Multilayer Data-Driven Cyber-Attack Detection System for Industrial Control Systems Based on Network, System, and Process Data

Fan Zhang, *Student Member, IEEE*, Hansaka Angel Dias Edirisinghe Kodituwakku, J. Wesley Hines, and Jamie Coble, *Senior Member, IEEE*

***Abstract*—The growing number of attacks against cyber-physical systems in recent years elevates the concern for cybersecurity of industrial control systems (ICSs). The current efforts of ICS cybersecurity are mainly based on firewalls, data diodes, and other methods of intrusion prevention, which may not be sufficient for growing cyber threats from motivated attackers. To enhance the cybersecurity of ICS, a cyber-attack detection system built on the concept of defense-in-depth is developed utilizing network traffic data, host system data, and measured process parameters. This attack detection system provides multiple-layer defense in order to gain the defenders precious time before unrecoverable consequences occur in the physical system. The data used for demonstrating the proposed detection system are from a real-time ICS testbed. Five attacks, including man in the middle (MITM), denial of service (DoS), data exfiltration, data tampering, and false data injection, are carried out to simulate the consequences of cyber attack and generate data for building data-driven detection models. Four classical classification models based on network data and host system data are studied, including k-nearest neighbor (KNN), decision tree, bootstrap aggregating (bagging), and random forest (RF), to provide a secondary line of defense of cyber-attack detection in the event that the intrusion prevention layer fails. Intrusion detection results suggest that KNN, bagging, and RF have low missed alarm and false alarm rates for MITM and DoS attacks, providing accurate and reliable detection of these cyber attacks. Cyber attacks that may not be detectable by monitoring network and host system data, such as command tampering and false data injection attacks by an insider, are monitored for by traditional process monitoring protocols. In the proposed detection system, an auto-associative kernel regression model is studied to strengthen early attack detection. The result shows that this approach detects physically impactful cyber attacks before significant consequences occur. The proposed multiple-layer data-driven cyber-attack detection system utilizing network, system, and process data is a promising solution for safeguarding an ICS.**

***Index Terms*—Cyber-attack detection, data-driven monitoring, defense-in-depth, industrial control system (ICS).**

I. Introduction

THE adoption of cyber-physical systems (CPSs) across a variety of industries is rapidly increasing. CPSs contain interactive physical assets and computational capabilities with information transfer [1]. An industrial control system (ICS) is a special class of CPS that involves the cyber aspect, which includes both a supervisory control and data acquisition (SCADA) system, and the physical industrial process system or facility. The rapid deployment of digitalization and development of CPSs leads to the widespread use of sensors, networked devices, and data acquisition systems. As ICS systems are deployed for high-value and safety-critical systems, security requirements are evolving to include both resilience to cyber attacks and situational awareness of cyber intrusions. Big data analytics can be used to distill information from the high volume of data available to make time sensitive decisions to detect threats and safeguard CPSs.

Fig. 1 shows the timeline of several major ICS targeted cyber attacks. In 2010, Stuxnet attacked Iranian nuclear enrichment centrifuges, causing severe equipment damage [2]. The malware was brought into the facility on a USB drive and exploited several zero-day vulnerabilities to inject its malicious code into Siemens programmable logic controllers to cause centrifuges to spin through their natural frequencies, causing them to wear at a much higher rate than expected. Meanwhile, the malware faked sensor outputs to mask the attack from operators. This early attack demonstrated security weaknesses and potential risk of cyber attacks against high-value ICSs. In 2015, BlackEnergy malware was used to attack the Ukrainian power grid and caused a large-scale power outage [3]. HatMan malware affected Triconex controllers through modifying inmemory firmware to add additional programming, which is believed to be behind the attack on critical infrastructure using Schneider Electrics Safety Instrumented Systems [4]. A report released in March 2018 revealed that around 40 percent of all ICS in energy organizations protected by Kaspersky Lab solutions were attacked by malware

Manuscript received September 29, 2018; revised November 26, 2018; accepted January 2, 2019. Date of publication January 7, 2019; date of current version July 3, 2019. This work was supported in part by the Lloyd's Register Foundation and the International Joint Research Center for the Safety of Nuclear Energy, and in part by Oracle Corporation. Paper no. TII-18-2564. *(Corresponding author: Jamie Coble.)*

F. Zhang, J. W. Hines, and J. B. Coble are with the Department of Nuclear Engineering, University of Tennessee, Knoxville, TN 37996 USA (e-mail: fan@utk.edu; jhines2@utk.edu; jamie@utk.edu).

H. A. D. E. Kodituwakku is with the Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, TN 37996 USA (e-mail: angelk@utk.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TII.2019.2891261

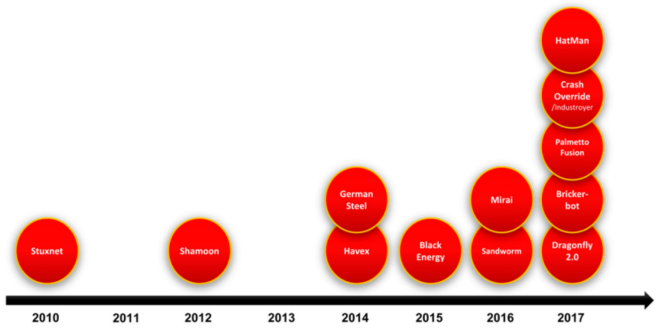

Fig. 1. Major ICS targeted cyber-attacks timeline.

at least once in the second half of 2017 [5]. There were many more ICS targeted cyber attacks in 2017, suggesting the trend in ICS cyber attacks will increase as these systems proliferate and vulnerabilities are identified. In April 2018, three U.S. natural gas pipeline companies claimed that their electronic communication systems with customers had been shut down for several days due to a cyber attack [6]. Even though this cyber event has not proved to cause data leakage, it indicates that the industry is vulnerable to cyber attacks.

Even though the commercial information technology (IT) cybersecurity methods are mature, they are not suitable to be adopted directly to ICS due to differences between ICS and IT. The security objectives of ICS are prioritized in the order of availability, integrity, and confidentiality, whereas the general IT are prioritized in the order of confidentiality, integrity, and availability [7]. ICS cyber attacks have potentially fatal and environmentally destructive consequences due to the close coupling with active physical processes. Therefore, ICS requires the highest security, including both intrusion prevention and intrusion detection. Given a motivated attacker with sufficient resources and time, cyber-attack success is highly likely even for the most robust intrusion prevention systems (IPS).

This research develops a multilayer, defense-in-depth-based intrusion detection systems (IDS) for robustly detecting intrusions in ICSs. The paper is organized as follows. Section II summarizes cyber attacks and the current state of IDS system development. Section III describes the structure of proposed system and machine learning algorithms investigated for the data-driven IDS. Section IV describes the ICS testbed architecture and composition, data collection, and the attack scenarios implemented to develop and demonstrate the proposed cyber-attack system. The results of the attack scenarios and their detection are given in Section V. The final section summarizes the conclusions of the current research and highlights some areas of future work to further develop the proposed cyber-attack detection system.

II. Background

A cyber attack may consist of several stages such as reconnaissance, denial of service (DoS), man in the middle (MITM), elevation of privilege, and data tampering [8]. Attackers first gather information about the target system during the reconnaissance phase to identify network topology, software versions, and critical targets to attack [9]. After attackers gain knowledge about the system, they plan further attacks by researching known vulnerabilities against the detected software versions; possible attack vectors and pivot points for bypassing firewalls and IPS; and options for removing evidence after the attack, such as deletion and manipulation of system logs. Based on this research and the ultimate goals of the attack, several types of attacks may be launched. DoS attacks against an ICS may aim to disrupt communication between the SCADA master and slaves, which could cause the SCADA master to lose control of local control systems and actuators. Privilege escalation may be needed to access the low-level hardware on a system or to read and write to protected system files; escalation can be achieved using zero-day attacks and known vulnerabilities in the operating system and software used. Interception of commands and sensor data can be performed using an MITM attack, while data tampering and false data injection attacks go a step further to modify sensor data in transit in order to mislead the monitoring systems and operators while the attack is in progress. Data tampering could alter SCADA masters commands to cause the actuator to actuate inappropriately; it could alter the feedback process data to manipulate the control; and it could alter data in a data historian to modify the operation log and system control-related data to obfuscate the details of the attack, which misleads the defender in postattack analysis.

IDS detect unauthorized access to the system. There are three types of IDS: signature-based, anomaly-based, and hybrid. Signature-based IDSs are developed to detect known attacks using their documented behavior. This class of IDS is very effective for known attacks with low false alarm rates but are not able to detect zero-day attacks since the IDS is not yet aware of this behavior. Anomaly-based IDSs, on the other hand, model the normal behavior using data mining techniques or machine learning algorithms and report deviations from normal behavior as an anomaly or potential attack. They are customized to the normal behavior of each system to detect attacks, including unknown attacks, making it difficult for attackers to learn the IDSs capabilities, further complicating attackers ability to launch undetectable attacks. The very nature of this makes anomaly-based IDSs result in a high number of false positives [10]. Auto-associative kernel regression (AAKR) models system has been previously investigated to analyze network data of a simulated SCADA for anomaly-based intrusion detection [11], [12]. The hybrid IDS is a combination of signature-based and anomaly-based detection; this approach combines the accuracy of signature-based approaches for known attacks with the generalizability of anomaly-based systems. Data-driven, hybrid IDSs are promising approaches to enhance ICS cybersecurity and defenders situational awareness. A recent survey of cybersecurity research using data mining and machine learning algorithms identified the following methods as effective in cyber-attack detection: clustering, decision tree (DT), genetic algorithms, naïve Bayes, support vector machine, neural network, and random forest (RF) [10].

IDSs can also be categorized according to network-based or host-based approaches. A network-based IDS monitors network traffic, while a host-based IDS monitors a specific host's process activities. However, IDSs that consider only network and host data may fail to detect sophisticated attacks and insider exploits.

Deep-packet-inspection-based IDSs review commands to alert when there is a malicious command, such as "stop" or "close" when such a command is not expected; however, this approach fails when the attack does not rely on sensitive commands [13]. State-based IDSs define secure states and critical states for a process and detect a cyber attack by comparing the current state with the critical state database [14]. This type of IDS requires a detailed analysis of the process to identify all critical states, which lacks generality and may fail when zero-day attacks exploit previously unknown critical states. If the network-based IDS did not detect the attacker when the attacker gained access to the system and the host-based IDS did not detect the intrusion while the attacker is in the system, the attacker could launch persistent command or process data tampering attacks that affect the physical system. With rapidly evolving cyber attacks, these network and host data-based methods are inadequate. Unsupervised models that incorporate process data provide complementary monitoring without relying on detailed knowledge of the exploit, potentially giving defenders time to take action before the attack leads to unrecoverable physical damage.

There is a limited research done in regard to cybersecurity using process data. Gawand, Bhattacharjee, and Roy developed an effective monitoring approach for detecting cyber attacks based on nuclear power plant process data using least square approximation and computational geometric approach [15]. Li and Huang proposed the concept of cyber-attack detection model using dynamic principal component analysis on process data [16].

Empirical monitoring models have been proven as an effective approach for detecting early-stage equipment degradation in industry [17]. This data-driven method may provide a complementary approach of early ICS cyber-attack detection for a variety of attack types that impact the physical system operation. Since the SCADA system already collects data, the cyber-attack analysis does not impact its normal operations.

Several ICS testbeds have been developed for cybersecurity research [18]. One such testbed consists of a physical two-loop forced flow experiment facility with controlled variables, a LabVIEW-based SCADA system, and an attacker computer running Kali Linux on the same local area network [19]. This system is capable of collecting network data, host system data, and process data. Five attack scenarios, including reconnaissance, DoS attacks, and a data tampering attack, have been conducted to demonstrate the possibility of cyber attacks and to generate data for studying IDS development.

A cyber-attack detection system utilizing a defense-in-depth concept to enhance ICS cybersecurity is presented in this paper. The proposed multilayer IDS improves overall cybersecurity by combining signature-based and anomaly-based analysis of network, host, and process data. The main contribution and innovation of this paper is that it integrates network traffic data, host system data, and process data into one system to provide multiple layers of cyber detection. The proposed system applies supervised and unsupervised models for network and system data as two barriers after firewall failure, and unsupervised model for industrial process data as the last cyber-attack defense line. In this paper, k-nearest neighbors (KNN), DT, bootstrap aggregating (bagging), and RF are investigated to differentiate normal operation and cyber-attack scenarios using network and host system data. AAKR is investigated with physical process data to detect attack scenarios that were undetected by the network and system data detection. The results suggest that these models are effective in detecting a variety of cyber attacks with high accuracy.

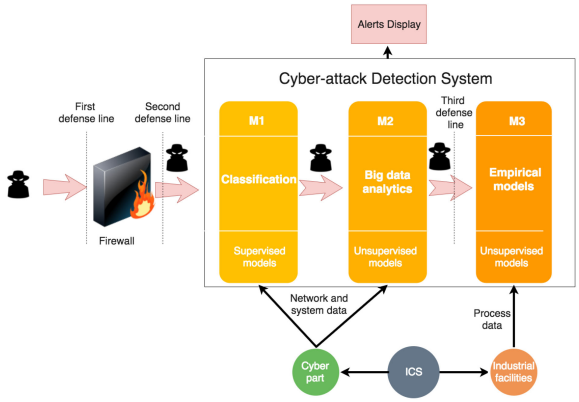

Fig. 2. Multiple-layer cyber-attack detection system.

III. Methodology

Fig. 2 shows the structure of the proposed cyber-attack detection system with a defense-in-depth concept. The first defense layer is the traditional intrusion prevention layer, including firewalls, data diodes, and gateways, which are already widely applied in the industry. However, there are situations that the attackers could bypass this defense line. The second defense layer consists of data-driven models for cyber-attack detection based on network traffic and system data, including the classification model indicated by M1 and big data analytics models indicated by M2. The classification models are based on supervised learning techniques, which can only detect attacks with behaviors similar to known attacks. Unsupervised big data analytics-based models will provide additional flexibility for intrusion detection; this is an area of ongoing research. M1 and M2 provide early detection of attackers when the attacks cause behavior deviation from normal operation. If the secondary layer fails to detect malicious activities, the last defense line monitors process data and uses empirical models indicated by M3 to detect abnormal operation, potentially due to cyber attack. This multilayer detection system improves the robustness of overall intrusion detection and is sensitive to both known and zero-day exploits.

A variety of classification techniques have been investigated on network traffic and host system data, which is shown as M1. Physical process data were analyzed using AAKR model, which is part of the M3. M2, the big data analytics models, are under development. The modeling algorithms employed are briefly discussed in the following subsections.

A. Cyber-Attack Detection Models Based on Network Traffic Data and Host System Data

All the features available in the Windows Performance Monitor were collected to identify which features may be most

useful for data-driven intrusion detection. On average, there are 47 000 features collected; the exact number of features depends on the counters configured in the Windows Performance Monitor. The data preprocessing procedure including removal of features with a large percentage of not-a-number values and near-constant values; these features typically do not convey useful information. Features related to software that is not directly relevant to the cyber operation and attacks in this research were also removed; these features do not statistically vary as a result of the postulated attack scenarios. A total of 142 features related to memory, computer process, and network behavior that contribute useful information for attack detection were selected for the final dataset. The dataset used for detection is collected under three cyber attacks: MITM, DoS attack to engineering workstation, and DoS attack to the National Instruments cDAQ (the data acquisition and control hardware). The observations collected while the system is under attack were labeled as 1, whereas the observations under normal condition were labeled as 0, indicating there is no cyber attack. The dataset was divided into training data and test data for training the classifier and for testing the performance, respectively.

KNN [20], DT [21], bagging [22], and RF [23] are four classification methods investigated in this research. These four approaches were selected based on the suggestion from the survey of the data mining methods for cybersecurity intrusion detection [10]. For brevity, the theoretical derivations of these four algorithms are not detailed here; the interested reader is referred to the seminal references given for a complete treatment of the algorithm. KNN classification is a simple technique that classifies an object by a majority vote of its neighbors. The object is assigned to the most common class among its KNN. A DT is a tree-like structure which represents classifiers and branches. The internal node represents a test for an attribute; the branch represents the test result; the leaf node represents a classification decision after considering all attributes [10]. For a given problem, KNN and DT search the hypothesis space to determine a hypothesis that makes good predictions. However, identifying a good hypothesis may be nontrivial. In contrast to that, an ensemble method is able to combine the predictions from multiple machine learning algorithms together to determine a better hypothesis than the best one alone. Bagging and RF are two ensemble methods combine several DT which have been applied to intrusion detection in former research [10].

Bagging reduces the variance of a DT by averaging votes when predicting a class. It creates m subsets with n samples per subset from the original dataset. These n individual samples are generated from the original dataset by uniformly sampling with replacement [22]. RF consists of many DT based on randomly picked subsets. RF yields the predictions from all of the subtrees, which have less correlation than bagging by constraining only one predictor out of a subset to be used in the split in a tree [23]. The prediction result is determined by majority voting or weighted voting. A single DT splits by selecting single features (variables), while RF splits by selecting multiple features at each split point.

Fig. 3 shows the procedure of analyzing data using these four different classification methods. The data were shuffled and randomly divided into 50% training data and 50% test data. The four classification models are trained using training data. Hyperparameters (e.g., the number of trees in the DT classification) for each model were optimized to obtain the lowest misclassification rate. After that, each model was evaluated using the test data. Comparing the predicted class and actual class, the accuracy of the model was obtained. The accuracy assessment considers true positive (correct detection), false positive (incorrect detection), true negative (correct normal), and false negative (missed detection) rates.

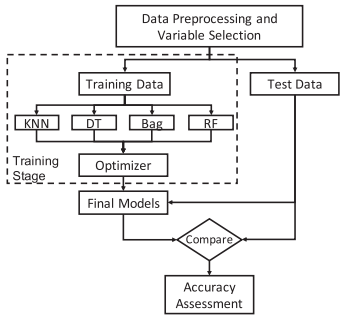

Fig. 3. Procedure for analyzing network and system data.

B. Cyber-Attack Detection Model Based on AAKR Using Process Data

AAKR is a memory-based, nonparametric, unsupervised error-correction algorithm; predictions are the expected values of sensed data under normal operating conditions. These predictions are calculated as a weighted average of historic memory vectors based on the distance between a query vector and the memory vectors, where the query vector is the observation under evaluation and the memory vectors are observations of past, error-free operation retained in a memory matrix. The weights are calculated using a kernel function, such as the Gaussian kernel in (1) [24].

$$w_i = e^{-\frac{d_i^2}{h^2}} \tag{1}$$

where the w_i is weight of the ith memory vector, d_i is the Euclidean distance between the query vector and the ith memory vector, and h is the bandwidth of the kernel function. The AAKR model was optimized by a brute-force search to select the number of memory vectors retained in the memory matrix and the bandwidth. The model with the lowest root mean square error for the intrusion-free test data was selected as the optimal model.

AAKR relies on the relationships between variables to perform error correction. Groups of sensors with Pearsons correlation coefficient magnitude greater than 0.3 are considered sufficiently well-correlated sensors to be included in a single AAKR model.

In this research, analysis of the physical process data is employed in the event that an attacker has made a malicious change to the system, the effects of which are expected to meet this persistence requirement.

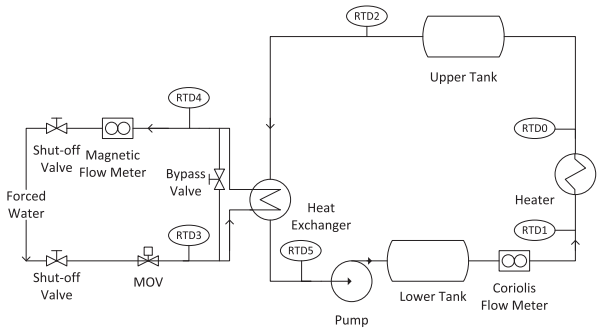

Fig. 4. Two-loop forced flow loop system layout [19].

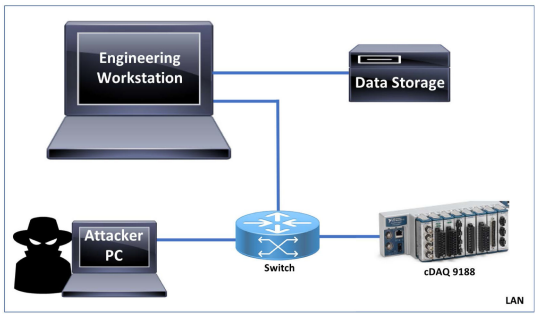

Fig. 5. Testbed SCADA system and other cyber components.

IV. Experimental Setup

A. Testbed Architecture

The ICS testbed includes a physical process system, a SCADA system and the cyber components to simulate the full CPS. Fig. 4 shows the physical part of the ICS testbed, which notionally simulates a two-loop nuclear power system. The primary loop (shown on the right) includes an adjustable 9 kW heater, a variable speed coolant pump, upper and lower delay tanks, and instrumentation, including a Coriolis flow meter and four resistance temperature detectors (RTDs). The heater represents the reactor core, whose power is controlled by SCADA master through an open loop controller. The pump represents the main coolant pump, whose speed is controlled by a variable frequency drive (VFD) and traditional proportional-integral (PI) control to maintain system setpoints. The secondary loop (shown on the left) contains a motor operated valve (MOV), a bypass valve, a magnetic flow meter, and two RTDs. The shell-and-tube heat exchanger (HX) thermally connects the primary loop to secondary loop. The MOV is controlled by a PI controller to maintain system setpoints by adjusting the secondary side flowrate. The control aim is to maintain the heater inlet and outlet temperatures (measured by RTD1 and RTD0) at their set points (26 and 30 °C, respectively) [19].

Fig. 5 shows the SCADA system and other cyber components, including data storage and an attacker computer, of the ICS testbed. The SCADA system consists of an engineering workstation as the SCADA master and a National Instruments cDAQ9188 chassis (referred to as cDAQ) with various data collection and control signal output modules as the SCADA slave. LabVIEW has been installed on the engineering workstation to record sensor data and send control commands to actuators. The cDAQ is the interface between the physical and cyber aspects of the system. Kali Linux operation system is used in the attacker PC, which has tools capable of launching different attacks, such as DoS and MITM [25]. Data storage is hosted on servers, which are connected directly to the engineering workstation to store sensor data and system log files.

B. Data Collection

The network traffic data and host system data were collected by the Windows Performance Monitor with the sample rate of one observation every 1 s, and process data were collected by LabVIEW with the sample rate of one observation every 8 s. To investigate data-driven models, normal behavior data were collected while the system operates under normal transients without any cyber attack; abnormal data are collected under postulated cyber-attack scenarios. Normal transient operation includes ramps in heater power between 100% and 50%, during which the process control system acts to maintain system setpoints. These operational transients are used to generate normal traffic between the engineering workstation and the cDAQ. Five attack scenarios are conducted to demonstrate the cyber attack to ICS and to generate data for developing an effective cyber-attack detection system.

C. Cyber-Attack Scenarios

Using the described two-loop cyber-physical testbed, five cyber-attack scenarios are conducted, namely packet sniffing using MITM, DoS, data exfiltration, false data injection and tampering, and simultaneous cyber attacks which leads to small loss of coolant accident (LOCA).

Cyber-attack scenario A is packet sniffing using MITM attack, which leverages a weakness in the address resolution protocol (ARP) to listen to the communication between two parties without their knowledge and potentially modify it [26]. It uses the lack of authentication in ARP to manipulate the ARP tables of the victims to reroute packets through attacker PC. Under this scenario, the attacker is able to obtain the traffic packets between the engineering workstation and the cDAQ. Then, the attacker may modify the packets to cause unintended actions in physical process.

Cyber-attack scenario B is a DoS attack against the cDAQ using a spoofed IP address. The DoS attack floods cDAQ with rapid superfluous requests to exhaust its capacity and render it unavailable, which prevents legitimate requests from executing, including those to collect data and to send control signals [27].

Cyber-attack scenario C is data exfiltration where the attacker steals and transmits critical information from the engineering workstation. For example, the attacker transmits the LabVIEW model or data collected from cDAQ via a server–client protocol, which is established between the engineering workstation and an external malicious server by the attacker. Using phishing and social engineering attack, the engineer workstation is compromised and a malicious client script is installed. A malicious server is deployed off the premises by the attacker that listens to any attempts to connect by a client. The engineering workstation

connects to the server via the existing network infrastructure. Critical operational data are then transmitted through the network periodically and silently as they become available.

Cyber-attack scenario D is tampering with command data and false data injection by an insider (such as a disgruntled employee) or an external attacker. An external attacker can use a spear phishing technique to deceive someone, referred to as the carrier, who has access to the intranet of the power plant or other high-value CPS to utilize as a vessel to bypass the air gap between the outside world and the intranet; for example, the carrier could be a subcontractor who is responsible for keeping the software up to date of the intranet computers. The carrier inadvertently downloads a virus by clicking a link sent by a specially crafted email by the attacker in the spear phishing attack. This virus has the capability to obfuscate its code to not trigger antivirus software in the carriers laptop. Once executed, it will change the system hosts configuration to point to a malicious update website that is designed to mimic the authentic website of the software vendor. It also updates the certificate authorities to accept a faux HTTPS certificate. When the carrier visits the updated website to download the updates, he is redirected to the attackers mock website with the malicious program, which will be downloaded to his laptop. Since the host configuration and certificate authorities are altered, it gives no warnings about the malicious download.

When the carrier visits the power plant to update the software, he executes the malicious software update. Even though it is run as a standard user, it has an exploit to leverage a known and unpatched vulnerability of the operating system to escalate its privileges. Using these higher privileges, it sets the ethernet interface of the laptop to the promiscuous mode, which enables the interface to see all the network packets regardless of the intended recipient. It performs a host scan to identify all the hosts that are up in the network including their interface MAC addresses. Using a vendor MAC address look-up table embedded with the malicious software update, it identifies the cDAQ and its IP address. Then, it launches an MITM attack to redirect the traffic between all other available hosts and the cDAQ through ARP poisoning.

Then, the external attacker listens to the connection between the engineering workstation and cDAQ long enough to identify which command it needs to alter and then modifies the commands to achieve its malicious goals. In this research, the false command injection first brings the heater power to 0 kW, which simulates a forced reactor scram. Then, the attacker brings the heater power to 6.3 kW and alters the core outlet temperature set point from 30 to 28.4 °C, which aims at accelerating the degradation or damage of critical assets.

Cyber-attack scenario E is an extended attack of scenario D to give a demonstration of simultaneous cyber attacks which leads to small LOCA. In this attack, the attacker stops the pump and closes the MOV, resulting in no flow in both the primary and secondary loop, while the heater power remains at 6.3 kW. Because there is no flow through the system, the heat cannot be removed from the primary coolant, eventually causing the joints of the pipes to expand due to overheating and resulting in a small LOCA.

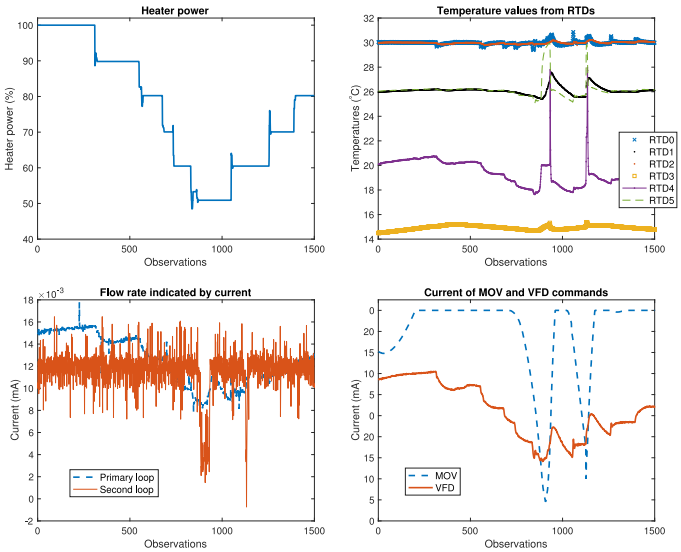

Fig. 6. Process data during normal operational transients.

V. Results

The ICS testbed was operated under normal operational transients and postulated cyber-attack scenarios to collect network traffic, host system data, and process data to train the four classification models and the AAKR process monitoring model and to test the cyber-attack detection framework effectiveness. This section shows that the combination of the second and third layers in the proposed cyber-attack detection system can provide robust cyber-attack detection coverage and functionality.

A. Normal Operational Transients

Fig. 6 shows the various process data under normal operational transients with changing heater power. The top-left subplot shows the heater power level profile specified by the operator, stepping from 100% nominal power to 50% and back to 100% in 10% power increments. The top-right subplot shows the temperatures in different locations, as indicated in Fig. 4. RTD1 and RTD0 are core inlet and outlet temperature, which are maintained at set points of 26 and 30 °C, respectively. RTD2 is located between the outlet of upper delay tank and inlet of HX primary side inlet; it follows the temperature of core outlet (RTD0) with a short time delay. RTD5 is located between the HX primary side outlet and lower delay tank inlet; it is followed by the temperature of core inlet (RTD1) with a short time delay. The water in HX secondary side is forced water from the building system. RTD3 measures the HX secondary side inlet temperature coming from the building supply; the RTD3 temperature is not controllable by the physical system and fluctuates according to the other activities in the building. RTD4 shows the HX secondary outlet temperature, which depends on the temperature measured by RTD3 as well as the heat removed from the primary side coolant to maintain heater inlet and outlet temperatures at their set points. The bottom-right subplot shows the control commands sent to the MOV and VFD determined by the PI controllers. The bottom-left subplot shows the primary and secondary flow rate meter currents; the magnetic flow

TABLE I
PERFORMANCE COMPARISON OF KNN, DECISION TREE, BAGGING, AND RANDOM FOREST

Classification Methods	True Positive	False Negative	False Positive	True Negative
KNN	98.84%	1.16%	0.54%	99.46%
Decision tree	94.80%	5.20%	1.25%	98.75%
Bagging	98.27%	1.73%	0.0%	100.0%
Random forest	97.69%	2.31%	0.0%	100.0%

TABLE II
COMPUTING COST COMPARISON OF KNN, DECISION TREE, BAGGING, AND RANDOM FOREST

Classification Method	User time (ms)
KNN	0.1558
Decision tree	0.0473
Bagging	0.0612
Random forest	0.0779

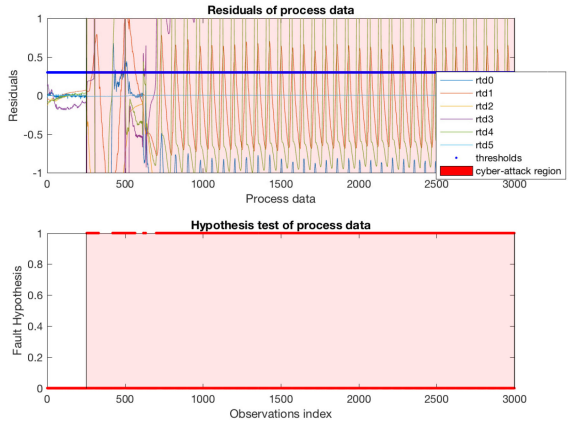

Fig. 7. Intrusion detection through simple thresholding of process monitoring system residuals.

meter on the secondary side has significantly more noise than the Coriolis meter on the primary side due to the measurement modality used.

The results show that temperatures of RTD0 and RTD1 are maintained at their set points, which demonstrates that under the normal operation transients, the functionality of SCADA system meets expectation.

B. Cyber-Attack Scenarios and Detection

Network data and host system data under the MITM and DoS attacks were analyzed using the four classification methods. The cyber-attack dataset contains 1438 observations, 179 observations of which are MITM attacks, 53 observations are DoS attack to engineering workstation, and 105 observations are DoS attack to cDAQ. The remaining 76.36% of observations are normal operation between specific attacks. The performances of the four models are shown in Table I. KNN gave the best true positive rate of 98.84%, while the DT gave the lowest true positive rate of 94.8%. The bagging and RF yielded a false positive rate of zero. Therefore, these two methods are very suitable for the application requiring false alarms be as low as possible. False negative rate results show that the KNN yielded the lowest missed alarm rate. The DT yielded the poorest performance with lowest true positive rate, highest false negative rate, and highest false positive rate. The improved performance of bagging and RF over DT is expected since these methods are techniques to improve accuracy based on DT.

In addition to performance, the computing cost is another important issue in cyber-attack detection. Table II shows computing costs (shown in milliseconds) of four models, which was measured by user time for R code to determine whether a signal observation is from attack scenarios or not. R version 3.4.3 is used to do the classification on a MacBook Air with dual core 1.7 GHz with 8 GB memory. All other active programs were terminated prior to the execution. Among the four methods investigated, the DT has the lowest computing cost in this case. The classification speed of bagging, RF, and KNN is much faster than the sample rate of the network data in this application; their computing costs are sufficient.

Under attack scenario D, at about 600 observations, the attacker changed the heater power to 70% and changed the core outlet temperature set point to 28.4 °C. The PI controller proportional and integral gains were tuned to control the core outlet temperature at 30 °C; once the set point changed, the PI controller did not function well. This led to cycling of both the MOV valve and the VFD frequency and resulted in unstable flowrates and temperatures. This fairly simple change to the system could allow the attacker to accelerate degradation or damage of critical assets.

The AAKR model was applied to the process data under this attack scenario. Nine highly intercorrelated process sensors are selected as the inputs for the AAKR monitoring model in this study: primary flowrate (Coriolis flow meter); temperatures measured by RTD0, RTD1, RTD2, RTD4, and RTD5; and control signals given by VFD frequency, MOV current, and heater power level. HX secondary side inlet temperature (RTD3) is not sufficiently correlated to the other signals, because it is nearly constant depending on the building supply water temperature. Fig. 7 shows the residuals of the temperature values of RTD0-5 between the monitored values and the predicted values in the top subplot and the fault hypothesis in the bottom subplot. In the top subplot, the threshold of the residuals is 0.3, indicated by the blue line; this threshold was selected based on the monitoring system residuals seen under normal operation. In the bottom subplot, the resulting alarms detected by simple thresholding are shown, where an intrusion is detected if any one residual crosses the detection threshold. The shaded red area indicates the time the system is under cyber attack. The first 250 observations are under normal operation; at observation 251, the attacker built the connection to cDAQ and changed the heater power to zero. At observation 600, attacker tampered with the core outlet set point and injected the command to change power to 70% of nominal power, which is within normal operational parameters. The process monitoring results suggest that the proposed cyber-attack detection based on process data can effectively detect attacks prior to significant physical effects.

VI. Conclusion

CPSs are able to improve both productivity and economic efficiency, resulting in rapid and widespread utilization growth. However, the evolving cyber threats present new challenges to CPS cybersecurity. Because of its critical nature, ICS cybersecurity needs to be enhanced. Most current cybersecurity efforts in industry may be not capable of detecting advanced cyber attacks before significant damage is done to the physical system. To address this concern, a multilayer, data-driven cyber-attack system was developed to enhance ICS cybersecurity by providing wider attack detection coverage by applying the defense-in-depth concept. In the proposed cyber-attack detection system, the first defense line contains firewalls and data diodes; the second defense line includes supervised classification models indicated by M1 and unsupervised big data analytics models indicated by M2 based on network traffic and system data; the last defense line uses unsupervised empirical models indicated by M3 based on physical process data for cyber-attack detection before significant physical consequences occur. In this research, four classification models were evaluated for M1 intrusion detection and an AAKR model with residual thresholding detection was implemented in the M3 defense layer. The detection results of M1 and M3 using data generated from the physical testbed show that the proposed cyber-attack detection system has a high detection accuracy and a wide attack coverage. In order to detect unknown attacks using network and host system data, the unsupervised big data analytics models in M2 will be studied to further enhance the second defense line.

Acknowledgment

The authors acknowledge the Lloyd's Register Foundation and the International Joint Research Center for the Safety of Nuclear Energy for partial funding of this research. Lloyd's Register Foundation helps to protect life and property by supporting engineering-related education, public engagement, and the application of research. The authors also acknowledge Oracle Corporation for partial funding of this research.

References

[1] Y. Ashibani and Q. H. Mahmoud, "Cyber physical systems security: Analysis, challenges and solutions," *Comput. Secur.*, vol. 68, pp. 81–97, 2017.

[2] R. Langner, "Stuxnet: Dissecting a cyberwarfare weapon," *IEEE Secur. Privacy*, vol. 9, no. 3, pp. 49–51, May/Jun. 2011.

[3] ICS-CERT, "Cyber-attack against ukrainian critical infrastructure," Feb. 2016. [Online]. Available: https://ics-cert.us-cert.gov/alerts/IR-ALERT-H-16-056-01

[4] ICS-CERT, "Hatman—Safety system targeted malware," Mar. 2017. [Online]. Available: https://ics-cert.us-cert.gov/MAR-17-352-01-HatMan-Targeted-Malware

[5] Kaspersky Lab ICS-CERT, "Threat landscape for industrial automation systems in h2 2017," Mar. 2018. [Online]. Available: https://ics-cert.kaspersky.com/reports/2018/03/26/threat-landscape-for-industrial-automation-systems-in-h2-2017/

[6] N. S. Malik, R. Collins, and M. Vamburkar, "Cyber-attack, pings data systems of at least four gas networks," Apr. 2018. [Online]. Available: https://www.bloomberg.com/news/articles/2018-04-03/day-after-cyber-atta ck-a-third-gas-pipeline-data-system-shuts

[7] Homeland Security Centre for the Protection of National Infrastructure, "Cyber security assessments of industrial control systems," Apr. 2011. [Online]. Available: https://www.ccn-cert.cni.es/publico/InfraestructurasCriticaspublico/CPNI-Guia-SCI.pdf

[8] G. Loukas, *Cyber-Physical Attack Steps*. London, U.K.: Butterworth-Heinemann, 2015, ch. 5, pp. 145–180.

[9] S. Han, M. Xie, H.-H. Chen, and Y. Ling, "Intrusion detection in cyber-physical systems: Techniques and challenges," *IEEE Syst. J.*, vol. 8, no. 4, pp. 1052–1062, Dec. 2014.

[10] A. L. Buczak and E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection," *IEEE Commun. Surv. Tut.*, vol. 18, no. 2, pp. 1153–1176, Apr.–Jun. 2016.

[11] D. Yang, A. Usynin, and J. W. Hines, "Anomaly-based intrusion detection for SCADA systems," in *Proc. 5th Nucl. Plant Instrum. Control and Human, Mach. Interface Technol.*, Albuquerque, NM, Nov. 2006, pp. 12–16.

[12] B. Jeffries, J. W. Hines, and K. C. Gross, "Behavior-based approach to misuse detection of a simulated SCADA system," in *Proc. 10th Nucl. Plant Instrum. Control and Human, Mach. Interface Technol.*, San Francisco, CA, USA, Jun. 2017, pp. 1761–1771.

[13] W. Gao and T. H. Morris, "On cyber attacks and signature based intrusion detection for MODBUS based industrial control systems," *J. Digit. Forensics, Secur. Law*, vol. 9, no. 1, 2014, Art. no. 3.

[14] A. Carcano, I. N. Fovino, M. Masera, and A. Trombetta, "State-based network intrusion detection systems for SCADA protocols: A proof of concept," in *Proc. Int. Workshop Crit. Inf. Infrastructures Secur.*, 2009, pp. 138–150.

[15] H. L. Gawand, A. Bhattacharjee, and K. Roy, "Securing a cyber physical system in nuclear power plants using least square approximation and computational geometric approach," *Nucl. Eng. Technol.*, vol. 49, no. 3, pp. 484–494, 2017.

[16] J. Li and X. Huang, "Cyber attack detection of I&C systems in NPPS based on physical process data," in *Proc. 24th Int. Conf. Nucl. Eng.*, Charlotte, NC, Jun. 2016, pp. V002T07A011; 4 pages, Paper No. ICONE24-60773.

[17] J. Coble, P. Ramuhalli, L. J. Bond, J. Hines, and B. Upadhyaya, "A review of prognostics and health management applications in nuclear power plants," *Int. J. Prognostics Health Manage.*, vol. 6, 2015, Art. no. 016.

[18] H. Holm, M. Karresand, A. Vidström, and E. Westring, "A survey of industrial control system testbeds," in *Secure IT Systems*. New York, NY, USA: Springer, 2015, pp. 11–26.

[19] F. Zhang, J. W. Hines, and J. B. Coble, "Industrial control system testbed for cybersecurity research with industrial process data," in *Proc. Int. Congr. Adv. Nucl. Power Plants*, Charlotte, NC, Apr. 2018, pp. 279–284.

[20] B. K. Samanthula, Y. Elmehdwi, and W. Jiang, "K-nearest neighbor classification over semantically secure encrypted relational data," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 5, pp. 1261–1273, May 2015.

[21] Y. Zhang *et al.*, "Comparison of machine learning methods for stationary wavelet entropy-based multiple sclerosis detection: Decision tree, k-nearest neighbors, and support vector machine," *Simulation*, vol. 92, no. 9, pp. 861–871, 2016.

[22] B. Wang and J. Pineau, "Online bagging and boosting for imbalanced data streams," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 12, pp. 3353–3366, Dec. 2016.

[23] M. Belgiu and L. Drăguţ, "Random forest in remote sensing: A review of applications and future directions," *ISPRS J. Photogrammetry Remote Sens.*, vol. 114, pp. 24–31, 2016.

[24] F. Zhang, S. Boring, J. W. Hines, J. B. Coble, and K. C. Gross, "Combination of unquantization technique and empirical modeling for industrial applications," in *Proc. Amer. Nucl. Soc. Winter Meeting*, Washington, D.C., USA, Nov. 2017, pp. 449–452.

[25] L. Allen, T. Heriyanto, and S. Ali, *Kali Linux–Assuring Security by Penetration Testing*. Birmingham, U.K.: Packt Publishing Ltd., 2014.

[26] M. Conti, N. Dragoni, and V. Lesyk, "A survey of man in the middle attacks," *IEEE Commun. Surv. Tut.*, vol. 18, no. 3, pp. 2027–2051, Jul.–Sep. 2016.

[27] K. K. Oo, K. Z. Ye, H. Tun, K. Z. Lin, and E. Portnov, "Enhancement of preventing application layer based on DDOS attacks by using hidden semi-Markov model," in *Genetic and Evolutionary Computing*. Cham, Switzerland: Springer, 2016, pp. 125–135.

Authors' photographs and biographies not available at the time of publication.