathcal{R}_{s,2}$.

Step 2. Threshold setup for trustworthiness judgement

1) Calculate the threshold of the T^2 statistics.
$J_{th,T^2,a} = \chi_\beta(m)$ where β denotes the confidence level.
2) Set the threshold $\Delta_{th,c}$ for the error in the decrypted check code.

ONLINE IMPLEMENTATION PHASE

Step 1. Encryption at the sender end

1) Generate random numbers $sigGen$, and calculate
$x_{a,new} = sigGen \cdot M_{AC}$ and $y_{a,new} = x_{a,new} \cdot M_{GC}$.
2) Generate check code $x_{c,new}$.
3) Collect online sensor measurement data $x_{m,new}$.
4) According to the superposition rule $\mathcal{R}_{s,2}$, add the sensing data and check code to different channels of the auxiliary masking data.
5) Add arbitrary uncorrelated disturbance signals to all the vacant channels.
6) Perform nonlinear transformation using $\mathcal{G}(\cdot)$.
7) Use the encrypted data for network transmission.

Step 2. Data decryption and trustworthiness judgement

1) Generate the check code x_c^r at the receiver end.
2) Performance (inverse) nonlinear transformation using $\mathcal{G}^{-1}(\cdot)$.
3) Decrypt the received data x_r by $x_{dec} = x_r \cdot M_o$.
4) Extract the check codes x_c^{dec}, the measurement signals x_m^{dec}, and the auxiliary masking signals x_a^{dec} from the corresponding channels of x^{dec}.
5) Determine whether the measurements are trustworthy and record into logs: If $||x_c^{dec} - x_c^r||_2^2 \leq \Delta_{th,c}$ and $J_{T^2,a} \leq J_{th,T^2,a}$, then trustworthy; Otherwise, untrustworthy.

suppressed by simply adding a module to determine whether or not the malfunction is caused by packet dropout.

E. Complete Implementation Procedures

Based on the above proposal and analysis throughout Sections III-A–III-D, the complete implementation procedures are summarized as an algorithm in Table V.

If the distribution of measurement data deviates far from χ^2, to reduce conservatism, nonparametric approaches such as kernel density estimation can be used to calculate the online indexes.

IV. Validation and Performance Evaluation

Due to the limit of space, a part of the definitions of the symbols, the modeling process, and enlarged result figures are provided in the "supplementary material" available at https://drive.google.com/file/d/1S97-3zG-5M7mGFthbPu5x88Nh6iuFZdh/view?usp=sharing.

A. Experiment Design

In this part, a power grid system that relies on long-range measurement data transmission is used for simulation [37]–[39]. As shown in Fig. 3, the overall system constitutes of two local areas: 1) a wide-area measurement unit and 2) a control center. In each area, there are two generators (GEN), an equivalent load (LOAD), and an energy storage system (ESS). The two areas are connected by a high-voltage alternating current (HVAC) tie line and a high-voltage direct current (HVDC) link that work in parallel. These components are interconnected for load frequency control.

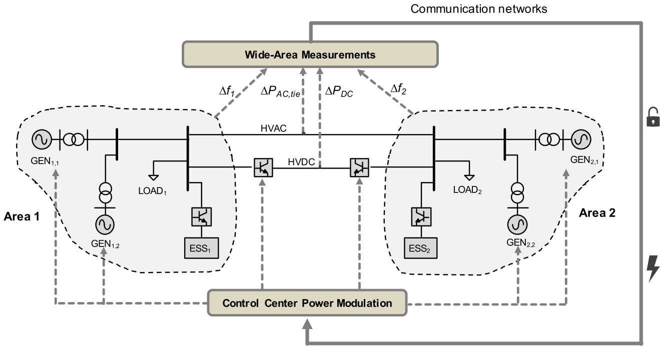


Fig. 3. Schematic of the two-area power grid with HVAC, HVDC, and ESS.

TABLE VI
DEFINITION OF THE INTEGRITY ATTACK SCENARIOS

Attack ID	Type of attack	Channel under attack
$Attack_1$	DoS	Measurement
$Attack_2$	DoS	Check code
$Attack_3$	DoS	Vacant
$Attack_4$	FDIA	Measurement
$Attack_5$	FDIA	Check code
$Attack_6$	FDIA	Vacant
$Attack_7$	Replay	Measurement
$Attack_8$	Replay	Check code
$Attack_9$	Replay	Vacant

The symbols and parameters used in this study are listed in the supplementary material (Tables I and II). Four signals (Δf_1, Δf_2, ΔP_{DC}, and $\Delta P_{AC,\text{tie}}$) are transmitted via communication networks such as the SCADA system, which are directly exposed to the attacks.

The measurement data are generated based on the models described in the supplementary material (Sections I-C and I-D).

The sampling frequency is 50 Hz. The system runs for a total of 100 s and correspondingly, 5000 samples are collected for each scenario.

For a comprehensive performance analysis, six types of loads and nine types of integrity attacks are defined, which covers all the typical scenarios of attack-free, DoS attack, false data injection attack, and replay attack, to different channels of transmission. The nine attack scenarios are defined in Table VI. The training dataset and the test dataset are built following Table VII. A set of different loads is designed and defined in Table VIII. Our intention to set the duration of Load_6 as 30–100 s is to simulate the condition of a sudden constant load change at the 30th second. It can be seen from Fig. 4(k) and (l) that such condition does not cause additional false alarms. The mathematical formulation are as follows:

$$\mathbf{y}_i^{\text{DoS}}(k) = \begin{cases} \mathbf{y}_i^{\text{DoS}}(k-1), & \text{under DoS attack} \\ \mathbf{y}_i(k), & \text{attack-free} \end{cases} \tag{22}$$

$$\mathbf{y}_i^{\text{FDIA}}(k) = \begin{cases} \mathbf{y}_i(k) + \text{rand}, & \text{under FDIA attack} \\ \mathbf{y}_i(k), & \text{attack-free} \end{cases} \tag{23}$$

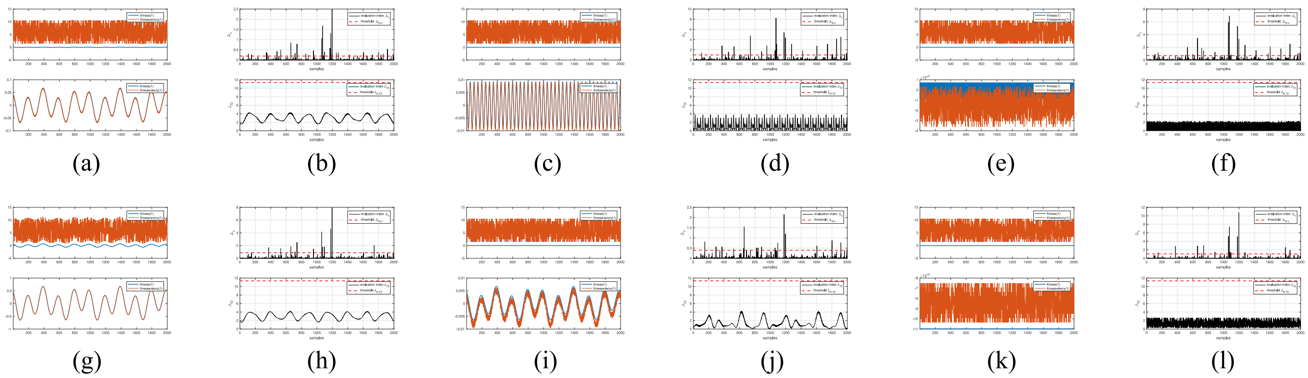

(a) (b) (c) (d) (e) (f)
(g) (h) (i) (j) (k) (l)

Fig. 4. Integrity attack not applied (zoomed-in plots are available in the supplementary material II-B). (a) Load_1. (b) Load_1. (c) Load_2. (d) Load_2. (e) Load_3. (f) Load_3. (g) Load_4. (h) Load_4. (i) Load_5. (j) Load_5. (k) Load_6. (l) Load_6.

TABLE VII
CONFIGURATION FOR DATASET CONSTRUCTION

Time	Sample index	Interpretation
0–2s	1–100	Establish steady state
2s–60s	101–3,000	Training dataset
60s–100s	3,001–5,000	Testing dataset
76s–96s	3,801–4,800	Integrity attack applied

TABLE VIII
LOAD SIGNAL

Load ID	ΔP_{d_1}	ΔP_{d_2}	Duration
$Load_1$	$0.05\sin(2\pi k/500)$	$0.2\cos(2\pi k/200)$	0-100s
$Load_2$	$0.05\sin(2\pi k/50)$	$0.2\cos(2\pi k/20)$	0-100s
$Load_3$	$0.05\sin(2\pi k/5)$	$0.2\cos(2\pi k/2)$	0-100s
$Load_4$	$0.5\sin(2\pi k/500)$	$2\cos(2\pi k/200)$	0-100s
$Load_5$	$0.005\sin(2\pi k/500)$	$0.02\cos(2\pi k/200)$	0-100s
$Load_6$	0	0.03	30-100s

$$\mathbf{y}_i^{\text{replay}}(k) = \begin{cases} \mathbf{y}_i(k-\kappa), & \text{under replay attack} \\ \mathbf{y}_i(k), & \text{attack-free.} \end{cases} \tag{24}$$

For the generation of check codes, the following Lorenz system is employed, which is a typical chaotic system:

$$\begin{cases} \dot{h}_1 = 10(h_2 - h_1) \\ \dot{h}_2 = 28h_1 - h_2 - h_1h_3 \\ \dot{h}_3 = h_1h_2 - \frac{8}{3}h_3 \end{cases} \tag{25}$$

where the initial condition is set as $h(0) = [0.99 \;\; 1 \;\; 1]^T$, and the MATLAB solver is ode45. The check code is selected as $h_1(k)$ where k is the index of the sample. The applied nonlinear transformation and inverse transmission: $\mathcal{G}(x) = x^3$, $\mathcal{G}^{-1}(\bar{x}) = \bar{x}^{1/3}$. The adopted superposition rule $\mathcal{R}_s$ is to use the first six channels to carry measurement data, and use the seventh channel to carry check code. The trustworthiness judgement decision logics are defined as follows.

1) *Combined Index:* Attack occurs iff $\Delta_c > \Delta_{th,c}$ and $J_{T^2} > J_{th,T^2}$.
2) *Individual Index:* Attack occurs iff $\Delta_c > \Delta_{th,c}$.
3) *Individual Index:* Attack occurs iff $J_{T^2} > J_{th,T^2}$.

B. Results Analysis

Figs. 4 and 5 show the effects of encryption and decryption, and the results of trustworthiness judgement corresponding to different loads and integrity attacks, respectively. Table IX summarizes the key (statistical) performance evaluators, including the false alarm rates (FARs) and the miss detection rates (MDRs) based on different decision-making logics, as well as the root-mean-squared errors (RMSEs) and the mean absolute errors (MAE) in signal reconstruction. By studying these, the following conclusions are drawn.

1) All the encrypted data used for transmission deviate significantly from the actual measurement values. The external attackers cannot reconstruct the measurements without the decryption matrix and the check codes. Thus, the proposed scheme can effectively prevent eavesdropping attacks. (Learned from Figs. 4 and 5).
2) Differences in the loads have little influence on the decision-making performance (FAR) or on the signal reconstruction performance (RMSE/MAE) (learned from Table IX, *No attack*, Load_{1-6}).
3) Attacks to the measurement channels and the check code channels can be effectively detected, whereas attacks to the vacant channels cannot be detected (learned from Table IX, Attack_{1-9}).
4) There is no significant difference in the influence from the type of attacks (DoS, FDIA, and Replay attack) (learned from Table IX, Attack_{1-9}).
5) Regarding the decision-making results of FAR and MDR, the combined index performs better than the individual index of Δ_c (learned from Table IX, $\text{Attack}_{1,4,7}$).
6) The T^2 index performs well in terms of identifying whether the attacks are applied to the measurement channels (learned from Table IX, Attack_{1-9}).

Fig. 6 shows the introduced time delays corresponding to different batch sizes while Table X lists the exact values of the time delays at each stage. It can be learned as follows.

1) The total introduced time delay is around 8.16 μs. This is well acceptable because it is negligible compared with the sampling time (0.04 s).
2) The decision-making stage causes the longest delay compared with the encryption and the deception stages.

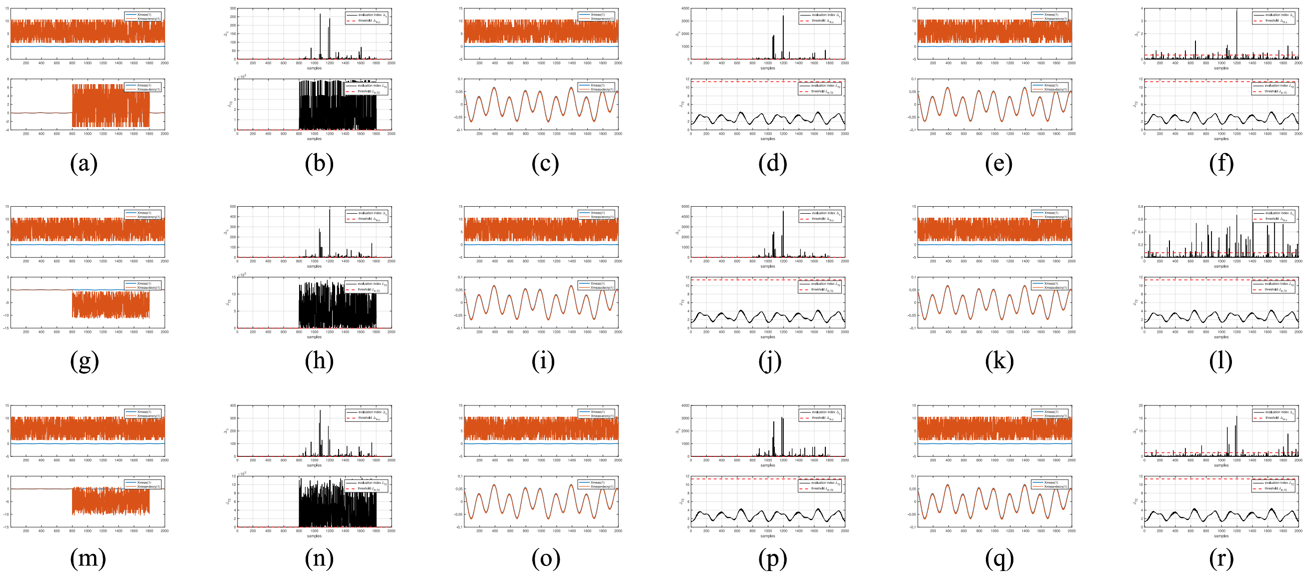


Fig. 5. Integrity attacks applied (zoomed-in plots are available in the supplementary material II-B). (a) Attack$_1$. (b) Attack$_1$. (c) Attack$_2$. (d) Attack$_2$. (e) Attack$_3$. (f) Attack$_3$. (g) Attack$_4$. (h) Attack$_4$. (i) Attack$_5$. (j) Attack$_5$. (k) Attack$_6$. (l) Attack$_6$. (m) Attack$_7$. (n) Attack$_7$. (o) Attack$_8$. (p) Attack$_8$. (q) Attack$_9$. (r) Attack$_9$.

TABLE IX
PERFORMANCE OF TRUSTWORTHINESS JUDGEMENT AND SIGNAL RECONSTRUCTION

Configuration	FAR (Δ_c)	MDR (Δ_c)	FAR (J_{T2})	MDR (J_{T2})	FAR (Combined)	MDR (Combined)	RMSE	MAE	Plot
$No\ attack,\ Load_1$	0.012	-	0	-	0.012	-	0.0031	0.0028	Fig. 4(b)
$No\ attack,\ Load_2$	0.014	-	0	-	0.014	-	0.0016	0.0014	Fig. 4(d)
$No\ attack,\ Load_3$	0.013	-	0	-	0.013	-	0.0013	0.0012	Fig. 4(f)
$No\ attack,\ Load_4$	0.007	-	0	-	0.007	-	0.0079	0.0071	Fig. 4(h)
$No\ attack,\ Load_5$	0.009	-	0	-	0.009	-	0.0022	0.0020	Fig. 4(j)
$No\ attack,\ Load_6$	0.008	-	0	-	0.008	-	0.0029	0.0026	Fig. 4(l)
$Attack_1$	0.022	0.058	0	0.025	**0.022**	**0.019**	0.0032	0.0029	Fig. 5(b)
$Attack_2$	0.009	0.117	0	1	0.009	0.117	0.0026	0.0024	Fig. 5(d)
$Attack_3$	0.011	0.984	0	1	0.011	0.984	0.0028	0.0025	Fig. 5(f)
$Attack_4$	0.014	0.502	0	0	**0.014**	**0**	0.0030	0.0027	Fig. 5(h)
$Attack_5$	0.011	0.068	0	1	0.011	0.068	0.0033	0.0029	Fig. 5(j)
$Attack_6$	0.021	0.973	0	1	0.021	0.973	0.0026	0.0023	Fig. 5(l)
$Attack_7$	0.023	0.148	0	0.004	**0.023**	**0.004**	0.0028	0.0026	Fig. 5(n)
$Attack_8$	0.012	0.067	0	1	0.012	0.067	0.0034	0.0031	Fig. 5(p)
$Attack_9$	0.009	0.989	0	1	0.009	0.989	0.0027	0.0025	Fig. 5(r)

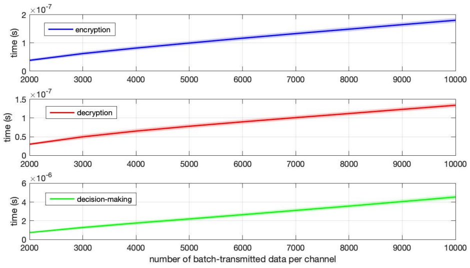


Fig. 6. Characteristic of time delay: the introduced delays by each stage per transmission per channel. The mean values and the standard deviations are calculated with 1000 Monte Carlo simulations.

3) An increase in the size of the batch, that is, used for transmission leads to slightly larger time delays. This results from inefficiencies in large-matrix computations.

Table XI summarizes the adjustable elements, which span the keyspace of encryption and contribute to the design DoF.

TABLE X
AVERAGE TIME DELAY CAUSED BY EACH STAGE (BATCH SIZE = 1, CHANNEL NUMBER = 10)

Stage	Encryption	Decryption	Decision-making	Total
Time delay*	0.38μs	0.30μs	7.48μs	8.16μs

*Average of $1,000$ Monte Carlo simulations.

Compared with signature-based approaches and the anomaly detection-based approaches, auxiliary signals provide an additional DoF. Besides, the flexibility in choosing superposition rules, which could be time varying, significantly enhances the system's security and robustness where external parties cannot launch social engineering attacks to the designers/system engineers. This can be treated as an additional DoF as most existing approaches do not have the concept of superposition rules [11], [22], [26].

TABLE XI
DESIGN DOF

Module	**Design parameter**	**Example**
Auxiliary Signal Generator	Type	Randi(10)
	Mean	0.5
	Variance	1
Correlative matrices	M_{AC} value	$[1:10]$
	M_{GC} value	$[15:24; 11:20]$
	M_{GC} channel	2
Encryption	Superposition rule	Use the first $2l$ channels to carry measurement data; use the next n_c channels to carry check codes
	Nonlinear transformation	$\mathcal{G}(x) = x^3$
Check code	Type	Output of a Lorenz system
	Number/channel	1

In future work, it is meaningful to put more effort into: 1) exploring other possible auxiliary signals; 2) the design of the superposition rules; and 3) the optimal selection of the type of the check codes.

V. Conclusion

In this article, an integrated data-driven framework is proposed to improve the cybersecurity of the industrial data transmitted through networks. It is designed to defend against eavesdropping attacks by masking the measurement data with randomized auxiliary signals, and in the meantime, perform trustworthiness judgement at the receiver end. The proposed auxiliary signal-based approach leads to a good level of security because a necessary part of the information (Y_a) for decryptor design is never transmitted. On this basis, the detectability and sensitivity are further guaranteed with rigorous mathematical derivations, based on which essential tricks like "dual transmission" and "nonlinear transformation" are included in the implementation procedures to further improve security. Tests on a simulated two-area power grid validate the theoretical outcomes and provide additional performance evaluation results with respect to the introduced time delay.

References

[1] A. W. Colombo, S. Karnouskos, O. Kaynak, Y. Shi, and S. Yin, "Industrial cyberphysical systems: A backbone of the fourth industrial revolution," *IEEE Ind. Electron. Mag.*, vol. 11, no. 1, pp. 6–16, Mar. 2017.

[2] V. C. Gungor *et al.*, "A survey on smart grid potential applications and communication requirements," *IEEE Trans. Ind. Informat.*, vol. 9, no. 1, pp. 28–42, Feb. 2013.

[3] Y.-G. Li and G.-H. Yang, "Optimal stealthy innovation-based attacks with historical data in cyber-physical systems," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 6, pp. 3401–3411, Jun. 2021.

[4] M. Khari, A. K. Garg, A. H. Gandomi, R. Gupta, R. Patan, and B. Balusamy, "Securing data in Internet of Things using cryptography and steganography techniques," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 50, no. 1, pp. 73–80, Jan. 2020.

[5] H. Pearce, S. Pinisetty, P. S. Roop, M. M. Y. Kuo, and A. Ukil, "Smart I/O modules for mitigating cyber-physical attacks on industrial control systems," *IEEE Trans. Ind. Informat.*, vol. 16, no. 7, pp. 4659–4669, Jul. 2020.

[6] M. Cheminod, L. Durante, and A. Valenzano, "Review of security issues in industrial networks," *IEEE Trans. Ind. Informat.*, vol. 9, no. 1, pp. 277–293, Feb. 2013.

[7] Y. Xia, S. Su, H. Su, X. Zhang, W. Luo, and W. Yang, "Detection of data integrity attacks in distributed state estimation," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 12, pp. 7735–7744, Dec. 2021.

[8] D. Ding, Q.-L. Han, Z. Wang, and X. Ge, "Recursive filtering of distributed cyber-physical systems with attack detection," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 10, pp. 6466–6476, Oct. 2021.

[9] X.-Y. Shen and X.-J. Li, "Data-driven output-feedback LQ secure control for unknown cyber-physical systems against sparse actuator attacks," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 9, pp. 5708–5720, Sep. 2021.

[10] V. C. Gungor *et al.*, "Smart grid technologies: Communication technologies and standards," *IEEE Trans. Ind. Informat.*, vol. 7, no. 4, pp. 529–539, Nov. 2011.

[11] S. M. Dibaji *et al.*, "A systems and control perspective of CPS security," *Annu. Rev. Control*, vol. 47, pp. 394–411, Jun. 2019.

[12] R. Deng, G. Xiao, and R. Lu, "Defending against false data injection attacks on power system state estimation," *IEEE Trans. Ind. Informat.*, vol. 13, no. 1, pp. 198–207, Feb. 2017.

[13] M. Usman, M. A. Jan, A. Jolfaei, M. Xu, X. He, and J. Chen, "DaaC: A distributed and anonymous data collection framework based on multilevel edge computing architecture," *IEEE Trans. Ind. Informat.*, vol. 16, no. 9, pp. 6114–6123, Sep. 2020.

[14] P. C. M. Arachchige, P. Bertok, I. Khalil, D. Liu, S. Camtepe, and M. Atiquzzaman, "A trustworthy privacy preserving framework for machine learning in industrial IoT systems," *IEEE Trans. Ind. Informat.*, vol. 16, no. 9, pp. 6092–6102, Sep. 2020.

[15] S. Yin, J. J. Rodriguez-Andina, and Y. Jiang, "Real-time monitoring and control of industrial cyberphysical systems with integrated plant-wide monitoring and control framework," *IEEE Ind. Electron. Mag.*, vol. 13, no. 4, pp. 38–47, Dec. 2019.

[16] Y. Jiang, S. Yin, K. Li, H. Luo, and O. Kaynak, "Industrial applications of digital twins," *Philos. Trans. R. Soc. A Math. Phy. Eng. Sci.*, vol. 379, no. 2270, 2021, Art. no. 20200360s.

[17] S. A. Foroutan and F. R. Salmasi, "Detection of false data injection attacks against state estimation in smart grids based on a mixture gaussian distribution learning method," *IET Cyber Phys. Syst. Theory Appl.*, vol. 2, no. 4, pp. 161–171, 2017.

[18] X. Li and K. W. Hedman, "Enhancing power system cyber-security with systematic two-stage detection strategy," *IEEE Trans. Power Syst.*, vol. 35, no. 2, pp. 1549–1561, Mar. 2020.

[19] M. Porter, P. Hespanhol, A. Aswani, M. Johnson-Roberson, and R. Vasudevan, "Detecting generalized replay attacks via time-varying dynamic watermarking," *IEEE Trans. Autom. Control*, vol. 66, no. 8, pp. 1–16, Aug. 2021, doi: 10.1109/TAC.2020.3022756.

[20] O. A. Beg, T. T. Johnson, and A. Davoudi, "Detection of false-data injection attacks in cyber-physical DC microgrids," *IEEE Trans. Ind. Informat.*, vol. 13, no. 5, pp. 2693–2703, Oct. 2017.

[21] S. Wu, Y. Jiang, H. Luo, J. Zhang, S. Yin, and O. Kaynak, "An integrated data-driven scheme for the defense of typical cyber-physical attacks," *Rel. Eng. Syst. Safety*, vol. 220, Apr. 2022, Art. no. 108257.

[22] D. Wang, J. Huang, Y. Tang, and F. Li, "A watermarking strategy against linear deception attacks on remote state estimation under K-L divergence," *IEEE Trans. Ind. Informat.*, vol. 17, no. 5, pp. 3273–3281, May 2021, doi: 10.1109/TII.2020.3009874.

[23] X. Ge, Q.-L. Han, M. Zhong, and X.-M. Zhang, "Distributed krein space-based attack detection over sensor networks under deception attacks," *Automatica*, vol. 109, Nov. 2019, Art. no. 108557.

[24] W. He, T. Luo, Y. Tang, W. Du, Y.-C. Tian, and F. Qian, "Secure communication based on quantized synchronization of chaotic neural networks under an event-triggered strategy," *IEEE Trans. Neural. Netw. Learn. Syst.*, vol. 31, no. 9, pp. 3334–3345, Sep. 2020.

[25] H. Tao *et al.*, "TrustData: Trustworthy and secured data collection for event detection in industrial cyber-physical system," *IEEE Trans. Ind. Informat.*, vol. 16, no. 5, pp. 3311–3321, May 2020.

[26] D. Ding, Q.-L. Han, X. Ge, and J. Wang, "Secure state estimation and control of cyber-physical systems: A survey," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 1, pp. 176–190, Jan. 2021.

[27] M. T. Khan and I. Tomic, "Securing industrial cyber-physical systems: A run-time multi-layer monitoring," *IEEE Trans. Ind. Informat.*, vol. 17, no. 9, pp. 6251–6259, Sep. 2021, doi: 10.1109/TII.2020.3032968.

[28] F. Zhang, H. A. D. E. Kodituwakku, J. W. Hines, and J. Coble, "Multilayer data-driven cyber-attack detection system for industrial control systems based on network, system, and process data," *IEEE Trans. Ind. Informat.*, vol. 15, no. 7, pp. 4362–4369, Jul. 2019.

[29] F. Farivar, M. S. Haghighi, A. Jolfaei, and M. Alazab, "Artificial intelligence for detection, estimation, and compensation of malicious attacks in nonlinear cyber-physical systems and industrial IoT," *IEEE Trans. Ind. Informat.*, vol. 16, no. 4, pp. 2716–2725, Apr. 2020.

[30] Y. Liu, A. Liu, X. Liu, and M. Ma, “A trust-based active detection for cyber-physical security in industrial environments,” *IEEE Trans. Ind. Informat.*, vol. 15, no. 12, pp. 6593–6603, Dec. 2019.

[31] Y. Xie, G. Zeng, R. Kurachi, H. Takada, and G. Xie, “Security/timing-aware design space exploration of can FD for automotive cyber-physical systems,” *IEEE Trans. Ind. Informat.*, vol. 15, no. 2, pp. 1094–1104, Feb. 2019.

[32] M. S. Rahman, M. A. Mahmud, A. M. T. Oo, and H. R. Pota, “Multi-agent approach for enhancing security of protection schemes in cyber-physical energy systems,” *IEEE Trans. Ind. Informat.*, vol. 13, no. 2, pp. 436–447, Apr. 2017.

[33] R. Deng, G. Xiao, R. Lu, H. Liang, and A. V. Vasilakos, “False data injection on state estimation in power systems—Attacks, impacts, and defense: A survey,” *IEEE Trans. Ind. Informat.*, vol. 13, no. 2, pp. 411–423, Apr. 2017.

[34] Y. Jiang and S. Yin, “Recent advances in key-performance-indicator oriented prognosis and diagnosis with a MATLAB toolbox: DB-KIT,” *IEEE Trans. Ind. Informat.*, vol. 15, no. 5, pp. 2849–2858, May 2019.

[35] G. Wang, H. Luo, and K. Peng, “Quality-related fault detection using linear and nonlinear principal component regression,” *J. Franklin Inst.*, vol. 353, no. 10, pp. 2159–2177, 2016.

[36] Y. Jiang and S. Yin, “Recursive total principle component regression based fault detection and its application to vehicular cyber-physical systems,” *IEEE Trans. Ind. Informat.*, vol. 14, no. 4, pp. 1415–1423, Apr. 2018.

[37] K. Pan, J. Dong, E. Rakhshani, and P. Palensky, “Effects of cyber attacks on AC and high-voltage DC interconnected power systems with emulated inertia,” *Energies*, vol. 13, no. 21, p. 5583, 2020.

[38] K. Pan, E. Rakhshani, and P. Palensky, “False data injection attacks on hybrid AC/HVDC interconnected systems with virtual inertia—Vulnerability, impact and detection,” *IEEE Access*, vol. 8, pp. 141932–141945, 2020.

[39] E. Rakhshani and P. Rodriguez, “Inertia emulation in AC/DC interconnected power systems using derivative technique considering frequency measurement effects,” *IEEE Trans. Power Syst.*, vol. 32, no. 5, pp. 3338–3351, Sep. 2017.

Yuchen Jiang (Member, IEEE) received the B.E. degree in automation and the Ph.D. degree in control science and engineering from the Harbin Institute of Technology, Harbin, China, in 2016 and 2021, respectively.

He was an intern with the Industrial Intelligence Academician Studio, Peng Cheng Laboratory, Shenzhen, China. He is currently a Lecturer with the School of Astronautics, Harbin Institute of Technology. His research interests include data-driven process monitoring, fault diagnosis and prognosis, industrial cyber–physical systems, and artificial intelligence.

Shimeng Wu received the B.E. degree in automation from Harbin Engineering University, Harbin, China, in 2020. She is currently pursuing the M.S. degree in control science and engineering with the Harbin Institute of Technology, Harbin.

Her research interests include fault diagnosis and prognosis, security of cyber-physical systems, and artificial intelligence.

Hongyan Yang (Member, IEEE) received the B.S. degree in mathematics and applied mathematics and the M.S. degree in optimization and automatic control theory from the College of Mathematics and Physics, Bohai University, Jinzhou, China, in 2013 and 2016, respectively, and the Ph.D. degree in control theory and control engineering from the Harbin Institute of Technology, Harbin, China, in 2020.

She is currently a Lecturer with the Beijing University of Technology, Beijing, China. Her current research interests include fault diagnosis and fault-tolerant control of nonlinear systems, Markovian jump systems, and cyber-physical systems.

Hao Luo (Senior Member, IEEE) received the B.E. degree in electrical engineering from Xi’an Jiaotong University, Xi’an, China, in 2007, and the M.Sc. and Ph.D. degrees in electrical engineering and information technology from the University of Duisburg–Essen, Duisburg, Germany, in 2012 and 2016, respectively.

He is currently a Professor with the School of Astronautics, Harbin Institute of Technology, Harbin, China. His research interests include model-based and data-driven fault diagnosis, fault-tolerant systems, and their plug-and-play application on industrial systems.

Zhiwen Chen (Member, IEEE) was born in Hunan, China. He received the B.S. degree in electronic information science and technology and the M.S. degree in electronic information and technology from Central South University, Changsha, China, in 2008 and 2012, respectively, and the Ph.D. degree in electrical engineering and information technology from the University of Duisburg–Essen, Duisburg, Germany, in 2016.

He is currently an Associate Professor with Central South University, and also a part-time Associate Researcher with Peng Cheng Laboratory, Shenzhen, China. His research interests are model-based and data driven-fault diagnosis, data analytics.

Shen Yin (Senior Member, IEEE) received the M.Sc. degree in control and information system and the Ph.D. degree in electrical engineering and information technology from the University of Duisburg-Essen, Duisburg, Germany, in 2007 and 2012, respectively.

He is currently a DNV-GL Professor with the Department of Mechanical and Industrial Engineering, Faculty of Engineering, Norwegian University of Science and Technology, Trondheim, Norway. His research interests include safety, reliability of complicated systems, system and control theory, data-driven and machine-learning approaches, applications in large-scale systems, and industrial cyber-physical systems.

Okyay Kaynak (Life Fellow, IEEE) received the B.Sc. (First-Class Hons.) and Ph.D. degrees in electronic and electrical engineering from the University of Birmingham, Birmingham, U.K., in 1969 and 1972, respectively.

From 1972 to 1979, he held various positions within the industry. In 1979, he joined Bogazici University, Istanbul, Turkey, where he is currently a Professor Emeritus, holding the UNESCO Chair on Mechatronics. He has held long-term (near to or more than a year) Visiting Professor/Scholar positions at various institutions in Japan, Germany, USA, Singapore, and China. He has authored three books and edited five and authored or coauthored more than 450 papers that have appeared in various journals, books, and conference proceedings. His current research interests include the fields of intelligent control and CPS.

Dr. Kaynak received the Chinese Government Friendship Award and Humboldt Research Prize in 2016. Most recently, in 2020, he received the International Research Prize of the Turkish Academy of Sciences. He is active in international organizations, has served on many committees of IEEE, and was the President of IEEE Industrial Electronics Society from 2002 to 2003. He is a member of Turkish Academy of Sciences.