Raman M. R. *et al. Cybersecurity* (2021) 4:27
https://doi.org/10.1186/s42400-021-00095-5

Cybersecurity

RESEARCH Open Access

Machine learning for intrusion detection in industrial control systems: challenges and lessons from experimental evaluation

Gauthama Raman M. R.[1*], Chuadhry Mujeeb Ahmed[2] and Aditya Mathur[3]

Abstract

Gradual increase in the number of successful attacks against Industrial Control Systems (ICS) has led to an urgent need to create defense mechanisms for accurate and timely detection of the resulting process anomalies. Towards this end, a class of anomaly detectors, created using data-centric approaches, are gaining attention. Using machine learning algorithms such approaches can automatically learn the process dynamics and control strategies deployed in an ICS. The use of these approaches leads to relatively easier and faster creation of anomaly detectors compared to the use of design-centric approaches that are based on plant physics and design. Despite the advantages, there exist significant challenges and implementation issues in the creation and deployment of detectors generated using machine learning for city-scale plants. In this work, we enumerate and discuss such challenges. Also presented is a series of lessons learned in our attempt to meet these challenges in an operational plant.

Keywords: Industrial control systems, ICS security, Machine learning, Intrusion detection, Testbed and experimental Study

Introduction

Industrial Control Systems (ICS) are part of modern Critical Infrastructures(CI) such as water treatment plants, oil refineries, power grids, and nuclear and thermal power plants. ICS refers to a system obtained by integrating computing and communication components to control a physical process. An ICS consists of devices and subsystems such as sensors, actuators, Programmable Logic Controllers (PLCs), Human Machine Interfaces (HMIs), and a Supervisory Control and Data Acquisition (SCADA) system.

An abstract view of an ICS is shown in Fig. 1. The field devices, i.e., sensors and actuators in the physical layer, monitor and regulate the underlying industrial process. The current state of the process is sampled through sensors and communicated to the PLCs in the distributed control layer. The PLCs generate control actions and transfer the actions to the actuators such as pumps, valves, generators, and circuit breakers. Other devices, such as the SCADA and HMIs in the supervisory control layer, enable communication between a plant operator and the PLCs for implementing human-assisted control actions. Recent studies indicate that a majority of ICS systems use proprietary communication protocols (Drias et al. 2015; Feng et al. 2016; Mirian et al. 2016).

The deployment of ICS in Critical Infrastructure (CI) makes them an attractive target for the adversaries. A skilled adversary could interfere with any of these systems to manipulate over time the sensor readings or actuator controls until their malicious intent is realized. Past incidents, such as the Stuxnet worm (Langner 2011) and Blackenergy (Case 2016), indicate that targeted attacks are possible in practice and may remain undetected for long (Ahmed and Zhou 2020). Such persistent threats have led to research in the development of defense-in-depth mechanisms for detection and mitigation. One class

*Correspondence: gauthama_mani@sutd.edu.sg
[1]iTrust, Singapore University of Technology and Design (SUTD), 8 Somapah Rd, Singapore 487372, Singapore
Full list of author information is available at the end of the article

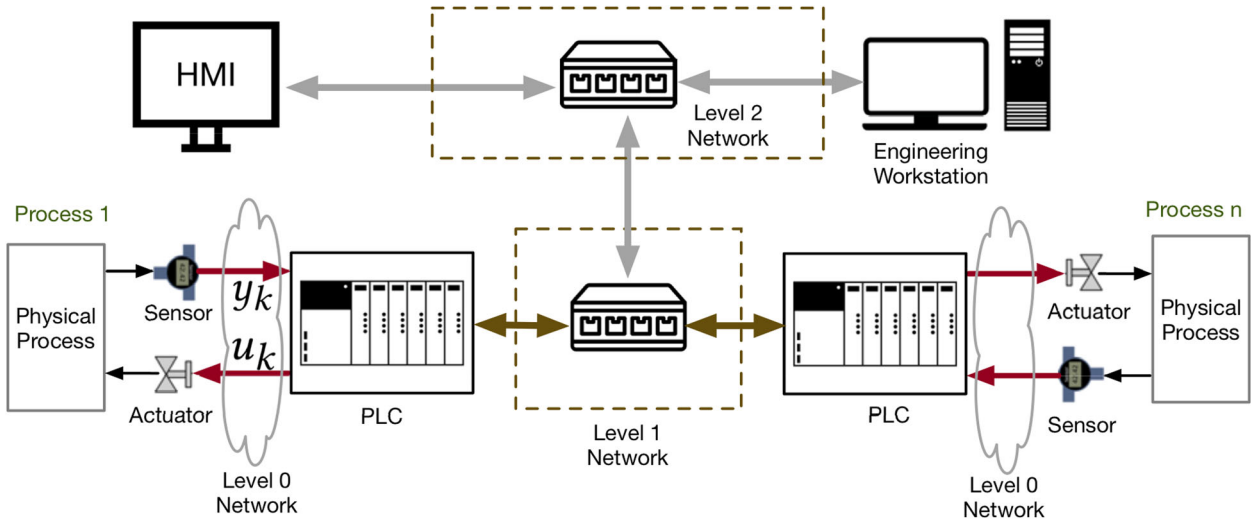

Fig. 1 An abstract view of an Industrial Control System

of such mechanisms is the anomaly detector that raises an alert when the physical process controlled by an ICS moves from its expected behavior to an undesirable state, or anomalous, state. An anomalous state could be due to faulty components, temporal glitches, misconfiguration, cyber-attacks, or a combination thereof. In this work, we assume that the existence of anomalies is due to cyber-attacks.

Approaches used to build anomaly detectors can be broadly categorized as design-centric and data-centric. Design-centric approaches make use of physical relationships, captured as invariants, among the ICS components obtained from the plant design for anomaly detection (Adepu and Mathur 2018; Ahmed et al. 2020). In data-centric approaches such relationships are learned and modeled through the application of machine learning and computational intelligence techniques (Gauthama Raman et al. 2017; Raman et al. 2017; Ahmed et al. 2020; Ahmed et al. 2017). Both approaches come with their pros and cons.

Data-centric approaches are attractive due to their automated feature learning ability through the application of machine learning algorithms. Further, the increasing availability of data and advanced computational resources makes it practical to develop and deploy anomaly detectors so created. Despite the advantages, one faces several challenges while creating and testing such detectors in an operational plant. It is useful to understand and address such challenges before an anomaly detector, designed using a data-centric approach, is deployed in large-scale systems such as a 100 Million Gallons/Day water treatment plant or a distributed power grid.

There exist surveys (Bhamare et al. 2020; Mitchell and Chen 2014; Han et al. 2014) related to the data-centric approaches in ICS security. However, to the best of our knowledge, this paper is the first of its kind that discusses the challenges one faces in the design and deployment of real-time ML-based anomaly detectors in operational city-scale plants. The learning approaches used in the design of the anomaly detectors are characterized here as data-based and behavioral-based. Data-based learning deploys the direct application of machine learning algorithms on the data collected from the operational plant for anomaly detection. In addition to using the data, a behavioral-based learning approach incorporates prior knowledge of the plant to detect anomalies. The primary focus of this article is to discuss the challenges related to the applicability of both approaches for anomaly detection in an ICS. Practical solutions are proposed to overcome these challenges that might be useful for researchers and practitioners. A related prior work (Ahmed et al. 2020) presented preliminary data collected from a water treatment testbed. This article examines a real-world city-scale water system and highlights additional challenges offering insights into what is required to scale a data-centric solution to a real critical infrastructure. Possible solutions to meet such challenges are proposed and supported by numerical results. A key contribution of this work is that the experiments are carried out on live ICS, in contrast to the existing studies, which are based on the historical dataset.

Organization: The remainder of this article is organized as follows. In "Materials and methods" section we discuss the difference between the characteristics of an ICS and traditional IT infrastructure followed by a brief introduction to the SWaT plant against which multiple anomaly detectors have been tested. Challenges related to the design of data-based and behavior-based learning

approaches for an operational plant are enumerated and explained in "Challenges in the design of anomaly detectors" section. Research directions aimed at the development of methods to overcome the challenges are summarized in"Future outlook and recommendations" section.

Materials and methods

Characteristics of an ICS

The term "cyber-attack" is not new and there exist plenty of security solutions including firewalls, access control, and encryption techniques, to thwart those. However, these IT-centric solutions, while necessary, are not sufficient to safeguard an ICS. IT-centric solutions are designed to deal only with the security issues found in typical IT systems where the protection of data against theft and manipulation are some of the primary concerns. Such solutions fail to address the vulnerabilities that could be exploited during the interaction of IT systems with physical devices or the environment. A breach of a firewall, for example, could remain undetected while the attacker manipulates process data leading the PLCs to issue undesirable commands and moving the plant into an anomalous state. Therefore, it is necessary to understand, and account for, the key characteristics of ICS while designing an anomaly detector. Below we enumerate a few unique characteristics of and requirements for ICS (Stouffer et al. 2014; Wang et al. 2019).

1. An ICS is often required to operate uninterrupted for long periods without any downtime for activities such as code patching in the controllers. Components in an ICS require deterministic responses with an acceptable level of jitter or delay whereas IT systems can tolerate a higher level of delay in network traffic without noticeable impact on the system performance.
2. The physical process controlled by an ICS is continuous and hence unexpected outage of the systems that monitor and control it is unacceptable. In IT systems, rebooting or temporary shutdown of the systems occurs much more often than it does in a physical plant that provides continuous critical services.
3. In a typical IT system, the CIA triad, i.e., data confidentiality, integrity, and availability, includes the primary concerns to ensure availability. On the contrary, in ICS the CIA triad is instead perceived as AIC wherein priority is accorded to the availability of data followed by integrity and confidentiality. For example, it might be desired to ensure the integrity of sensor data while the confidentiality of data itself might not be a major concern.
4. In an IT system, security focus is on safeguarding the IT assets through which data transfer takes place. In ICS, the primary focus safeguarding the edge clients such as the PLCs, sensors, and actuators.
5. A successful attack on an ICS may have severer impact than one on an IT system. The damage in the physical components of an ICS could lead to service disruption and may even impact human life. In contrast, a successful attack on an IT system generally leads to information loss.
6. The behavior of an IT system is highly uncertain and varied whereas that of ICS components is much more stable and predictive.
7. The payload of ICS data is shorter than IT data due to delay-tolerance requirements. Unlike in an IT system, the data generated in an ICS is highly correlated and obeys the system design specifications.
8. A typical ICS operates in a significantly resource-constrained environment and the usage of third-party applications is restricted. On the contrary, the computational efficiency of the IT system can be frequently updated based on user requirements.
9. Communication protocols used for data transfer among the ICS components are proprietary and different from the well-known protocols used in traditional IT environments (Drias et al. 2015; Feng et al. 2016; Mirian et al. 2016).

SWaT: secure water treatment plant

Most of the experiments reported in this paper were performed on a state of the art testbed referred to as Secure Water Treatment (SWaT) plant (Mathur and Tippenhauer 2016). SWaT has been used extensively by researchers to test defense mechanisms for CI (Goh et al. 2016). A brief introduction is provided in the following to aid in understanding the challenges described in this article.

SWaT is a scaled-down version of a modern water treatment process. It produces 5 gallons/minute of water purified using six stages in SWaT. Each stage is equipped with a set of sensors and actuators. Sensors include level meters, pressure guage, and those to measure water quality parameters such as pH, oxidation-reduction potential, and conductivity. Examples of actuators include motorized valves and electric pumps. A pictorial view of the SWaT testbed is in Fig. 2. A more detailed explanation of the testbed can be found in Mathur and Tippenhauer (2016).

SWaT uses a layered architecture (Williams 1993). As shown in Fig. 1, there are three levels of communications. Level 0 is the field communication network and is composed of field devices, e.g., remote I/O units and communication interfaces to send/receive information to/from PLCs. Using the level 0 network, sensors send the physical process state to the PLCs and in turn, PLCs send the control commands to the actuators. Level 1 is the communication layer where PLCs communicate with each

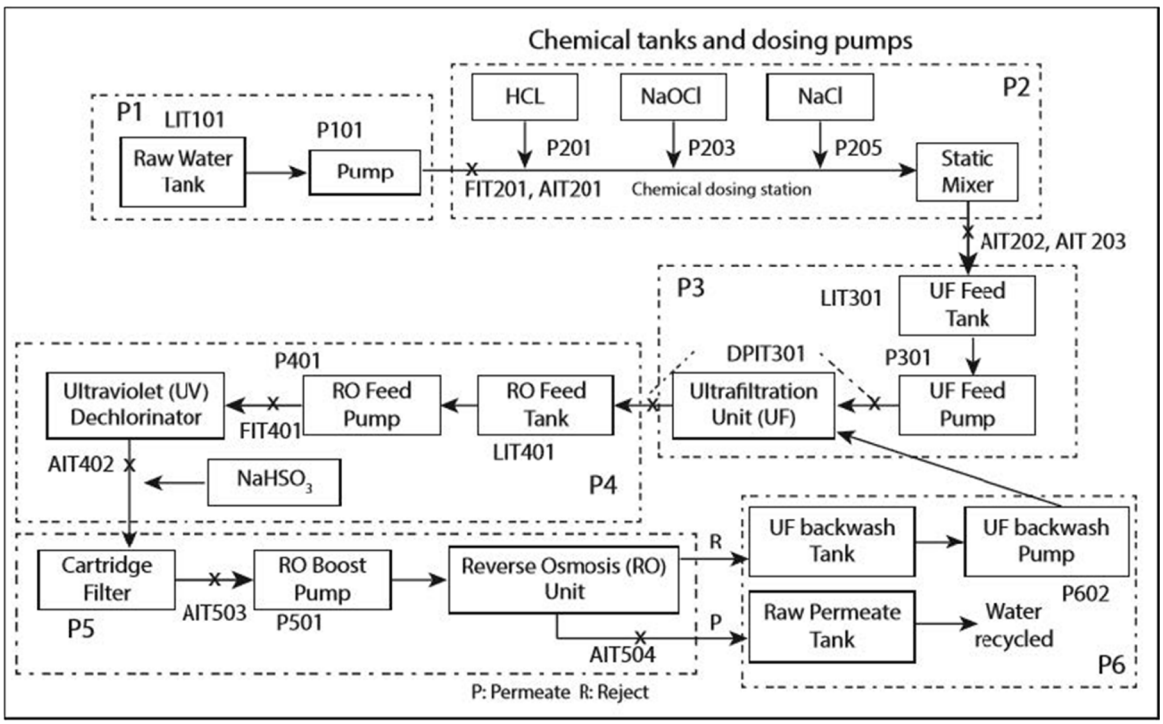

Fig. 2 SWaT testbed processes overview

other to exchange data to make control decisions. Level 2 is where PLCs communicate with the SCADA workstation, HMI, historian server; this is the supervisory control network.

The communication protocols in an ICS have been proprietary until recently when the focus shifted to using the enterprise network technologies for ease of deployment and scalability, such as the Ethernet and TCP/IP. A survey of communication protocols in an ICS can be found in Gaj et al. (2013). In Fig. 1, a specific example of a water treatment testbed used in this study is shown. Common Industrial Protocol (CIP) is used in SWaT. CIP is an application layer protocol on top of Ethernet/IP (ENIP) to exchange data at level 1 and level 2 (Schiffer et al. 2006; Brook 2001). The messages between the devices can use either wired media, i.e. IEEE 802.3, Ethernet, or wireless media i.e. IEEE 802.11 WiFi standard. There are two generic types of messages in the CIP/ENIP standard,i.e., explicit messaging and implicit messaging (Schiffer et al. 2006). Explicit messages use CIP as an application layer protocol and use TCP/IP service to establish a connection. An example is a PLC sending a request message for the exchange of data to another PLC. Implicit messaging, also known as I/O messaging, is used to communicate between PLC and I/O devices.

Anomaly detector for ICS using data-centric approach

An anomaly detector created using a data-centric approach is considered a black box wherein the underlying model is generated using the data received from an operational plant. Based on the characteristics of an ICS as discussed in "Characteristics of an ICS" section, there exist at least two possible ways for the construction of an anomaly detector using data-centric approaches.

1 The physical process controlled by an ICS must remain within the prescribed design limits (Adepu and Mathur 2018). For example, the water level in a tank should not fall below a predefined threshold and, similarly, the ultrafiltration membranes must be cleaned every 30-minutes to maintain the pressure drop across the filtration unit to within safety limits. Thus, a physical model can be created and trained by applying one or more ML techniques to learn the interactions among the feature vectors.
2 Several aspects of the physical process controlled by an ICS are predictive (Wang et al. 2019), e.g., where the sensor measurements are continuous-valued. Thus, the prediction of such sub-processes can be considered as a time series forecasting problem. For such sub-processes, well known ML technique can

be used to learn the behavior of one or more components from historical data and predict their future behavior with minimal forecasting error. Further, using a statistical approach, the discrepancies in the actual and predicted behavior can be analyzed for alert generation.

In both the cases mentioned above, the development of an anomaly detector cycles through three distinct phases, namely, model creation, deployment, and tuning or retraining. Figure 3 illustrates these activities in each stage for the design of an effective anomaly detector. Additional information on these phases is found in Ahmed et al. (2020).

Challenges in the design of anomaly detectors

Next, we enumerate and discuss the challenges faced during the design and deployment of data-based and behavior-based approaches for creating anomaly detectors. The presentation below is organized into tuples as (challenge, positions, lessons). Thus, a challenge is described and the corresponding positions, taken by the authors, discussed. Corresponding to each position the lessons learned from experiments performed on SWaT are enumerated. Note that several lessons here are known to researchers in the machine learning community, and hence not novel. However, the cited experiments offer evidence that will likely enable researchers to decide whether or not to use an approach.

Challenges in data-based learning approaches

Most intrusion detection systems for ICS use the data based learning approaches. These techniques use historical data collected from the operational ICS. Without prior design knowledge, these ML algorithms are trained by fine-tuning the intrinsic parameters to learn the process dynamics of the underlying ICS. Although these approaches offer simplicity in the design and deployment of detectors, they suffer from huge computational complexity due to the existence of heterogeneous components with variable operating ranges. Such complexity leads to the following challenges.

Challenge : Type of machine learning algorithm: Machine learning algorithms can be categorized into the following three types, supervised, semi-supervised, and unsupervised techniques. The type of algorithm utilized to create an anomaly detector depends on, among other factors, data characteristics. There exist several works that cover the use of all the above-mentioned types. The challenges in deploying such techniques in an operational plant are described next.

Position 1: In comparison with the unsupervised learning algorithms, the supervised learning algorithms return a higher detection rate and a lower rate of false alarms.

A Probabilistic Neural Network (PNN) based anomaly detector is presented in Gauthama Raman et al. (2019) for detecting anomalies in SWaT. During the training process, several parameters of PNN are fine-tuned using historical data collected from SWaT. This dataset represents the behavior of the plant under normal and attack scenarios. During testing, the overall detection rate for known attacks was greater than 99.93% and no false alarms were raised. However, the detector was unable to detect novel attacks, i.e., for which signatures were not in the training dataset.

Lessons learned:

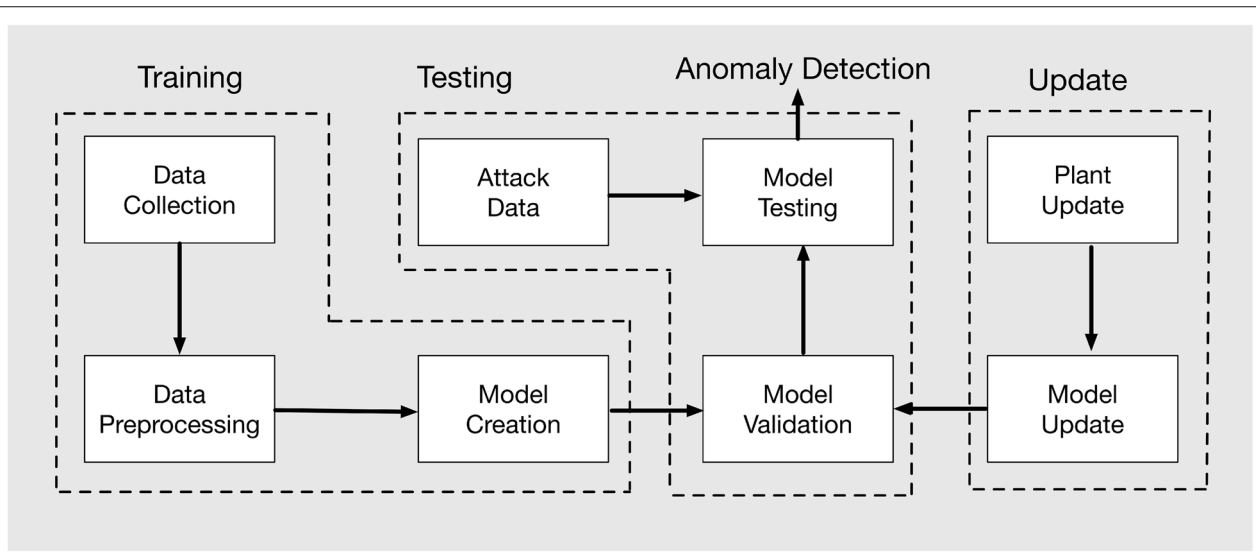

Fig. 3 Stages in the development of an anomaly detector using methods from machine learning: model creation (training phase), model deployment (testing phase), and retraining (update phase)

1 Supervised learning algorithms are blind to zero-day vulnerabilities.
2 The generation of attack signatures for many, if not all, possible combinations of ICS components and testing against them is practically infeasible.

Position 2: Unsupervised learning algorithms possess the ability to detect attacks that exploit zero-day vulnerabilities.

Continuously operational plants offer the luxury of large amounts of process data. Such data capture the behavior of, and interactions among, the ICS components during normal operation. Thus, the design of an anomaly detector can be considered as a "One-Class Classification problem." This aspect of ICS enables the effective use of several unsupervised machine learning algorithms. Hence, a boundary region, corresponding to the normal operation of the plant, can be constructed through the concept of feature learning and behavior that lies outside the boundary can be declared as an anomaly. We have experimented with several unsupervised learning algorithms including One-Class Support Vector Machine (OCSVM), Isolation Forest (IF), K-Nearest Neighbour (K-NN), and Principle Component Analysis (PCA) on data collected from SWaT. Data in Table 1 indicates that these algorithms suffer from an unacceptable number of false positives due to the multi-variate nature of the training data. For example, of all the alerts generated by the detector created using OCSVM, 56.32% were labeled as false alarms. Furthermore, the voluminous data makes it challenging to fine-tune the hyper-parameters associated with these algorithms (Narayanan and Bobba 2018). Moreover, some of the related works also concluded that using similar techniques as mentioned above, makes it harder to localize the attack (Lin et al. 2018). Unsupervised learning is advocated from other works as well on the SWaT dataset (Schneider and Böttinger 2018).

Lessons learned:

1 Detectors designed using unsupervised learning algorithms can detect novel attacks but raise an unacceptable number of false alarms.
2 It is challenging to localize the anomalies when using such detectors.

Table 1 Performance analysis of anomaly detectors (Gauthama Raman et al. 2019)

S.No	Algorithm	False alarms(%)
1	OCSVM	56.32
2	IF	48.71
3	K-NN	34.87
4	PCA	64.21

Position 3: Semi-supervised approaches raise the least number of false positives and can localize the anomalies.

The physical process controlled by an ICS possesses dynamic behavior. There exists a temporal dependency among the sensor measurements collected at different time instances. Such dependency can be learned effectively using the application of regression models, i.e., a semi-supervised approach. One such approach uses a Multi-Layer Perceptron (MLP) based anomaly detector (Raman MR et al. 2020). In this work, the sensor measurements are predicted using their past values through MLP, and the difference between the actual and predicted values is analyzed using the well-known CUSUM approach for anomaly detection. Using this approach individual models are created for each flow meter and water level sensors of SWaT, and their behaviors are monitored for anomaly detection. Using this approach, around 99.91% of anomalies against these components were detected; no false alarms were raised.

Similar observations are made by other studies on using unsupervised or semi-supervised machine learning algorithms to detect attacks in an ICS (Kravchik and Shabtai 2018; Ahmed et al. 2018; Goh et al. 2017; Inoue et al. 2017; Huda et al. 2018; Filonov et al. 2017; Filonov et al. 2016). In particular, some of them have used data from Secure Water Treatment (SWaT) testbed (Mathur and Tippenhauer 2016). The design of an anomaly detector for ICS is treated as a "one-class classification problem" and several unsupervised learning methods are effectively employed (Inoue et al. 2017). Unsupervised learning approaches construct a baseline for normal behavior through feature learning and monitor whether the current behavior is within the specified range or not. Although these techniques can detect zero day vulnerabilities, they generate high false alarms due to the existence of several hyper-parameters and multivariate nature of ICS data. Similarly, for one class SVM, authors in Inoue et al. (2017) have fine-tuned the parameters, namely c and γ for better performance on the SWaT dataset. Although there exist several automated approaches, such as grid search, randomized search, and metaheuristic optimization techniques for fine tuning, a significant challenge we face is overfitting. Generally, the error rate during the validation process should be less for the trained model; higher validation error for the model trained with a large volume of data implies that the model is over-fitted. In Shalyga et al. (2018) the authors have investigated the performance of several unsupervised neural network models for anomaly detection in SWaT testbed and proposed various statistical anomaly scoring techniques to achieve minimal false alarms.

Lessons learned:

1 As these approaches need to be developed for individual state variables, they enable localization of

anomalies, i.e., the identification of components that may have resulted in the detected anomaly.
2 The detectors so created fails to detect stealthy attacks due to a lack of knowledge regarding interactions among the plant components.
3 The applicability of such detectors is limited to continuous-valued state variables.

Challenges in behavioral-based learning approaches

The existence of the high dimensional nature of ICS data, and heterogeneous components with different operating ranges, degrades the detection precision of data-based learning approaches. Contrasted with the data-based learning approach, the authors of Gauthama Raman et al. (2020); Raman MR et al. (2020) focus in the development of a behavioral-based learning approaches that include DAE, I-DCNN, and AICrit[1]. These methods capture the spatio-temporal dependencies among the state variables using the design knowledge of the plant and the historical data. As the highly correlated state variables are extracted from the plant design, modeling the functional dependencies is simplified through the application of ML techniques. Further, these approaches are found to be computationally attractive with better detection rates and can locate the area or component under threat to the plant operator for forensics. Once such a detector is built, tested, validated, and deployed in a live plant, its performance may degrade over time. This observation leads to the following challenges.

Challenge 2: Design knowledge: "Design knowledge" of a plant refers to information such as its architecture, specification of components, computing devices, and communication infrastructure. Thus, the amount and nature of design knowledge available and used impacts the performance of a behavioral-based detector. In DAE (Gauthama Raman et al. 2020), the authors designed and evaluated three variants of deep autoencoders with varying amounts of design knowledge. These are: (i) DAE_{IAD}- six AE models monitoring each stage independently, (ii) DAE_{CAD} - three AE models independently monitoring stage 1-2-3, stage 3-4-5, and stage 5-6, and (iii) DAE_{OAD} - one AE model monitoring the entire SWaT plant. These models were implemented and tested against several attacks launched during plant operation. Interestingly, DAE_{IAD} outperforms the other two variants since each AE model captures the sensor dependencies within its host stage more effectively. Further, its computational complexity is low due to its deployment across the plant. Similar observations are reported in Kravchik and Shabtai (2018) when using LSTM based autoencoders. A similar approach was used in I-DCNN and AICrit where the interactions among the components are extracted from the P&ID diagram of SWaT and modeled through the application of deep learning algorithms. I-DCNN and AICrit exhibited better performance in terms of their respective detection accuracy when compared against that of DAE_{IAD}.

Lessons learned:

1 The improvement in accuracy is due to the focus on the relationship across sensors and actuators, operational within and across different process stages.
2 The computational complexity of the ML algorithms reduces significantly due to the incorporation of design knowledge.

Challenge 3: Operational drift: Although the physical process controlled by an ICS must be kept within the specified design limits, one can expect dynamic behavior due to the time-varying operational characteristics and requirements of plant components. Generally, a plant can be operated in several modes and one such mode is the manual mode. In the manual mode, a plant operator can modify the operating range of selected components for reasons such as volatility in demand, availability of resources, and maintenance. Such changes cause the detectors to raise alerts although the affected behavior is acceptable.

As an example of operational drift, we refer to an instance where during the CISS2020-OL event[2], the storage capacity of tank T401 in SWaT was kept between 250mm to 1000mm. This was different from the actual data available in Goh et al. (2016) since the operating range indicated by the level sensor for T401, i.e., LIT401, was between 800mm to 1000mm. In such a case, AICrit raised alarms since the behavior of LIT401 did not conform to the expected. In Fig. 4, we have compared the change in the behavior of LIT401 from the year 2015 to 2020 in terms of distribution wise. Since the *P*-Value obtained from the K-S test is 0, implies both distributions are not identical. Thus AICrit trained with the 2015 dataset raises a false alarm while testing with the 2020 dataset. Due to the absence of training data corresponding to the change in behavior, the reference models of such detector become redundant and need to be updated. Another study (Zizzo et al. 2019a) independently has shown similar operational drift in the SWaT data, thus strengthening our argument. Moreover, it is important to highlight that other model based studies would suffer from the sensor drift (Kim et al. 2019; Ahmed et al. 2017). Another independent study (Kravchik and Shabtai 2021) conducted a statistical analysis using the Kolmogorov-Smirnov test (K-S test) on SWaT, WADI, and the BATADAL datasets to quantify the similarity between the probability distributions of the training and testing data. The outcome of this work has led to the avoidance

[1] https://itrust.sutd.edu.sg/research/technologies/

[2] https://itrust.sutd.edu.sg/ciss-2020-ol/

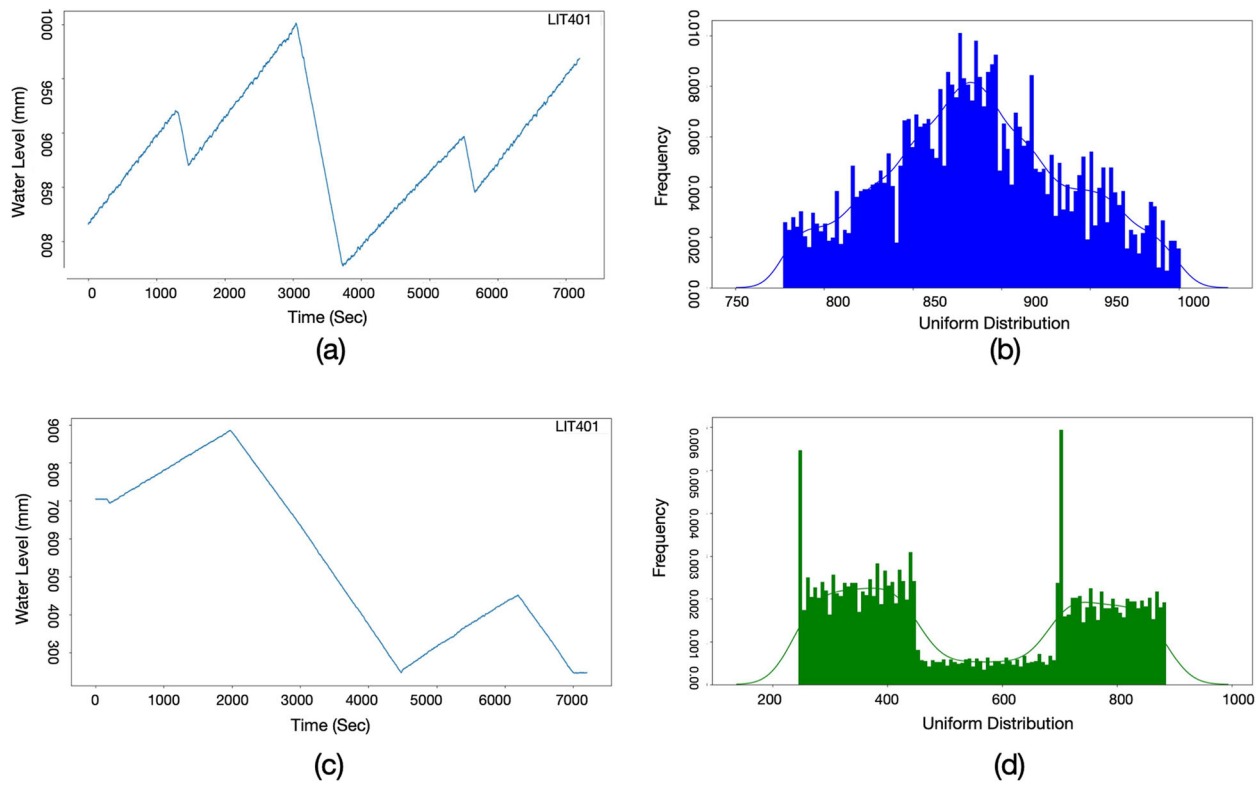

Fig. 4 Comparison of feature statistics: **a** Behavior of LIT401 in 2015; **b** normal distribution plot of **a** with mean=0.42 and variance= 0.248; **c** behaviour of LIT401 in CISS2020-OL event; **d** normal distribution plot of **c** with mean=0.53 and variance=0.37; *P*-value of K-S test (Karson 1968) is 0, hence rejecting null hypothesis i.e., the distributions are not identical

of several features (ICS components) for model creation since there exists a difference between the distribution in training and testing samples. Further, the authors claim that the absence of these features forms an important reason for the reduced false alarm rate of the proposed model.

Lessons learned:

1 Anomaly detectors should be capable of updating their reference model at regular intervals through online learning.
2 There should be an automated mechanism that initiates the retraining process when there exists a notable difference in the distribution of past data from the current dataset.

Challenge 4: Component ageing: Processes controlled by an ICS contain heterogeneous components (i.e., discrete and continuous), e.g., an OPEN-CLOSE valve and a variable speed generator. The performance of these components degrades with time and use leading to a direct impact on the detector performance. For example, a motorized valve (MV101) in SWaT connected to the inlet of tank (T101) does not close or open immediately when a PLC issues a command to change its state.

From the data available in Goh et al. (2016), it was found that the time delay for MV101 to close or open completely was 7 to 9 seconds. Using this, AICrit modeled the relationship between the MV101 and FIT101, the sensor measuring inflow rate, through the application of a decision tree. However, over five years, the delay in the change of state of MV101 increased to 12-15 seconds. Due to this change referred to as "sensor drift," AICrit generated false alarms during the CISS2020-OL event.

Lessons learned:

1 We need to deploy an automated drift detection mechanism similar to the one proposed in Baena-Garcıa et al. (2006); Zenisek et al. (2019) based on the predictive machine learning approaches.
2 Such a detection mechanism monitors the behavior of the components in real-time and reports to the plant operators when its performance degrades below an acceptable level.

Challenge 5: Noisy data and temporal glitches: The current state of the plant, i.e., sensor measurements and actuator states, are saved in a data historian at regular intervals in the supervisory control layer. This information

serves to act as a source for the anomaly detectors. Due to a variety of reasons such as human error, transmission delay, and network packet loss, there might exist noisy data or temporal glitches which lead to false alarms. It has been demonstrated that an attacker can "hide" in the noise distribution of the data (Ahmed et al. 2018). In Feng et al. (2017) the authors conclude that often machine learning algorithms miss the attacks in the noisy process data. For such a stealthy attacker it is important to consider the process noise distribution to train the detector.

Lessons learned: To overcome such issues, we introduced several parameters including a time slack variable, time window, and window size (Gauthama Raman et al. 2020). These parameters act as a buffer and if the discrepancy between the actual and prediction behavior exists for more than a specified time limit, then the alerts are generated, otherwise, they are considered as noise or glitches. We also developed an automatic packet validator that exists between the data historian and the detector for neglecting the packets with an invalid payload. By doing this, the detectors are provided with correct data to ensure the current system state is under control.

Challenge 6: Model based Learning: Taking all the process data and using it as input to machine learning algorithms is susceptible to adversarial attacks as demonstrated in Zizzo et al. (2019b); Kravchik et al. (2021). It is challenging to design the model based detectors given the persistent threat of adversarial learning. A recent work on SWaT data has deployed neural network based stealthy attack generator (Feng et al. 2017). Synthetic data spoofing is learnt for the popular process based attack detectors (Erba and Tippenhauer 2020).

Lessons learned: It is important to test not only the accuracy of model based machine learning techniques for intrusion detection but it is also critical to test the robustness against the adversarial manipulation of data input to the detector itself. It is to say that the threat model shall not only focus on the naive attacks, an attacker can execute but the more advanced stealthy attacks as highlighted above. To raise the bar and defend against an advanced attacker capable of learning the process, few solutions are proposed inspired by the classical challenge-response paradigm to ensure the non-deterministic behavior in the data (Mujeeb Ahmed et al. 2021; Ahmed et al. 2020).

Future outlook and recommendations

The challenges mentioned above, and the lessons learned from experiments on an operational plant, lead to new research directions. In the following, we make recommendations for future work based on these challenges.

Recommendation 1: Improve the transparency of the anomaly detectors: From a plant operator's point of view, most of the detectors created and deployed in operational plants behave like a black-box that inputs the current state of the plant and generate alerts indicating a process anomaly. These approaches fail to explain the semantics of the system state, i.e., "Why does the reported anomaly exist" or "Where does it exist?" or "Is the anomaly due to a cyber-attack or due to one or more faulty components?". As pointed out in Adepu and Mathur (2016), there exist several ways in which an adversary can compromise the ICS components to realize a malicious intent. As an ICS consists of several coordinated sub-processes that are monitored and controlled by multiple components, the transparency of the anomaly detector becomes an important issue. The interpretation of the detection results is crucial for plant engineers who need to make decisions to protect the underlying process from entering an undesirable state. Transparency also supports the discovery of vulnerabilities in the plant and process and aids in subsequent forensics.

Recommendation 2: Are the detection and false alarm rates adequate for evaluating anomaly detectors for ICS? Traditionally, the performance of machine learning algorithms was evaluated using metrics such as classification accuracy, precision, recall, and F1 score. Further, these metrics are computed from the values of true positive, true negative, false positive, and false negatives. In particular, in an anomaly detector, the two most significant metrics that we utilized are rates of detection and false alarms. Several works mentioned in this article aim to have a higher detection rate with minimal false alarms. However, these two metrics alone cannot comprehensively evaluate the performance of the anomaly detector designed for deployment in large continuously operational plants.

A successful attack on an ICS may cause catastrophic failures with a substantive impact on the national economy or even on human life. Thus, it is necessary to detect the anomaly due to an attack as early as possible and certainly, before the adversary's intent is realized. Hence, the detection latency should also be used as one of the evaluation metrics (Athalye et al. 2020). We have compared the performance of several statistical approaches namely CUSUM, permutation entropy, residual skewness, and Gaussian distribution integrated with the forecasting model in terms of timely detection of stealthy attacks. Several single and multi-point coordinated attacks were launched against the operational SWaT and it was found that the CUSUM approach, combined with other forecasting methods, possesses the least detection latency and can detect attacks in less than 9 seconds from the time of launch.

In Raman et al. (2019) a recommended metric, referred to as Conflict index Factor (CiF), is proposed. CiF computes the trade-off between the two conflict parameters, i.e., detection rate and false alarm rate. This metric can also be used as an evaluation metric for an anomaly

detector designed for ICS. Lower CiF values indicate better detector performance in terms of higher detection rates and the low rate of false alarms.

Recommendation 3: Base the design of an anomaly detector on domain constraints: Research reported in Priyanga et al. (2019); Krithivasan et al. (2020), focuses on the design of a generic anomaly detection system for ICS operating in different domains. Due to the similarity in the nature of the data, the design of such detectors appears feasible. However, one can argue about the generality of the ML-based detector after it is been deployed and tested across several operational ICS. Evaluating the detectors through a simulation-based environment, and validating their accuracy, does not necessarily lead to generalizable results. Further, the merits of utilizing the design knowledge in an ML-based anomaly detector are briefly discussed in Gauthama Raman et al. (2020). Thus, it is better to design an application-specific anomaly detector, than a generic one, for specific ICS to achieve better performance.

Recommendation 4: Anomaly detectors should be capable of distinguishing faults from the cyber-attacks: A physical process could enter an anomalous state due to one or more reasons. For example, it might be due to a human error, component fault, misconfiguration, and a cyber-attack. It is challenging to determine whether the reported anomaly is due to a cyber-attack or some other reason. Most ML-based anomaly detectors model the behavior of the process dynamics and detect anomalies based on the residual series generated by comparing the actual and predicted behavior. We believe that through the deep inspection of residual series, one may be able to identify the cause of an anomaly.

Summary

We are witnessing a rise in use of machine learning to design anomaly detectors for deployment in critical infrastructure such as Industrial Control Systems. While the use of machine learning enables the relatively rapid creation of the detectors when compared to the design-centric approaches, they also come with their own challenges. Several such challenges faced by the authors in their research are summarized in this article. The challenges surfaced while the authors conducted experiments with such detectors on an operational water treatment plant. To solve each challenge, additional experiments were conducted. Lessons learned from a multitude of experiments are summarized. Lastly, we make recommendations that may be useful for researchers and practitioners in the design of secure critical infrastructure.

Acknowledgements
This work was supported in part by the National Research Foundation (NRF), Prime Minister's Office, Singapore, under its National Cybersecurity R&D Programme (Award No. NRF2016NCR-NCR002-023 and NRF2018NCR-NSOE005-0001) and administered by the National Cybersecurity R&D Directorate.

Authors' contributions
All authors carry equal contributions. The author(s) read and approved the final manuscript.

Funding
Not applicable

Availability of data and materials
NA

Declarations

Competing interests
The authors declare that they have no competing interests.

Author details
[1] iTrust, Singapore University of Technology and Design (SUTD), 8 Somapah Rd, Singapore 487372, Singapore. [2] University of Strathclyde, 16 Richmond St, Glasgow G1 1XQ, United Kingdom. [3] iTrust, Singapore University of Technology and Design (SUTD), 8 Somapah Rd, Singapore 487372, Singapore.

Received: 18 October 2020 Accepted: 29 April 2021
Published online: 02 August 2021

References

Adepu S, Mathur A (2016) Generalized attacker and attack models for cyber physical systems. In: 2016 IEEE 40th annual computer software and applications conference (COMPSAC), vol 1. IEEE. pp 283–292
Adepu S, Mathur A (2018) Distributed attack detection in a water treatment plant: Method and case study. IEEE Trans Dependable Secure Comput:1–1
Ahmed CM, Gauthama Raman MR, Mathur AP (2020) Challenges in machine learning based approaches for real-time anomaly detection in industrial control systems. In: Proceedings of the 6th ACM on Cyber-Physical System Security Workshop
Ahmed CM, Mathur AP, Ochoa M (2020) Noisense print: detecting data integrity attacks on sensor measurements using hardware-based fingerprints. ACM Trans Priv Secur(TOPS) 24(1):1–35
Ahmed CM, Murguia C, Ruths J (2017) Model-based attack detection scheme for smart water distribution networks. In: Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security, ASIA CCS '17. ACM, New York, NY, USA. pp 101–113. https://doi.org/10.1145/3052973.3053011
Ahmed CM, Prakash J, Qadeer R, Agrawal A, Zhou J (2020) Process skew: Fingerprinting the process for anomaly detection in industrial control systems. In: Proceedings of the 13th ACM Conference on Security and Privacy in Wireless and Mobile Networks, WiSec '20. Association for Computing Machinery, New York, NY, USA. pp 219–230. https://doi.org/10.1145/3395351.3399364
Ahmed CM, Zhou J (2020) Challenges and opportunities in CPS security: A physics-based perspective. arXiv preprint arXiv:2004.03178
Ahmed CM, Zhou J, Mathur AP (2018) Noise matters: Using sensor and process noise fingerprint to detect stealthy cyber attacks and authenticate sensors in CPS. In: Proceedings of the 34th Annual Computer Security Applications Conference, ACSAC 2018, San Juan, PR, USA, December 03-07, 2018. pp 566–581
Athalye S, Ahmed CM, Zhou J (2020) A tale of two testbeds: A comparative study of attack detection techniques in CPS. In: Rashid A, Popov P (eds). Critical Information Infrastructures Security. Springer, Cham. pp 17–30
Baena-Garcıa M, del Campo-Ávila J, Fidalgo R, Bifet A, Gavalda R, Morales-Bueno R (2006) Early drift detection method. In: Fourth International Workshop on Knowledge Discovery from Data Streams Vol. 6. pp 77–86
Bhamare D, Zolanvari M, Erbad A, Jain R, Khan K, Meskin N (2020) Cybersecurity for industrial control systems: A survey. Comput Secur 89:101677
Brook P (2001) Ethernet/IP Industrial Protocol White Paper. IEEE EFTA
Case, Defense Use (2016) Analysis of the cyber attack on the Ukrainian power grid. Electricity Information Sharing and Analysis Center (E-ISAC):388

Drias Z, Serhrouchni A, Vogel O (2015) Taxonomy of attacks on industrial control protocols. In: 2015 International Conference on Protocol Engineering (ICPE) and International Conference on New Technologies of Distributed Systems (NTDS). IEEE

Erba A, Tippenhauer NO (2020) No Need to Know Physics: Resilience of Process-based Model-free Anomaly Detection for Industrial Control Systems. arXiv preprint arXiv:2012.03586

Feng C, Li T, Chana D (2017) Multi-level anomaly detection in industrial control systems via package signatures and LSTM networks. In: 2017 47th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN). IEEE. pp 261–272

Feng X, Li Q, Wang H, Sun L (2016) Characterizing industrial control system devices on the internet. In: 2016 IEEE 24th International Conference on Network Protocols (ICNP). pp 1–10. https://doi.org/10.1109/ICNP.2016.7784407

Feng C, Li T, Zhu Z, Chana D (2017) A deep learning-based framework for conducting stealthy attacks in industrial control systems. arXiv preprint arXiv:1709.06397

Filonov P, Kitashov F, Lavrentyev A (2017) Rnn-based early cyber-attack detection for the tennessee eastman process. arXiv preprint arXiv:1709.02232

Filonov P, Lavrentyev A, Vorontsov A (2016) Multivariate industrial time series with cyber-attack simulation: Fault detection using an lstm-based predictive data model. arXiv preprint arXiv:1612.06676

Gaj P, Jasperneite J, Felser M (2013) Computer communication within industrial distributed environment–A survey. IEEE Trans Ind Inform 9(1):182–189. https://doi.org/10.1109/TII.2012.2209668

Gauthama Raman MR, Dong W, Mathur A (2020) Deep autoencoders as anomaly detectors: Method and case study in a distributed water treatment plant. Comput Secur 99:102055. https://doi.org/10.1016/j.cose.2020.102055

Gauthama Raman MR, Somu N, Kirthivasan K, Liscano R, Shankar Sriram VS (2017) An efficient intrusion detection system based on hypergraph - genetic algorithm for parameter optimization and feature selection in support vector machine. Knowl-Based Syst 134:1–12. https://doi.org/10.1016/j.knosys.2017.07.005

Gauthama Raman MR, Somu N, Mathur AP (2019) Anomaly detection in critical infrastructure using probabilistic neural network. In: Shankar Sriram VS, Subramaniyaswamy V, Sasikaladevi N, Zhang L, Batten L, Li G (eds). Applications and Techniques in Information Security. Springer, Singapore. pp 129–141

Goh J, Adepu S, Tan M, Lee ZS (2017) Anomaly detection in cyber physical systems using recurrent neural networks. In: 2017 IEEE 18th International Symposium on High Assurance Systems Engineering (HASE). IEEE. pp 140–145

Goh J, et al. (2016) A dataset to support research in the design of secure water treatment systems. International conference on critical information infrastructures security. Springer, Cham

Han S, Xie M, Chen H-H, Ling Y (2014) Intrusion detection in cyber-physical systems: Techniques and challenges. IEEE Syst J 8(4):1052–1062

Huda S, Yearwood J, Hassan MM, Almogren A (2018) Securing the operations in SCADA-IoT platform based industrial control system using ensemble of deep belief networks. Appl Soft Comput 71:66–77

Inoue J, Yamagata Y, Chen Y, Poskitt CM, Sun J (2017) Anomaly detection for a water treatment system using unsupervised machine learning. In: 2017 IEEE International Conference on Data Mining Workshops (ICDMW). IEEE. pp 1058–1065

Karson M (1968) Handbook of Methods of Applied Statistics. Volume I: Techniques of Computation Descriptive Methods, and Statistical Inference. Volume II: Planning of Surveys and Experiments. IM Chakravarti, RG Laha, and J. Roy, New York, John Wiley; 1967, $9.00:1047–1049

Kim J, Yun JH, Kim HC (2019) Anomaly detection for industrial control systems using sequence-to-sequence neural networks. In: Computer Security. Springer, Cham. pp 3–18

Kravchik M, Biggio B, Shabtai A (2021) Poisoning attacks on cyber attack detectors for industrial control systems. In: Proceedings of the 36th Annual ACM Symposium on Applied Computing. pp 116–125

Kravchik M, Shabtai A (2018) Detecting cyber attacks in industrial control systems using convolutional neural networks. In: Proceedings of the 2018 Workshop on Cyber-Physical Systems Security and PrivaCy

Kravchik M, Shabtai A (2021) Efficient cyber attack detection in industrial control systems using lightweight neural networks and pca. IEEE Trans Dependable Secure Comput

Krithivasan K, Priyanga S, Shankar Sriram VS (2020) Detection of Cyberattacks in Industrial Control Systems Using Enhanced Principal Component Analysis and Hypergraph-Based Convolution Neural Network (EPCA-HG-CNN). IEEE Trans Ind Appl 56(4):4394–4404

Langner R (2011) Stuxnet: Dissecting a cyberwarfare weapon. IEEE Secur Priv 9(3):49–51. https://doi.org/10.1109/MSP.2011.67

Lin Q, Adepu S, Verwer S, Mathur A (2018) Tabor: A graphical model-based approach for anomaly detection in industrial control systems. In: Proceedings of the 2018 on Asia Conference on Computer and Communications Security, ASIACCS '18. Association for Computing Machinery, New York, NY, USA. pp 525–536. https://doi.org/10.1145/3196494.3196546

Mathur AP, Tippenhauer NO (2016) SWaT: A water treatment testbed for research and training on ICS security. In: International Workshop on Cyber-physical Systems for Smart Water Networks (CySWater). IEEE, USA. pp 31–36

Mirian A, Ma Z, Adrian D, Tischer M, Chuenchujit T, Yardley T, Berthier R, Mason J, Durumeric Z, Halderman JA, Bailey M (2016) An internet-wide view of ics devices. In: 2016 14th Annual Conference on Privacy, Security and Trust (PST). pp 96–103. https://doi.org/10.1109/PST.2016.7906943

Mitchell R, Chen I-R (2014) A survey of intrusion detection techniques for cyber-physical systems. ACM Comput Surv (CSUR) 46(4):1–29

Mujeeb Ahmed C, Ochoa M, Zhou J, Mathur A (2021) Scanning the Cycle: Timing-based Authentication on PLCs. arXiv e-prints. Feb:arXiv-2102

Narayanan V, Bobba RB (2018) Learning based anomaly detection for industrial arm applications. In: Proceedings of the 2018 Workshop on Cyber-Physical Systems Security and PrivaCy, CPS-SPC '18. Association for Computing Machinery, New York, NY, USA. pp 13–23. https://doi.org/10.1145/3264888.3264894

Priyanga S, Gauthama Raman M, Jagtap SS, Aswin N, Kirthivasan K, Shankar Sriram V (2019) An improved rough set theory based feature selection approach for intrusion detection in SCADA systems. J Intell Fuzzy Syst 36:1–11

Raman MR G, Somu N, Mathur AP (2020) A multilayer perceptron model for anomaly detection in water treatment plants. Int J Crit Infrastruct Prot 31:100393. https://doi.org/10.1016/j.ijcip.2020.100393

Raman MRG, Somu N, Jagarapu S, Manghnani T, Selvam T, Krithivasan K, Sriram VSS (2019) An efficient intrusion detection technique based on support vector machine and improved binary gravitational search algorithm. Artif Intell Rev 53:3255–3286

Raman MRG, Somu N, Kirthivasan K, Sriram VSS (2017) A hypergraph and arithmetic residue-based probabilistic neural network for classification in intrusion detection systems. Neural Netw 92:89–97. https://doi.org/10.1016/j.neunet.2017.01.012

Schiffer V, Vangompel DJ, Voss R (2006) The common industrial protocol (CIP) and the family of CIP networks. ODVA, Milwaukee

Schneider P, Böttinger K (2018) High-performance unsupervised anomaly detection for cyber-physical system networks. In: Proceedings of the 2018 Workshop on Cyber-Physical Systems Security and PrivaCy, CPS-SPC '18. Association for Computing Machinery, New York. pp 1–12. https://doi.org/10.1145/3264888.3264890

Shalyga D, Filonov P, Lavrentyev A (2018) Anomaly detection for water treatment system based on neural network with automatic architecture optimization. In: ICML Workshop for Deep Learning for Safety-Critical in Engineering Systems. pp 1–9

Stouffer K, et al. (2014) NIST special publication 800-82, revision 2: Guide to industrial control systems (ICS) security. National Institute of Standards & Technology

Wang Q, Chen H, Li Y, Vucetic B (2019) Recent advances in machine learning-based anomaly detection for industrial control networks. In: 2019 1st International Conference on Industrial Artificial Intelligence (IAI). pp 1–6

Williams TJ (1993) The purdue enterprise reference architecture. In: Proceedings of the JSPE/IFIP TC5/WG5.3 Workshop on the Design of Information Infrastructure Systems for Manufacturing, DIISM '93. North-Holland Publishing Co., Amsterdam, The Netherlands, The Netherlands. pp 43–64. http://dl.acm.org/citation.cfm?id=647134.716786

Zenisek J, Holzinger F, Affenzeller M (2019) Machine learning based concept drift detection for predictive maintenance. Comput Ind Eng 137:106031

Zizzo G, Hankin C, Maffeis S, Jones K (2019) Intrusion detection for industrial control systems: Evaluation analysis and adversarial attacks. arXiv preprint arXiv:1911.04278

Zizzo G, Hankin C, Maffeis S, Jones K (2019) Invited: Adversarial machine learning beyond the image domain. In: 2019 56th ACM/IEEE Design Automation Conference (DAC). pp 1–4

Publisher's Note