Using datasets from industrial control systems for cyber security research and education

Lin, Qin; Verwer, Sicco; Kooij, Robert; Mathur, Aditya

DOI
10.1007/978-3-030-37670-3_10

Publication date
2020

Document Version
Final published version

Published in
Critical Information Infrastructures Security - 14th International Conference, CRITIS 2019, Revised Selected Papers

Using Datasets from Industrial Control Systems for Cyber Security Research and Education

Qin Lin[1], Sicco Verwer[1], Robert Kooij[1,2(✉)], and Aditya Mathur[2,3]

[1] Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Delft, The Netherlands
{q.lin,s.e.verwer}@tudelft.nl
[2] iTrust Centre for Research in Cyber Security, Singapore University of Technology and Design, Tampines, Singapore
{robert_kooij,aditya_mathur}@sutd.edu.sg
[3] Computer Science, Purdue University, West Lafayette, USA

Abstract. The availability of high-quality benchmark datasets is an important prerequisite for research and education in the cyber security domain. Datasets from realistic systems offer a platform for researchers to develop and test novel models and algorithms. Such datasets also offer students opportunities for active and project-centric learning. In this paper, we describe six publicly available datasets from the domain of Industrial Control Systems (ICS). Five of these datasets are obtained through experiments conducted in the context of operational ICS while the sixth is obtained from a widely used simulation tool, namely EPANET, for large scale water distribution networks. This paper presents two studies on the use of the datasets. The first study uses the dataset from a live water treatment plant. This study leads to a novel and explainable anomaly detection method based upon Timed Automata and Bayesian Networks. The study conducted in the context of education made use of the water distribution network dataset in a graduate course on cyber data analytics. Through an assignment, students explored the effectiveness of various methods for anomaly detection. Research outcomes and the success of the course indicate an appreciation in the research community and positive learning experience in education.

Keywords: Cyber security · Research and education · Cyber-physical systems · Industrial Control Systems · Cyber Data Analytics · Anomaly detection

1 Introduction

A broad section of cyber security experts from the government, industry and the academia tend to agree that cybercrime has long evolved from an emerging threat to one that is urgent and critical. This is reflected in the estimates of economic losses due to cybercrime. Two estimates of the global annual cost of cybercrime

S. Nadjm-Tehrani (Ed.): CRITIS 2019, LNCS 11777, pp. 122–133, 2020.
https://doi.org/10.1007/978-3-030-37670-3_10

by Symantec [29] and McAfee [16] in 2017 range from $172B to $600B, respectively. In another study [2], set up as a scientific framework for computing the economic cost of cybercrime, the estimated cost was $225B in 2012.

Cyber threats have now infiltrated the domain of cyber-physical systems (CPS) [11,33]. Such systems consist of an Industrial Control System (ICS) that monitors and controls the behavior of the underlying physical process in a CPS through interactions with a network of sensors and actuators. The focus of this work is on distributed ICS found specifically in critical infrastructure such as electrical power grids, water treatment and distribution systems, and transportation systems.

Most countries have responded to the increasing cyber threats by establishing a National Cyber Security Strategy (NCSS). We refer to [18] for a comparison between 19 NCSSs. Typically, such strategies include strengthening the resilience of the Critical Information Infrastructure, the development of a vibrant cyber security ecosystem comprising a skilled workforce, technologically-advanced companies, and strong research collaborations. As cyber threats are borderless, such strategies also include efforts to forge strong international partnerships at a continental or even global level (see for instance [5,7,8]) An important pillar of a cyber security strategy is the ramping up of efforts on research and education. For instance, Singapore's National Research Foundation launched a $130 million, 5-year programme in 2014, with the aim to develop R&D expertise and capabilities in cyber security [25].

In recent years cyber security solutions have started to deploy big data analytics to correlate security events across multiple data sources, providing, amongst others, early detection of suspicious activities. According to a recent survey [3], 90% of those working in cyber security are certain that in a few years Cyber Data Analytics will play a critical role in their field.

Methods employed in the field of Cyber Data Analytics are predominantly based on Machine Learning (ML). Significant amounts of data is needed to train the ML models. While one could generate such data synthetically through simulations, data generated from operational plants is likely to offer more realistic scenarios and challenges to algorithms for training ML models. However, concerned about the safety and privacy of their plants and customers, plant owners are often reluctant to share their data. Though a number of CPS testbeds have been created, e.g., [10,20], the majority of these contain simulated and emulated components. In addition, according to [1], none of the available testbeds openly share data for research and education.

To support research and education in the design of secure CPS, iTrust, a Centre for Research in Cyber Security [13], which belogs to the Singapore University of Technology and Design (SUTD), designed and built three testbeds, that are functional replicas of their larger counterparts. These testbeds are Secure Water Treatment (SWaT), Water Distribution (WADI), and Electric Power Intelligent Control (EPIC). Although the testbeds are scaled down replicas, they contain the essential elements of fully operational critical infrastructure that support cities. The data generated at these testbeds is available [14] for use in research and education.

A recent survey summarized several public datasets from the ICS domain, that can be used for cyber security research [4]. Morris describes datasets with data collected from power systems, gas pipelines, water storage systems and energy management systems [24]. The majority of the Morris datasets [21–23], along with others such as [6,15], contain only network traffic data. The most related data is contained in the Morris-1 dataset [27]; it has been used to apply machine learning to anomaly detection at the power system's process level. We have not found any evidence that the datasets described above, have been downloaded at a large scale, been applied in research competitions or have been used for educational purposes, as is the case for the iTrust datasets.

The aim of this paper is to showcase how the availability of data from live ICS, as well as from realistic simulations, contributes to research and education in the field of cyber security. Towards this end we describe how the dataset from SWaT, a live water treatment plant, has been used to construct an innovative anomaly detection method. In addition we show how simulated data from an international competition around a fictional C-Town water distribution system, has been used effectively in cyber security education.

Contributions: (a) A summary of the impact of six publicly available datasets from operational ICS in both research and education. (b) A case study to understand the impact of using data from ICS for constructing a machine learning model for anomaly detection. (c) A case study to understand the impact of using data from ICS on the learning outcomes in a cyber data analytics class.

Organization: The remainder of this work is organized as follows. Realistic datasets available for education and research are described in Sect. 2. Two datasets, namely the SWaT and the BATADAL datasets, are described in detail in this section. A case study about the use of ICS datasets for research is shown in Sect. 3. Section 4 illuminates our datasets' education value by showcasing the use of the BATADAL dataset in a cyber data analytics course taught at Delft University of Technology. Our conclusions are reported in Sect. 5.

2 Description of Datasets

2.1 Overview

The testbeds hosted at iTrust are used for research, experimentation and training, aimed at the design of secure critical infrastructure. As a contribution to the on-going effort to improve the security of legacy and new critical infrastructure, iTrust generates a large amount of data from the testbeds. The data so generated is made available to researchers across the world[1]. In this section we briefly discuss the six datasets that are currently made available by iTrust, and which can be downloaded upon request. Of these, two datasets, namely SWaT [19] (Secure Water Treatment) and BATADAL (BATtle of Attack Detection Algorithms) [31], were used to showcase the use of real-life data for research and education in cyber security.

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

Secure Water Treatment (SWaT) Dataset. The data collected from the testbed consists of 11 days of continuous operation. Seven days' worth of data was collected under normal operation while 4 days' worth of data was collected while the testbed was under attack. During the data collection, all network traffic, sensor and actuator data were stored in the historian.

S317 Dataset. An event named SUTD Security Showdown (S3) has been organized consecutively for two years since 2016. S3 has enabled researchers and practitioners to assess the effectiveness of methods and products aimed at detecting cyber attacks launched in real-time on SWaT. During S3, independent attack teams design and launch attacks on SWaT while defence teams protect the plant passively and raise alarms upon attack detection, but are refrained from blocking the attacks. Attack teams are scored according to how successful they are in performing attacks based on specific intents while the defence teams are scored based on the effectiveness of their methods to detect the attacks.

WADI Dataset. Similar to the SWaT dataset, the data collected from the Water Distribution testbed consists of 16 days of continuous operation, of which 14 days' worth of data was collected under normal operation and 2 days with attack scenarios. During data collection, all network traffic, sensor and actuator data were collected.

EPIC Dataset. The data collected from the EPIC testbed consists of 8 scenarios under normal operation, where for each scenario, the facility is running for about 30 min. Sensor and actuator data were collected and recorded in an Excel spreadsheet, while network traffic was saved in "pcap" files.

Blaq_0 Dataset. Blaq_0 Hackathon was first organized in January 2018 for SUTD undergraduate students. Independent attack teams design and launch attacks on the EPIC testbed. Attack teams were scored according to how successful they were in performing attacks based on specific intents.

BATADAL Dataset. This dataset is not based on real-life data though is considered realistic as it was constructed using the de facto standard simulation tool for water distribution system modeling, namely the open source software package EPANET [28]. This dataset was constructed for the BATtle of Attack Detection Algorithms (BATADAL), a competition to objectively compare the performance of algorithms for the detection of cyber attacks on water distribution systems [31].

The datasets became available in 2016. As of August 30, 2019, a total of 450 download requests were received and processed. The requests originated from 52 countries, 88% of the requests originated from universities and research institutes. The remaining 12% came from industry. Given the distribution in Table 1,

SWaT dataset is the most requested. Note that some requests were for downloading multiple datasets and hence the sum of entries in the downloads column in Table 1 is more than 450.

Table 1. Downloads of iTrust datasets

Dataset	Number of downloads
SWaT	412
S317	165
WADI	178
EPIC	147
Blaq_0	81
BATADAL	95

2.2 SWaT Testbed and the Dataset

SWaT is a scaled down water treatment plant with a small footprint that produces 5 gallons/minute of doubly filtered water. The SWaT dataset was collected over 11 days of continuous operation. The first 7 days of data was collected under normal operation (without any attacks) while the remaining 4 days of data were collected with 36 designed attack scenarios. All network traffic and physical data (sensor and actuator) were collected. We focus on the detection of attacks through the analysis of physical data, hence the network traffic data is ignored. The physical data was recorded from 22/12/2015 4:00:00 PM to 2/1/2016 2:59:59 PM. The dataset contains a total of 53 columns: 1 for *timestamp*, 1 for *label* ("Attack" and "Normal"), and the remaining 51 are numeric values showing recorded data from 51 sensors and actuators. The sensors and actuators were sampled every second. The description of all 36 attack scenarios can be found on the iTrust website[2].

2.3 BATADAL Event and the Dataset

Recently Taormina et al. [30] have enhanced EPANET with a Matlab® toolbox, which enables the user to design cyber-physical attacks (CPAs) and then assess their impact on the hydraulic behavior of water distribution systems. This toolbox is dubbed epanetCPA. Using data generated with epanetCPA, Taormina et al. [31] organized the BATtle of Attack Detection ALgorithms (BATADAL). BATADAL makes use of the fictional C-Town water distribution network, first introduced for the Battle of the Water Calibration Networks by Ostfeld et al. [26]. C-Town is based on a real-world medium-sized network which contains 388 nodes, 429 pipes, 7 tanks, 11 pumps, and one actionable valve. The BATADAL

[2] https://itrust.sutd.edu.sg/dataset/.

dataset consists of three subsets. The first set contains six months of data, whose characteristics (no attacks) can be used to study the normal system operations and is labeled accordingly. The second set consists of three months of data. This dataset contains three attacks, leading to anomalous low levels in one tank, high levels in another tank and overflow in the same tank. In this set, all data are also labeled. The third set consists of only unlabeled data, while the system was running both under normal operations and during attack.

3 Case Study: Using SWaT Data to Construct a Machine Learning Model for Anomaly Detection

One of the drawbacks of general machine learning approaches is that the usage of high-dimensional data leads to opaque models. In this section, we discuss TABOR [17], a novel machine learning model for detecting cyber intrusions of ICS. TABOR is explainable due to its graphical nature, which is based upon the use of Timed Automata (TA) and Bayesian Networks (BNs). The TA is learned as a model of regular behavior of sensor signals, such as fluctuations of water levels in tanks. The BN is learned to discover dependencies between sensors and actuators. As a result, the model is easily readable and verifiable for experts and system operators. Any detection results are tractable and localizable to abnormal nodes in the model. The workflow of TABOR is as follows:

1. Sub-processes of the entire ICS are modeled. Sets of sensors and actuators in the ICS are partitioned into groups according to their locally governing PLCs for the sake of dimension and complexity reduction.
2. Signals from the sensors and actuators are symbolically represented. By doing so, on one hand, the large amount of continuous data is further compressed; on the other hand, meaningful symbols lay the foundation of learning insightful state machine models.
3. The states in the TA are associated with other actuator's states by causality inference using the BN. For example, the status (open or closed) of pumps are associated with the changes of the water level.
4. In the detection phase, irregular patterns and dependencies that do not adhere to the learned model from normal behavior, are considered anomalies.

Figure 1 shows the TA learned from the water level sensor LIT101 in the sub-process P1. In the SWaT system, the function of P1 is just raw water supply and storage - pumping raw water into the tank and pumping the water out to the next sub-process. In Fig. 1 we can observe some repeating regular behaviors such as the state transition path $S0-S1-S2-S4-S7-S1$ with the events: 3 (SU, water level Slowly goes Up)−4 (QU, water level Quickly goes Up)−2 (SC, water level Stays Constant)−1 (QD, water level Quickly goes Down)−3 (SU, water level Slowly goes Up) are discovered by the model. This typical loop is essentially a complete description of how the raw water flows into the empty tank until it is full and then flows out of the tank onto the next sub-process. The timed information is used for constructing branches with timed-varied behaviors, but with the same symbolical representation. For instance, due to the different control strategies, the water level may stay for a short or long time at its highest level.

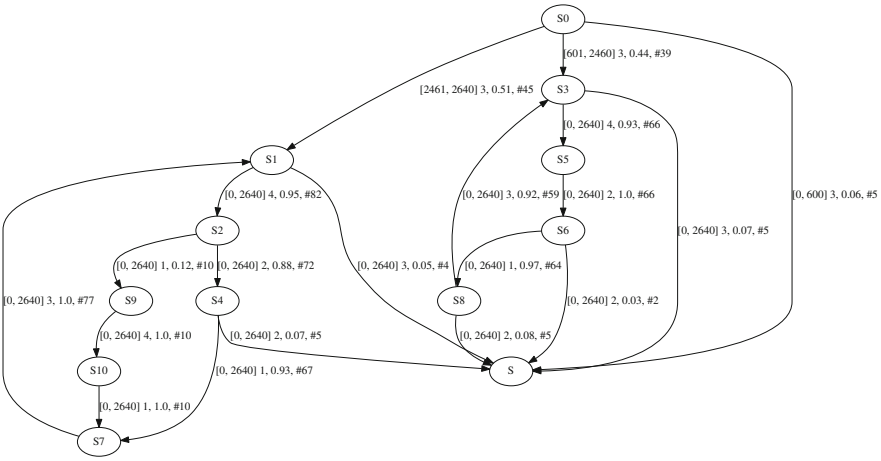

Fig. 1. Timed automaton learned from LIT101. S is the sink state, which is introduced due to fact that some sequences in the training data have very low frequencies of occurrence.

Figure 2 shows the learned BN representing the causalities among the sensors and actuators in P1. In the figure, dependencies are represented by arrows. The conditional probability distribution shows the probability distribution of a node given its parents.

In the testing phase, the new incoming data are represented by discrete events and then they are executed in the TA and BN models. Any abnormal events i.e., invalid transition/state in the TA and zero probability in the BN, are reported as anomalies. The explanation and localization of such detection results are achieved by identifying the nodes, where the anomalies occur. The explanation can be verified by the scenarios description of the SWaT dataset, where the ground-truth (starting/ending time of attacks and names of sensors/actuators under attacks) of every attack scenario is discussed in detail.

Thanks to the public availability of the SWaT dataset, it is possible to directly compare the detection performance and training/testing runtime with two published papers, whose detection methods are based upon deep neural networks [9] and SVM [12], in which exactly the same dataset was used. The results demonstrate that TABOR outperforms the other two methods in terms of effectiveness and efficiency.

4 Case Study: Using BATADAL Data in a Cyber Data Analytic Course

Since 2016 Delft University of Technology is offering a *cyber data analytics* course as part of the Data Science & Technology Track for the master's degree in Computer Science. The course provides a theoretical and practical background

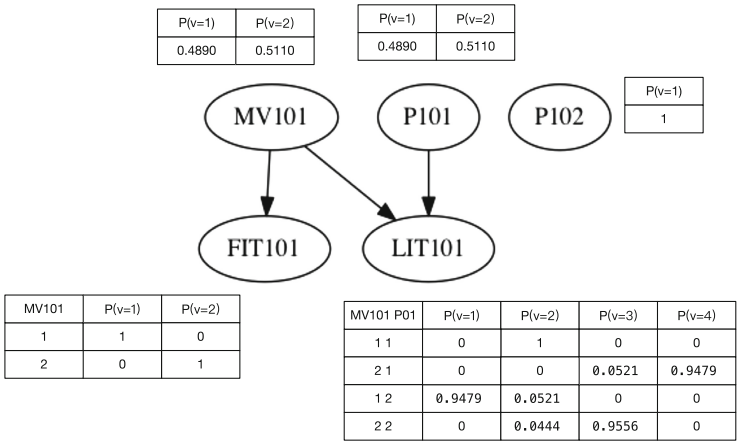

Fig. 2. Bayesian network learned from P1. The first column and the second row of the table about LIT101 indicates that given both MV101 and P101 are closed, the probability that water level quickly decreases (QD) is 0. Note that the actuators' states: open and closed, are denoted as 2 and 1, respectively.

for applying data analytics in the field of cyber security. In 2018, about 150 students registered for the course, which contained an assignment based on the BATADAL dataset. As part of the assignment, the students were asked to apply a machine learning method to detect cyber attacks on the water distribution system.

This section describes the assignments based on the BATADAL dataset and the outcomes based on submissions by the students. A total of 51 student groups took the BATADAL assignment. The results from the groups were scored in the same way as in the BATADAL competition. The score is the average of the time to detect an attack and the detection accuracy [31]. The Time-To-Detection (TTD) is the time needed by the algorithm to recognize a threat and is defined as the difference between the time t_d at which the attack is detected and the time t_0 at which the attack started:

$$TTD = t_d - t_0. \tag{1}$$

Score S_{TTD} is defined as follows to evaluate the performance of the detection algorithm for several attacks,

$$S_{TTD} = 1 - \frac{1}{m} \sum_{i}^{m} \frac{TTD_i}{\Delta t_i}, \tag{2}$$

where TTD_i is the TTD score in $i-$th attack, Δt_i is the duration of $i-$th attack scenario and m is the total number of attack scenarios. Note that for the BATADAL competition $m = 7$.

The detection accuracy S_{CM} is defined as the mean of true positive rate (TPR) and true negative rate (TNR):

$$S_{CM} = \frac{TPR + TNR}{2}. \tag{3}$$

Finally, the overall score is computed as the average S_{TTD} and S_{CM}:

$$S = \frac{S_{TTD} + S_{CM}}{2}. \tag{4}$$

Surprisingly, one group of students achieved a competitively high score, namely $S = 0.924$. This group used a combination of results from a discrete Markov model and PCA (Principal Component Analysis). This implies that the student group would be placed fourth among the experienced participants in the original BATADAL competition [31], see Table 2.

Table 2. BATADAL competition ranking

Place	Team	Attacks detected	Score (S)
1	Housh and Ohar	7	0.970
2	Abokifa et al.	7	0.949
3	Giacomoni et al.	7	0.927
4	**Student group**	**7**	**0.924**
5	Brentan et al.	7	0.894
6	Chandy et al.	7	0.802
7	Pasha et al.	7	0.773
8	Aghashahi et al.	7	0.534

A survey was conducted to assess how the students valued the use of the BATADAL dataset in their assignment. The students were asked to answer five questions related to the use of the real-life data. Because the number of respondents was low (21 students) we have refrained from using a Likert five-point scale. Instead, the respondents simply answered either "yes" or "no" to the questions. They were also allowed to give comments on their answers.

The following five questions were addressed to the students:

Q1: It is important that during the course we use real-life data.
Q2: The use of real-life data increases my understanding of the models we learn in the course.
Q3: I have used real-life data from the cyber domain before.
Q4: I would like to apply the models we learn in the course to more real-life data.
Q5: Machine learning techniques are promising to solve real-life cyber security problems.

A vast majority of the students (86%) agree with the importance of using real-life data (Q1). One student commented: "because in practice we will also use real-life data." Opinions were mixed for Q2 (57%–43%), which polls whether using real-life data increases understanding of the ML models taught in the class. One student feels there is no correlation while another finds real-life data more interesting and hence pushes to study harder. A majority of the students (76%) had not used real-life data from the cyber domain prior to coming to this course (Q3). One student comments that this is a really nice addition to the cyber security track. A majority of the students (76%) would like to apply the taught ML models to more real-life data (Q4). One student deems this important because "techniques are categorized based on their efficiency for certain datasets." Lastly, a majority of the students (95%) find ML techniques promising to solve real-life cyber security problems. The results in this section indicate how the use of real-life data augments education in cyber security.

5 Summary and Conclusions

Data analytics and machine learning classes have sprung up across the research community and many universities. Researchers and instructors in their classes often make earnest attempts to obtain realistic datasets to conduct research and to teach the students. Unfortunately, there are few known ICS datasets available for use. iTrust makes available several such datasets, two of which were used in this work. The objective of the study reported here was to understand how the use of realistic datasets from live and simulated ICS enhance the research and student learning.

An analysis of the results from the two offerings answer some, though not all, questions a researcher or an instructor may pose. First, based on the fruitful research outcome using a live dataset such as SWaT, researchers appreciated a practical platform to develop and test their algorithms. The competition among variate machine learning techniques boosts the flourish of advanced intrusion systems protecting critical CPSs. Second, based on the responses from the survey, we can claim that students appreciated the use of realistic data from an ICS, the simulated BATADAL dataset. However, we do not have any statistical evidence that supports, or does not support, a claim that the use of live data enhances learning of data analytic techniques used in this study. However, we believe that the use of data from an operational plant, SWaT in this study, enhances student motivation and hence learning. Details of SWaT plant are public [14] and thus the students can discuss the pros and cons of using machine learning techniques in detecting process anomalies in a physical context. Such discussions also add to student's knowledge of how ICS operates and the inherent vulnerabilities that could be exploited leading to process anomalies.

Given the above conclusions, we believe that this work is a step towards a more detailed study that would focus on a better understanding of the impact of using live ICS data on research and student learning. Such a study would require both live ICS data as well as synthetic data. A significant amount of synthetic

data is already available in the public domain, e.g. the "Electrical Grid Stability Simulated Data" from UC Irvine [32]. Such data can be used, along with live datasets available from iTrust, to conduct a deeper study with research and educational objectives similar to those in the study reported here.

Acknowledgements. This work is partially supported by Technologiestichting STW VENI project 13136 (MANTA), NWO project 62001628 (LEMMA) and the 2+2 PhD program of TUD and SUTD. This work was also supported by the National Research Foundation (NRF), Prime Minister's Office, Singapore, under its National Cybersecurity R&D Programme (Award No. NRF2014NCR-NCR001-040) and administered by the National Cybersecurity R&D Directorate. The testbeds were made possible through funding from Ministry of Defence, Singapore, NRF and the SUTD-MIT International Design Centre (IDC). The authors thank Mark Goh for maintaining the iTrust datasets and processing requests for downloads.

References

1. Almgren, M., et al.: RICS-el: building a national testbed for research and training on SCADA security (short paper). In: Luiijf, E., Žutautaitė, I., Hämmerli, B.M. (eds.) CRITIS 2018. LNCS, vol. 11260, pp. 219–225. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-05849-4_17
2. Anderson, R., et al.: Measuring the cost of cybercrime. In: Proceedings of the 11th Workshop on Economics of Information Security (2012)
3. Balaganski, A., Derwisch, S.: Big data and information security. KuppingerCole and BARC Joint Study, Report No.: 7400 (2016)
4. Choi, S., Yun, J.-H., Kim, S.-K.: A comparison of ICS datasets for security research based on attack paths. In: Luiijf, E., Žutautaitė, I., Hämmerli, B.M. (eds.) CRITIS 2018. LNCS, vol. 11260, pp. 154–166. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-05849-4_12
5. Council of the European Union: European council, council directive 2016/1148 of 6 July 2016 concerning measures for a high common level of security of network and information systems across the union (2016). https://eur-lex.europa.eu/eli/dir/2016/1148/oj
6. Digitalbond: S4x15 ICS village CTF dataset (2015). https://www.digitalbond.com/blog/2015/03/16/s4x15-ctf-ics-village-page/
7. G8: G8 principles for protecting critical information infrastructures (2003). http://www.cybersecuritycooperation.org/documents/G8_CIIP_Principles.pdf
8. GFCE: global forum on cyber expertise (2015). https://www.thegfce.com/about
9. Goh, J., Adepu, S., Tan, M., Lee, Z.S.: Anomaly detection in cyber physical systems using recurrent neural networks. In: 2017 IEEE 18th International Symposium on High Assurance Systems Engineering (HASE), pp. 140–145. IEEE (2017)
10. Holm, H., Karresand, M., Vidström, A., Westring, E.: A survey of industrial control system testbeds. In: Buchegger, S., Dam, M. (eds.) Secure IT Systems, vol. 9417, pp. 11–26. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-26502-5_2
11. ICS-CERT: Cyber-attack against Ukrainian critical infrastructure (2016). https://ics-cert.us-cert.gov/alerts/IR-ALERT-H-16-056-01
12. Inoue, J., Yamagata, Y., Chen, Y., Poskitt, C.M., Sun, J.: Anomaly detection for a water treatment system using unsupervised machine learning. In: 2017 IEEE International Conference on Data Mining Workshops (ICDMW), pp. 1058–1065. IEEE (2017)

13. iTrust: Centre for Research in Cyber Security (2015). https://itrust.sutd.edu.sg/
14. iTrust: Secure Water Treatment (SWaT) Testbed (2015). https://itrust.sutd.edu.sg/research/dataset/
15. Lemay, A., Fernandez, J.M.: Providing {SCADA} network data sets for intrusion detection research. In: 2016 9th Workshop on Cyber Security Experimentation and Test ({CSET}) (2016)
16. Lewis, J.A.: Economic impact of cybercrime-no slowing down (2018). https://www.csis.org/analysis/economic-impact-cybercrime
17. Lin, Q., Adepu, S., Verwer, S., Mathur, A.: 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, pp. 525–536. ACM (2018)
18. Luiijf, E., Besseling, K., De Graaf, P.: Nineteen national cyber security strategies. Int. J. Crit. Infrastruct. (IJCIS) **9**(1/2), 3–31 (2013)
19. Mathur, A.P., Tippenhauer, N.: SWaT: a water treatment testbed for research and training on ICS security. In: International Workshop on Cyber-physical Systems for Smart Water Networks (CySWater), pp. 31–36. IEEE, USA, April 2016
20. McLaughlin, S., et al.: The cybersecurity landscape in industrial control systems. Proc. IEEE **104**(5), 1039–1057 (2016)
21. Morris, T., Gao, W.: Industrial control system traffic data sets for intrusion detection research. In: Butts, J., Shenoi, S. (eds.) ICCIP 2014. IAICT, vol. 441, pp. 65–78. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-662-45355-1_5
22. Morris, T., Srivastava, A., Reaves, B., Gao, W., Pavurapu, K., Reddi, R.: A control system testbed to validate critical infrastructure protection concepts. Int. J. Crit. Infrastruct. Prot. **4**(2), 88–103 (2011)
23. Morris, T.H., Thornton, Z., Turnipseed, I.: Industrial control system simulation and data logging for intrusion detection system research. In: 7th Annual Southeastern Cyber Security Summit, pp. 3–4 (2015)
24. Morris, T.: Industrial control system (ICS) cyber attack datasets (2015). https://sites.google.com/a/uah.edu/tommy-morris-uah/ics-data-sets
25. NRF: Singapore, national cybersecurity R&D programme (2013). https://www.nrf.gov.sg/programmes/national-cybersecurity-r-d-programme
26. Ostfeld, A., et al.: Battle of the water calibration networks. J. Water Resour. Plan. Manag. **138**(5), 523–532 (2012)
27. Pan, S., Morris, T., Adhikari, U.: Developing a hybrid intrusion detection system using data mining for power systems. IEEE Trans. Smart Grid **6**(6), 3104–3113 (2015)
28. Rossman, L.A.: EPANET 2: User Manual (2000)
29. Symantec: Norton cyber security insights report, global results (2017). https://www.symantec.com/content/dam/symantec/docs/about/2017-ncsir-global-results-en.pdf
30. Taormina, R., Galelli, S., Tippenhauer, N.O., Salomons, E., Ostfeld, A.: Characterizing cyber-physical attacks on water distribution systems. J. Water Resour. Plan. Manag. **143**(5), 04017009 (2017)
31. Taormina, R., et al.: The battle of the attack detection algorithms: disclosing cyber attacks on water distribution networks. J. Water Resour. Plan. Manag. **144**(8), 1–11 (2018)
32. UC Irvine: Machine learning repository (2007). https://archive.ics.uci.edu/ml/index.php
33. Weinberger, S.: Computer security: is this the start of cyberwarfare? Nature **174**, 142–145 (2011)